Evolution and regulation of the *Early Lactation Protein* gene in an Australian marsupial, the tammar wallaby (*Macropus eugenii*)



Elizabeth Anne Pharo

orcid.org/0000-0002-5573-1746

Submitted in total fulfilment of the requirements of degree of

Doctor of Philosophy

April 2016

School of BioSciences,

Faculty of Science,

The University of Melbourne

Produced on archival quality paper

Abstract

Marsupials have developed a very different reproductive strategy to eutherians. The Australian marsupial, the tammar wallaby (*Macropus eugenii*) has a very short active pregnancy of ~26.5 days, (at the end of which they enter Phase 1 of lactation), with a comparatively long lactation of ~300-350 days. The tammar mother gives birth to an altricial ~400 mg young which spends the first 200 days postpartum (pp) in its mother's pouch, permanently (0-100 days pp, Phase 2A) and then intermittently (100-200 days pp, Phase 2B) attached to the teat. The commencement of Phase 3 marks the first exit from the pouch (in some ways akin to the birth of a precocious eutherian neonate) and the supplementation of milk with herbage. The marsupial mother progressively alters milk composition and individual milk constituents to provide nutrients that are presumed to be appropriate for the considerable neurological and physiological development and growth of her young.

During each Phase of this long and complex lactation, specific proteins are produced. One of these, early lactation protein (ELP), is the main subject of this thesis. The *ELP* gene is expressed in the tammar wallaby, South American opossum and stripe-faced and fat-tailed dunnarts. Here, this study has shown that tammar *ELP* expression is minimal during pregnancy, but is upregulated at parturition and expressed for 100-125 days pp in the mammary gland. It is then down-regulated for the remainder of lactation (Phases 2B and 3). As the Phase 2A early lactation period ~0-100 days pp has no eutherian equivalent, ELP was thought to be marsupial-specific. However, the presence of a single 51 amino acid Kunitz bovine pancreatic trypsin inhibitor (BPTI) domain in both ELP and bovine colostrum trypsin inhibitor (CTI) and the fact that bovine CTI is only secreted in the colostrum for a brief 24-48 h pp period suggests that these proteins may be related. Since they are produced at the time when the neonate and pouch young initially have only an innate immune system, they may have a function in immune protection.

Comparative genomic analyses of available marsupial and eutherian genomes, retrotransposon analyses and a PCR-based cloning approach showed that *ELP* and *CTI* share the conserved single-copy *phosphatidyl inositol glycan class T (PIGT)* and *whey acidic protein four disulphide core domain 2 (WFDC2)* as flanking genes. This suggests that they evolved from a common ancestral gene at least 160 million years ago, before the divergence of marsupials and eutherians. The status of *ELP*/*CTI* in monotremes is unclear from the platypus genome. Thus this thesis highlights that *ELP* is conserved in

all marsupials thus far, including the Australidelphia (tammar, stripe-faced and fat-tailed dunnarts, Tasmanian devil, brushtail possum, koala) and Ameridelphia (South American opossum), but demonstrates that not all eutherians have a functional *CTI* gene.

Amongst eutherians, a Kunitz domain-encoding *CTI* gene has been found in mammals of the Laurasiatherian orders Carnivora (Cape fur seal, Pacific walrus, Giant panda, dog, cat) and Cetartiodactyla (cow, sheep, goat, antelope, camel and others). Interestingly, *CTI* is functional in the white rhinoceros (Order Perissodactyla) but non-functional in other species studied of this order including the Thoroughbred and Prezwalski's horses. *CTI* is also a pseudogene in at least some species of the following clades: Afrotheria (elephant), Xenarthra (sloth, armadillo), Euarchontoglires (including humans and other primates and rodents) and the Laurasiatherian orders Chiroptera (flying fox) and Eulipotyphla (hedgehog). *CTI* gene loss was most probably caused by the deletion or transposition of *CTI* gene fragments elsewhere in the genome by active retrotransposons. Retro-elements may also be responsible for the evolution of bovine *PTI*, *spleen trypsin inhibitor* (*STI*) and the five ruminant placenta-specific *trophoblast Kunitz domain protein* genes (*TKDP1-5*) from bovine *CTI*. However, while *CTI* expression is specific to the mammary gland, the *TKDP*s are expressed in the ruminant placenta only and *PTI* and *STI* are ubiquitously expressed.

The *ELP* and *CTI* genes range from ~1.4-4.8 kb with a conserved 3-exon structure, but a variable-length first intron. The first exon encodes the signal peptide and N-terminus; the second, a low-complexity coiled region and the 51 aa Kunitz domain; and the third, a short C-terminus. Although tammar *ELP* has two TATA boxes and two transcription start sites they both encode an identical 104 amino acid precursor protein. The mature 75-84 amino acid marsupial ELP and eutherian CTI peptides share 44.6-64.6% similarity, but only 43.4-55.0% with BPTI. ELP and CTI most likely form 8.47-9.68 kDa α+β fold globular proteins with 3 disulphide bonds, characteristic to the Kunitz domain. However, the conservation of an N-glycosylation site at the 'base' of the pear-shaped ELP and CTI proteins produces larger ~10-20.5 kDa glycoproteins.

This study showed that a basic P_1 'warhead' residue (lysine or arginine) is conserved in eutherian CTI. This site may be involved in serine protease inhibition, suggesting that the eutherian proteins inhibit serine endopeptidases of the S1 chymotrypsin family such as trypsin and plasmin. Structural homology modelling of the porcine trypsin-human tissue factor pathway inhibitor (TFPI) Kunitz domain 2 complex (Protein Data Bank: 1TFX) suggests that Cape fur seal CTI inhibits trypsin in a canonical 1:1 interaction characteristic to the Kunitz domain. In contrast, the lack of conservation of the marsupial ELP P_1 residue suggests that these proteins either inhibit elastase, are inactive against

trypsin, possess no protease inhibitor function or that the protease inhibitor site may be situated elsewhere in the proteins.

The endocrine and transcriptional regulation of tammar *ELP* in both 2D mammary epithelial cell culture and 3D tammar mammary gland explant culture was maximally responsive to the lactogenic hormones; insulin (I), hydrocortisone (HC) and prolactin (PRL) in Phase 1 (day 24 of pregnancy) explants. However, the maintenance of *ELP* expression after 12-day treatments with either I, or I+HC, but without PRL, was quite unlike most mammalian milk protein genes. Furthermore, the inability to manipulate *ELP* expression *in vitro* unless the gene was expressed *in vivo* (Phases 1 and 2A only) demonstrates that its regulation is complex. The importance of ECM for *ELP* expression was suggested by the lack of *ELP* response to IHCPRL in primary cells isolated from the tammar mammary gland. Similarly, tammar *ELP* promoter constructs transfected into HC11 or HEK293T immortalised cells were not responsive to lactogenic hormones *in vitro* unlike rodent *β-casein* promoter constructs.

This study provides the first information on the transcriptional regulation of *ELP* and *CTI*. Six marsupial *ELP* and twenty eutherian *CTI* proximal promoters were identified. Activation of the tammar *ELP* gene appears to be the result of mammary cell-activating factor (MAF), an E26 transformation-specific (ETS) factor binding to the CTTCCT(-) motif. Furthermore, the PRL-responsiveness of *ELP*, but the absence of a conserved STAT5A binding site for the signalling of PRL via its receptor was surprising. Additionally, the temporal, developmentally-regulated expression of both *ELP* and *CTI* suggests that chromatin-binding controls gene transcription throughout the reproductive cycle. Hence, the binding of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), estrogen receptor-related-alpha or -beta (ERR-α/-β) and an ETS repressor elements, plus "open" chromatin permits low-level *ELP* expression during pregnancy. At parturition, chromatin remains "opens" and MAF, CCAAT/enhancer binding protein (C/EBP) and/or activating protein 2 gamma (AP2γ) bind to and activate high-level *ELP* expression during Phase 2A only. The lack of *ELP* expression for the remainder of lactation and in the involuting, non-pregnant and virgin mammary gland, nor in other tissues suggests that "closed" chromatin prevents transcription factor machinery access to the promoter during these periods.

Mammalian and bacterial expression systems were evaluated for the production of recombinant tammar ELP *in vitro*. Although recombinant tammar ELP was successfully produced *in vitro* using the maltose binding protein (pMAL) system, whether ELP was

biologically active was doubtful, particularly due to the inability of the *Escherichia coli* cytoplasm to produce disulphide bonds and the lack of attached oligosaccharides. Mammalian expression systems should provide the ideal method for tammar ELP production but these studies were hampered by the lack of an antibody to tammar ELP. Neither the pCMV-SPORT6 nor pcDNA3.1 expression vectors produced visually detectable recombinant tammar ELP. In contrast, recombinant tammar cystatin C was abundantly secreted into culture media.

The conservation of *ELP* in all marsupials investigated thus far suggests that tammar ELP may have an important role during the first third of post-natal growth and development. The presence of a Kunitz domain with attached oligosaccharides suggests that ELP and the orthologous CTI could have multiple roles in the maternal mammary gland and milk and/or in the gut and circulatory system of the young. For example, ELP and CTI may act as serine protease inhibitors, soluble receptor analogues which prevent gastrointestinal infections in the young and may as well provide an important amino acid and carbohydrate source. Furthermore, the timing of ELP secretion in the tammar correlates with the time of maximum transfer of immunoglobulins via milk to the dependent pouch young (which lacks an acquired immune system until mid-lactation). In the Cape fur seal, CTI may be critical for preventing the proteolytic degradation of immunoglobulins that are passively transferred from mother to young via colostrum and milk. The retention of CTI in certain eutherians, but its loss in others, suggests that alternative mechanisms may have evolved for immune protection in those species lacking CTI. Taken together, the results of this study suggests that these proteins (and indeed the whole of milk) play a vital role in the survival of the vulnerable mammalian young and of the marsupial young in particular.

Declaration

This is to certify that

- the thesis comprises only my original work towards the PhD except where indicated in the Preface,
- due acknowledgment has been made in the text to all other material used,
- the thesis is less than 100,000 words in length, exclusive of tables, bibliographies and appendices.

Signed: **……..**

Preface

Author contributions to submitted and published works.

This thesis contains published material and material that has been submitted for publication.

Published

Pharo, E.A., De Leo, A.A., Renfree, M.B., Thomson, P.C., Lefevre, C.M., and Nicholas, K.R. (2012). The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene. *BMC Evol Biol* **12**: 8.

Author contributions: Experiments were designed, conducted and analysed by EAP, with the exception of the fluorescence *in situ* hybridization of the tammar *ELP* gene, which was performed by AAD, the statistical analysis of microarray experiments which was performed by CML and PCT and the microarray experiments. KRN and EAP dissected mammary glands from animals provided by MBR and the Department of Primary Industries, Attwood, Victoria. EAP wrote the manuscript and all authors read, edited and approved the final manuscript.

Pharo, E.A. (2014) Expression of the mammary gland-specific tammar wallaby *early lactation protein* gene is maintained *in vitro* in the absence of prolactin. *Molecular and Cellular Endocrinology*. **382:** 871-880.

Author contributions: Experiments were designed by EAP and Kevin Nicholas and conducted and analysed by EAP. EAP wrote the manuscript and Kevin R Nicholas and Marilyn B Renfree provided critical evaluation of the manuscript.

Pharo, E.A., Cane, K.N., McCoey, J., Buckle, A.M., Oosthuizen, W.H., Guinet, C., and Arnould, J.P.Y. A colostrum trypsin inhibitor gene expressed in the cape fur seal mammary gland during lactation. *Gene*. **578:** 7-16.

Author contributions: EAP performed the experiments with the exception of the structural homology modelling which was done by AMB and JM and the preparation of the fur seal mammary gland membrane for Northern analysis. JPYA, KNC, and CG dissected seal tissues in the field and WHO provided logistical support. Marilyn B Renfree provided critical evaluation of the manuscript and all authors read, edited and approved the final manuscript.

Submitted

Pharo, E.A., Renfree, M.B., and Cane, K.N. Mammary cell-activating factor regulates the hormone-independent transcription of the *Early Lactation Protein* (*ELP*) gene in a marsupial. Submitted to *Molecular and Cellular Endocrinology*.

Author contributions: EAP designed and performed the experiments with the assistance of KNC. MBR provided input into experiment design and all authors read, edited and approved the final manuscript.

Other Publications During Candidature

De Leo, A.A., Lefevre, C., Topcic, D., Pharo, E., Cheng, J.F., Frappell, P., Westerman, M., Graves, J.A.M., and Nicholas, K.R. (2006). Characterization of two whey protein genes in the Australian dasyurid marsupial, the stripe-faced dunnart (*Sminthopsis macroura*). *Cytogenet Genome Research* **115**: 62-69.

Renfree, M.B., Papenfuss, A.T., Deakin, J.E., Lindsay, J., Heider, T., Belov, K., Rens, W., Waters, P.D., **Pharo, E.A.**, Shaw, G., Wong, E.S., Lefevre, C.M., Nicholas, K.R., Kuroki, Y., Wakefield, M.J., Zenger, K.R., Wang, C., Ferguson-Smith, M., Nicholas, F.W., Hickford, D., Yu, H., Short, K.R., Siddle, H.V., Frankenberg, S.R., Chew, K.Y., Menzies, B.R., Stringer, J.M., Suzuki, S., Hore, T.A., Delbridge, M.L., Patel, H.R., Mohammadi, A., Schneider, N.Y., Hu, Y., O'Hara, W., Al Nadaf, S., Wu, C., Feng, Z.P., Cocks, B.G., Wang, J., Flicek, P., Searle, S.M., Fairley, S., Beal, K., Herrero, J., Carone, D.M., Suzuki, Y., Sugano, S., Toyoda, A., Sakaki, Y., Kondo, S., Nishida, Y., Tatsumoto, S., Mandiou, I., Hsu, A., McColl, K.A., Lansdell, B., Weinstock, G., Kuczek, E., McGrath, A., Wilson, P., Men, A., Hazar-Rethinam, M., Hall, A., Davis, J., Wood, D., Williams, S., Sundaravadanam, Y., Muzny, D.M., Jhangiani, S.N., Lewis, L.R., Morgan, M.B., Okwuonu, G.O., Ruiz, S.J., Santibanez, J., Nazareth, L., Cree, A., Fowler, G., Kovar, C.L., Dinh, H.H., Joshi, V., Jing, C., Lara, F., Thornton, R., Chen, L., Deng, J., Liu, Y., Shen, J.Y., Song, X.Z., Edson, J., Troon, C., Thomas, D., Stephens, A., Yapa, L., Levchenko, T., Gibbs, R.A., Cooper, D.W., Speed, T.P., Fujiyama, A., Graves, J.A., O'Neill, R.J., Pask, A.J., Forrest, S.M., and Worley, K.C. (2011). Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol* **12**: R81. **[Cover article, Highly Accessed]**

Acknowledgements

I would like to extend my huge appreciation and gratitude to Marilyn Renfree who stepped in and provided a terrific mentor, supervisor. Thank you for your encouragement, generosity, patience, friendship and both gentle and not so gentle persuasion to get my papers submitted and thesis handed in. I really appreciate your effort in helping me improve my writing, to be more concise and to tell a story … even though I bled over each paragraph you cut! I couldn't have done it without you!

Thank you to Kevin Nicholas for providing a really interesting project on tammar wallaby lactation. I really enjoyed our discussions on tammar lactation and milk production. In particular, thank you for making me become an independent researcher. A thank you to Laura Parry for your help and advice.

A special mention to those at VIAS: Jane Whitley for teaching me how to make terrific RNA, Jo Argento for schooling me in molecular biology and to Michelle Wells, Sonia Mailer and Jenny Carfi for all their help with tissue and cell culture work. Thanks to members of my lab group at the Department of Zoology, The University of Melbourne for all your support and help, particularly Coralie Reich and Kylie Cane when things got tough. Sonia provided great help for microarrays and primary and immortalised cell culture techniques … some enjoyable days at the winery. To my mammary gland explant cutters: Jo Argento, Jenny Carfi, Kevin, Amelia Brennan, Kylie, Karensa Menzies, Sonia, Coralie; my hands are eternally grateful to you all for the help with cutting! I am also very grateful to Christophe Lefèvre for his help and advice on bioinformatics and to Danny Topcic for providing RNA and plasmids, as well as entertainment through his band Kretch.

To past and present members of the Walgroup, I really appreciate you guys for making me feel so welcome and for all your moral support. Special thanks to Geoff Shaw for his calming influence and computer help, to Andrew Pask for his boundless enthusiasm and energy, to Keng for his friendship and computer chats, to Brandon Menzies for interesting science discussions and to Terry Fletcher for his help with gel filtration. To other 'Walgroupies': Jenni Hetz, Karen Hansen, Stephen 'Frankers' Frankenberg, Danielle Hickford, Hongshi Yu, Natalie Calatayud, Monika Paranjpe, Cynthia Martin, Yu Chen, Yanqiu Hu, Shunsuke Suzuki Suzuki, Jess Stringer, Nanette Schneider, Bonnie Dopheide, Sue Osborn, Alison Bradfield, Eugene Fredericks, Francois Monty thanks for your support. A huge thanks to Scott Brownlees, Kerry Martin and Greg Cranston for looking after the wallabies so well.

Thanks to Jenni for your optimism and encouragement in the lab and to you and Victor for being such great beach walk companions! A special thanks to my office mates Helen Clark (Gehring), Mai Truong and Beth Finger for putting up with my constant piles of papers, my venting and getting coffees for me in the final stages of my write-up.

To the former heads of Zoology: David Macmillan and David Gardner, and to Mary Familari, I send my sincere appreciation for your support and advice during my PhD candidature. Other Zoology staff (now BioSciences 4) also provided invaluable support: Peter Krotsis for always being able to find desks, filing cabinets, etc, Garry Jolly Kim Knowles and for computer support and David Paul for providing excellent Photoshop and Illustrator tips and allowing me to use photograph of the young tammar wallaby on the front cover of my thesis!

Tim Adams (CSIRO, Parkville) provided HEK293T cells and the C/EBP, STAT5 and PRLR expression plasmids as well as helpful advice on mammalian protein expression, while Dr Shuliang Cui provided invaluable advice on bacterial protein expression. Thank you to Dr. Hong-Jian Zhu for the use of his GloMax-96 Microplate Luminometer and provision of MDA-MB-231 cells, and to Josie Iaria for her assistance with dual luciferase assays.

I am grateful to the Australian Government for providing an Australian Postgraduate Award (APA) stipend and to the Cooperative Research Centre for Innovative Dairy Products (CRC-IDP) for a top-up a stipend. The CRC supported this research and organised terrific annual conferences and interesting workshops for Postgraduate students and funded numerous courses including: microarray analysis, bioinformatics and speed reading.

Thanks to friends who have been extremely patient and understanding and who stopped asking the dreaded question, “so when are you going to hand in?” Thank you to my good friend Coralie Reich, for encouraging me to get out and to help break my workaholic tendencies! I really appreciated the stress relief! Thanks also to friends who read through various thesis and manuscript drafts: Genevieve, Kylie and Coralie. To Natalie and Kylie, Sancho, Mia and Grrr, thanks for your friendship, support and encouragement and providing a place to stay when I needed it! You guys have been terrific. To others: Justyna, Matt and Kylie, Dave, Clare, Emile and Kath, Gratiana, Phil, Bec, thanks for some fun times cruising the Murray, Friday drinks, etc. Thanks to my ski buddies Coralie, Anthony, Wenji, Simon and many others for some terrific stress-free holidays in the ski fields of Victoria and NSW and Japan.

Thanks to my family, particular my father John, a terrific English and History teacher who always insisted “you can do whatever you set your mind to” and to “never give up”. Dad, I hope wherever you are you are proud of me for finally handing in after a very, very ‘interesting’ PhD experience. I am so very grateful to you for all your moral (and financial support) and encouragement and your comments on my papers. Thanks to my brother David for breaking up the monotony of writing with requests for computer help and property advice and to my sister Helen, her husband Adam and the twins Zach and Miles for their support. A big thanks to the twins for being great fun … for the soccer games I never seem to win (the benefit of making up the rules as you go), backyard footy matches and bike rides, plus a chance to practice my carpentry skills on a two-storey cubby house. Lastly, thank you to the other Pharos: Jill Pharo and Charlie, the Harveys, Zoe and the Littles for making the past 6 months easier!

Table of Contents

Abstract	iii
Declaration	vii
Preface	viii
Acknowledgments	x
Table of Contents	xiii
List of Figures	xvi
List of Tables	xxv
List of Abbreviations	xxvii
Chapter One - General Introduction	**1-44**
Introduction	3
The Evolution of Lactation	4
Reproductive Strategies	6
Monotremes	6
Marsupials	6
Eutherians	7
Development of the Mammary Gland	8
The Secretion of Milk	11
The Composition of Milk	11
Colostrum	13
Mature Milk	14
Whey Proteins	18
Early Lactation Protein	18
The Kunitz-BPTI (I2) Family of Protease Inhibitors	23
Whey Acidic Protein	27
α-Lactalbumin and Lysozyme	28
The Milk Lipocalins: β-Lactoglobulin and Late Lactation Proteins A and B.	28
Caseins	29
Other Milk Constituents	30
Milk Fat Globules	30
The Regulation of Milk Protein Gene Expression	31
Endocrine Regulation of Mammary Gland-Specific Milk Protein Gene Expression	31
Prolactin	32
Prolactin Receptor	32
Growth Hormone	36
Insulin	36
Glucocorticoids	36
Thyroid Hormones	36
Oxytocin	37
Transcriptional Regulation of Mammary Gland-Specific Milk Protein Gene Expression	38
Signal Transducers and Activators of Transcription	38
Glucocorticoid Receptor (GR) and the Glucocorticoid Response Element (GRE)	39

CCAAT Box-Binding Transcription Factor/Nuclear Factor I (CTF/NFI) 39
Mammary Cell-Activating Factor (MAF) 39
CCAAT/Enhancer Binding Proteins (C/EBPs) 40
Other Transcription Factors 41
CoREs (Composite Response Elements) 41
Extracellular Matrix and Milk Protein Gene Expression – Lessons from in vitro Culture Systems 42
Mammary Gland Explant Culture 43
Primary Cells and Immortalised Cell Lines 44
Scope of the Thesis 46

Chapter Two - General Materials and Methods 49-76
Introduction 51
Animals 55
Tammar Wallabies 55
Cape and Subantarctic Fur Seal Samples 55
Miscellaneous Samples 56
Ethics Protocols 56
Histology 57
Milk Collection 58
Molecular Biology 59
DNA Oligonucleotide Design 59
Isolation of Genomic DNA 59
Reverse Transcriptase-PCR 60
Electrophoresis of DNA 60
Gel Extraction 60
Ligations 61
Electroporation or Transformation of DNA into E. coli 61
Preparation of Plasmid DNA 62
Restriction Enzyme Digestion 64
Sequencing of Plasmid DNA 64
Isolation of Total RNA from Tissue, Milk and Cells 64
Electrophoresis of Total RNA 66
Northern Blotting 67
[α^{-32}P] dCTP Labelling of cDNA Probes 67
Membrane Hybridisation 67
Quantitation of mRNA Expression 68
Microarray Analysis of the Tammar Wallaby Mammary Gland 68
Labelling of Microarray Probes 71
Hybridisation of Microarray Slides 71
Analysis of Microarray Slides 72
Protein 73
Extraction of Whey from Whole Milk 73
Determination of Protein Concentration 73
1D Gel Electrophoresis of Proteins (SDS-PAGE) 73
Staining of SDS-PAGE Gels 74
Coomassie Staining 74
Silver Staining 74
Cell Culture 74
Recovery of Cells from Liquid N_2 Storage 74

Transfection of Primary Cells Isolated from the Tammar Wallaby Mammary Gland and the HC11, HEK293T and KIM-2 Immortalised Cell Lines 75
Primary Cells 75
Immortalised Cells Lines 75
In silico Analyses 75
Bioinformatics 75
Statistics 76

Chapter Three - The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene **77-112**
Abstract 79
Background 79
Results 79
Conclusions 79
Background 79
Results 81
ELP/CTI evolved from a common ancestral gene 81
Selection pressure acting upon ELP and eutherian CTI 83
Marsupial ELP and eutherian CTI share common flanking genes 84
CTI has been lost in some eutherians 84
Transposable elements within the ELP/CTI gene 85
Bovine CTI, PTI, STI and the TKDPs share a common ancestral gene 85
Tammar ELP expression is up-regulated at parturition and is mammary-specific 86
Discussion 87
Mammalian ELP/CTI and the evolution of bovine PTI, STI and the TKDPs 87
Lack of conservation of the ELP/CTI putative P_1 reactive site residue 88
ELP/CTI – a conserved N-glycosylation site predicted within the Kunitz domain 89
Conclusions 91
Methods 91
Animals 91
Tissues 92
RNA extraction and northern analysis 92
RT-PCR and cloning of ELP/CTI 92
Genomic DNA isolation and cloning 92
Isolation of the tammar ELP gene from a genomic library 92
Fluorescence in situ hybridisation (FISH) 92
cDNA microarray analysis of tammar ELP gene expression 93
Sequence analysis 94
dN/dS analysis 95
Phylogenetic analysis 95
Additional files 95
Abbreviations 97
Competing interests 97
Authors' Contributions 97

Acknowledgements 97
Author details 97
Authors' note 97
References 97
Additional files 1-12 100-111

Chapter Four – A *colostrum trypsin inhibitor* gene expressed in the Cape fur seal mammary gland during lactation **113-142**
Abstract 115
Abbreviations 115
Introduction 115
Materials and methods 116
Animals 116
Isolation of Cape fur seal genomic DNA and cloning of the CTI gene 116
In silico identification of CTI 116
Preparation of total RNA 117
RT-PCR cloning of the fur seal transcript 117
Northern blot analyses 118
Sequence analysis 118
Structural homology modelling 118
Results 118
The Cape fur seal and Pacific walrus have a putative protein-coding CTI gene, but Weddell seal CTI encodes a pseudogene 118
Cape and subantarctic fur seal CTI are expressed in the mammary gland for at least 2-3 months postpartum 118
Pinniped CTI, putative N-glycosylated Kunitz serine protease inhibitors 118
Structural homology modelling of Cape fur seal CTI suggests that it inhibits trypsin 120
Discussion 120
Conservation of pinniped CTI gene structure 121
The unique expression pattern of Cape fur seal CTI 121
Cape fur seal CTI – a putative protease inhibitor 122
Putative CTI functions in the mammary gland and neonate 122
Potential role of oligosaccharides in CTI function 123
Conclusions 123
Acknowledgements 123
References 123
Appendix A. Supplementary Data: Supplementary files 1-13 125-141

Chapter Five – Expression of the mammary gland-specific tammar wallaby *Early Lactation Protein* gene is maintained *in vitro* in the absence of prolactin **143-158**
Abstract 145
Abbreviations 145
Introduction 145
Materials and methods 147
Explants experiment design 147
Animals 147

Preparation and culture of tammar mammary gland explants 147
RNA extraction and Northern blot analysis 147
Statistical analyses 147
Results 148
Endocrine regulation of the tammar *ELP* gene in mammary gland explants 148
Phase 1 (pregnancy) 148
Phase 2A (early lactation) 148
Phase 2B (mid-lactation) 149
Discussion 150
Conclusions 152
Acknowledgements 152
References 152
Appendix A. Supplementary Material: Supplementary files 1-13 155-161

Chapter Six – Mammary cell-activating factor regulates the hormone-independent transcription of the *Early Lactation Protein* (*ELP*) gene in a marsupial **163-235**
Abstract 167
Keywords 168
Abbreviations 168
Introduction 169
Materials and methods 172
Animals 172
Primer Extension analysis of tammar ELP 172
Creation of tammar ELP and murine CSN2 d2EGFP and luciferase promoter reporter constructs 172
Other plasmids used for transfection experiments 175
Cell Culture, stable and transient transfections 175
HEK293T Cell Culture 175
Creation of prolactin-responsive HEK293T cells 175
Minimal tammar ELP promoter required for transcription in HEK293T cells 176
Reporter assays: analysis of transgene expression 176
Detection of d2EGFP 176
Luciferase, β-Galactosidase and dual luciferase assays 176
Statistical analyses 177
Cloning and in silico identification of marsupial ELP and eutherian CTI promoters 177
Prediction of putative ELP and CTI transcription factor binding sites 177
Putative TFBS within genes with a similar expression profile to tammar ELP 177
Results 178
Tammar ELP has two transcription start sites 178
Identification of a putative milk protein gene-like composite response element (CoRE) within the tammar ELP promoter 179
Tammar ELP is unresponsive to lactogenic hormones in 2D cell culture model 183
Tammar ELP is not responsive to PRL alone 186

Identification of the minimal tammar ELP promoter required for hormone-independent gene transcription 186
Identification of highly-conserved putative MAF/ETS, AP2γ and overlapping C/EBP and NF-KB sites within the proximal ELP and CTI promoters 187
Conserved distal and intragenic elements in the ELP and CTI promoters 192
Conserved TFBS in Phase 2A-specific tammar gene promoters 195
Discussion 195
Activation and repression of the tammar ELP gene during the reproductive cycle 195
The hormone-independent activation of tammar ELP 200
Indirect activation of tammar ELP 201
The importance of extracellular matrix and a 3D tissue structure 201
Conclusions 201
References 203
Acknowledgements 211
Supplementary Material: Supplementary files 1-16 213-242

Chapter Seven – Developing recombinant protein expression of tammar ELP **243-284**
Introduction 245
Materials and methods 246
Mammalian expression of marsupial ELP and eutherian CTI 246
Preparation of ELP and CTI mammalian expression constructs 246
Human Embryonic Kidney (HEK293T) cell culture 247
Production of ELP and CTI in conditioned media 247
Preparation of HEK293T cell lysates 248
Sulforhodamine B colorimetric assay: measurement of cell biomass production 248
Bacterial expression of tammar ELP 249
Preparation of bacterial expression constructs to produce recombinant tammar ELP 249
Optimisation of MBP-tELP fusion protein production 255
SDS-PAGE, Western blotting and immunodetection of MBP-tELP 255
Purification of recombinant tELP from its cleaved MBP fusion-partner by gel filtration 256
Purification of insoluble MBP-tELP from E. coli inclusion bodies 257
Results 258
Evaluation of the pCMVSport6 and pcDNA3.1 mammalian expression vectors for ELP and CTI production 258
Overexpression of recombinant tammar ELP in *E. coli* cells with the pMAL system 262
Optimisation MBP-tELP fusion protein production 262
Expression of MBP-tELP in the ompT protease-deficient E. coli K12 UT5600 strain – reduced proteolytic degradation 262
Improved solubility of MBP-tELP in the E. coli cytoplasm 265
MBP-tELP amylose resin affinity binding - ~40% of MBP-tELP is lost 265

Factor Xa cleavage of MBP-tELP is optimal but incomplete with added detergent 272
Scaled-up production of MBP-tELP - amylose resin affinity purification 272
Gel purification of recombinant tammar ELP 275
Production and isolation of the MBP-tELP fusion protein from inclusion bodies 277
Discussion 278
Production of tammar CST3, but not ELP and CTI in mammalian expression systems 278
Successful production of the MBP-tELP and MBP-tCST3 fusion proteins 279
Replacement of rare E. coli codons in the protein of interest 280
Improvement of MBP fusion protein solubility 280
Disulphide bond formation in *E. coli* 281
Isolation of MBP-tELP from insoluble inclusion bodies using denaturation and refolding reduces E. coli protein impurities 282
Improving Factor Xa enzymatic cleavage of MBP-tELP 282
Production of N-glycosylated tammar ELP 283
To tag or not to tag? 283
Conclusions 284

Chapter Eight – General discussion and future directions 285-306
Discussion 287
The Evolution of the *Early Lactation Protein* gene 287
The regulation of tammar *ELP* 290
Marsupial ELP and eutherian CTI: putative protease inhibitors 297
The Evolution of Lactation 300
General Conclusions 305

Bibliography 307-374

Appendices 375-455
Appendix I: Animal List 376
Appendix I: Buffers and Solutions 377-386
Protein Solutions 377
RNA Solutions 379
DNA Solutions 381
Tissue and Cell Culture Solutions 383
Appendix III: Preparation of competent cells 387
Chemically competent cells 387
Electrocompetent cells 387
Appendix IV: Sequencing Reaction Setup, PCR Conditions and Clean-Up and Primers 388-392
Table 1. Sequencing reaction set-up for sequencing of plasmid DNA by the Department of Pathology, University of Melbourne 388
Table 2. PCR conditions for sequencing plasmid DNA 388
Sequencing Reaction Clean-Up Protocol 388
Sequencing Primers 389

Table 3. Primers used to sequence the tammar wallaby (*Macropus eugenii*) *ELP* distal promoter 389
Table 4. Primers used to sequence the Victorian Koala (*Phascolarctos cinereus victor*) *ELP* gene 389
Table 5. Primers used to sequence the Brushtail possum (*Trichosurus vulpecula*) *ELP* gen 390
Table 6. Primers used to sequence the Fat-tailed dunnart (*Sminthopsis crassicaudata*) *ELP* gene 390
Table 7. Primers used to sequence the Cape fur seal (*Arctocephalus pusillus pusillus*) *CTI* gene 391
Table 8. Primers used to sequence the Victorian Koala (*Phascolarctos cinereus victor*) *ELP* promoter 391
Table 9. Primers used to sequence the Fat-tailed dunnart (*Sminthopsis crassicaudata*) *ELP* promoter 391
Table 10. Primers used to sequence the Cape fur seal (*Arctocephalus pusillus pusillus*) *CTI* promoter 392
Table 11. Primers used to sequence the mammalian *ELP* and *CTI* and bacterial *ELP* and *CST3* expression constructs 392

Appendix V: Optimisation of MBP-tELP fusion protein production **393-395**

Appendix VI: Northern Analysis of the tammar *ELP* and *CST3* and bovine *CTI* transcripts produced by mammalian expression vectors transfected into HEK293Ts **396**

Appendix VII: Other Publications arising from this thesis **397-452**

Characterization of two whey protein genes in the Australian dasyurid marsupial, the stripe-faced dunnart (*Sminthopsis macroura*) 399-406
Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development 407-454

List of Figures

Chapter One

Figure 1.1	The evolution of the extant mammals, birds and squamates	5
Figure 1.2	The five transport pathways via which constituents enter milk	12
Figure 1.3	Changes in milk composition and the growth and development of the pouch young during the tammar wallaby lactation cycle	16
Figure 1.4	Expression of selected tammar milk proteins genes and enzymes throughout the lactation cycle	19
Figure 1.5	Alignment of the mature tammar ELP peptide with selected Kunitz protease inhibitors	20
Figure 1.6	Ribbon representation of the 3D structure of BPTI	23
Figure 1.7	The 'standard' (or 'Laskowski') reaction mechanism characteristic to the Kunitz inhibitor family	24
Figure 1.8	The two-step mechanism of serine protease inhibition	25
Figure 1.9	Structure of the BPTI-trypsin complex	26
Figure 1.10	Prolactin receptor signalling in the mammary gland	34

Chapter Two

Figure 2.1	The annual reproductive cycle of the tammar wallaby, *Macropus eugenii*	52
Figure 2.2	Tammar wallaby mammary gland explant culture system	57
Figure 2.3	Collection of milk from the mammary gland of a tammar wallaby	58
Figure 2.4	Measurement of tammar PY head length with calipers	58
Figure 2.5	cDNA microarray analysis of the tammar wallaby mammary gland from the virgin gland (V), through to pregnancy (P), lactation (L) and involution (I)	69
Figure 2.6	Creation of the commercially prepared tammar wallaby mammary gland cDNA library in the pCMV-SPORT6.ccdb mammalian expression vector (Invitrogen).	70

Chapter Three

Figure 1	Structure of the marsupial *ELP* and eutherian *CTI* genes	82
Figure 2	Alignment of the marsupial ELP and eutherian CTI precursor proteins	83
Figure 3	Localisation of the tammar ELP gene to Macropus eugenii chromosome 1q using FISH	85
Figure 4	Chromosomal location of the *ELP*/*CTI* gene in different species	86
Figure 5	VISTA plot of pairwise alignments for selected mammals in the region containing the *PIGT*, *ELP*/*CTI* and *WFDC2* genes	87
Figure 6	A phylogenetic tree of *ELP*/*CTI* and the *CTI*-like bovine *PTI*, *STI* and *TKDP1*, *2*, *3*, *4* and *5* family	88
Figure 7	*ELP* expression in the tammar mammary gland throughout the reproductive cycle	89
Figure 8	Tammar *ELP* expression is specific to the mammary gland	90
Figure 9	Evolution of the *ELP*/*CTI* gene in therians	90
Additional Figures		
Figure S1	Alignment of the marsupial *ELP* and eutherian *CTI* transcripts	101

Figure S2 Exon 1 and 2 mutations within selected putative eutherian *CTI* pseudogenes 104
Figure S3 Transposable elements and simple repeats located within the *PIGT* and *ELP*/*CTI* genes and flanking regions 105
Figure S4 Alignment of the bovine CTI, PTI, STI, TKDP1-5 and SPINT4 precursor proteins 106
Figure S5 Relationship between bovine CTI, PTI, STI, TKDP1-5 and SPINT4 107
Figure S6 Genomic arrangement and mVISTA plot of pairwise alignments for the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes 108
Figure S7 Transposable elements located within the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes 109

Chapter Four

Figure 1 Evolutionary history of the Pinnipedia 117
Figure 2 Structure of pinniped *CTI* and other eutherian *CTI* genes 119
Figure 3 Expression of *CTI* in the fur seal mammary gland 119
Figure 4 Alignment of CTI, BPTI and the second Kunitz domain of human TFPI 120
Figure 5 Panel showing salient features of the modelled Cape fur seal CTI (green) and its complex with Pacific walrus trypsin (cyan) 121
Figure 6 Alignment of selected eutherian trypsinogen precursor proteins 122

Supplementary Figures

Supp. File 2 Weddell seal *CTI* pseudogene - putative transcript and protein 126
Supp. File 5 Distribution of the eutherian *CTI* and marsupial *ELP* gene in mammals 129
Supp. File 6 Alignment of eutherian *CTI* and marsupial *ELP* mRNA transcripts 131
Supp. File 7 Phylogeny of the mature protein-coding *CTI* and *ELP* transcripts 133
Supp. File 8 Northern analysis of Cape and subantarctic fur seal mammary gland tissue probed with [α-32P] dCTP-labelled fur seal CTI cDNA 136
Supp. File 11 ClustalW2 alignment of eutherian CTI and marsupial ELP precursor proteins 139

Chapter Five

Figure 1 Expression of the major milk protein genes and development of the pouch young during the reproductive cycle of the tammar wallaby 146
Figure 2 Endocrine regulation of *ELP* and *LGB* in mammary explants cultured from late-pregnancy (Phase 1) 148
Figure 3 Endocrine regulation of *ELP*, *WAP* and *LGB* in mammary explants cultured from late-pregnancy (Phase 1) 149
Figure 4 Endocrine regulation of *ELP*, *WAP* and *LGB* in explants cultured from early lactation (Phase 2A) mammary tissue after down-regulation with insulin *in vitro* 149
Figure 5 Endocrine regulation of *ELP*, *WAP* and *LGB* in explants from mid-lactation (Phase 2B) mammary tissue after down-regulation with insulin *in vitro* 150

Supplementary Figures

Supp. File 1 Phase 1 Phase 1 (day 24 of pregnancy) explant experiment 155
Supp. File 2 Phase 1 (day 24 of pregnancy) explant experiment 156
Supp. File 3 Phase 2A (day 80 pp, early lactation) explant experiment 157
Supp. File 4 Endocrine regulation of *ELP* and *WAP* in mammary explants cultured from early lactation (Phase 2A) 158
Supp. File 5 Endocrine regulation of *ELP*, *WAP* and *LGB* in mammary explants cultured from mid-lactation (Phase 2B) 159
Supp. File 6 Phase 2B explant experiment 160
Supp. File 7 Phase 2B explant experiment 161

Chapter Six

Figure 1 Identification of two transcription start sites in the tammar *ELP* promoter 180
Figure 2 Putative transcription factor binding sites within the tammar wallaby *ELP* promoter and *ELP* promoter constructs 184
Figure 3 Luciferase activity of tammar *ELP* promoter constructs in prolactin-responsive HEK293T cells 188
Figure 4 Hormone-independent activity of tammar *ELP* promoter constructs in HEK293T cells 190
Figure 5 Conserved transcription factor binding sites within the marsupial *ELP* and eutherian *CTI* proximal promoters 191
Figure 6 Highly conserved putative *ELP*- and *CTI*-specific proximal promoter TFBS 193
Figure 7 VISTA plot of pairwise alignments for selected marsupial *ELP* and eutherian *CTI* genes and flanking regions 194
Figure 8 Putative regulatory mechanisms involved in the expression of tammar *ELP* 198

Supplementary Figures

Supp. File 1 Map of pBeloBACII-14.7kbtELP 213
Supp. File 2 Cell Culture Methods: HC11, KIM-2 and MDA-MB-231 immortalised cell lines and Tammar mammary gland primary cells 214
Supp. File 3 Maps of the pd2EGFP and pGL3-Basic tammar *ELP* promoter constructs 219
Supp. File 6 Identification of the transcription start sites of the tammar *ELP* gene 222
Supp. File 8 Prolactin-responsiveness of HC11 and KIM-2 murine mammary epithelial cells in an *in vitro* culture system 229
Supp. File 9 Activity of tammar *ELP* promoter constructs in KIM-2 cells 230
Supp. File 10 Growth and differentiation of tammar primary mammary gland cells (PMGCs) on tissue culture plastic 231
Supp. File 11 Lactogenic hormone-responsiveness of tammar primary mammary gland cells (PMGCs) 232

Supp. File 12 Hormone-independent activity of tammar *ELP* promoter constructs in MDA MB 231 cells 233
Supp. File 13 Alignment of marsupial *ELP* proximal promoters 234
Supp. File 14 Alignment of eutherian *CTI* proximal promoters 235
Supp. File 15 Conserved transcription factor binding sites within 6 marsupial *ELP* and 20 eutherian *CTI* proximal promoters 239

Chapter Seven

Figure 7.1 Production and purification of recombinant tammar ELP using the pMAL Protein Fusion and Purification System 250
Figure 7.2 Design of ELP and cystatin C constructs for expression of the MBP-tELP and MBP-tCST3 fusion proteins in the *E. coli* cytoplasm (pMAL-c2X) and periplasm (pMAL-p2X) 252
Figure 7.3 Optimization of recombinant fusion protein production with the pMAL Protein Fusion and Purification System 254
Figure 7.4 Production of tammar ELP and Cystatin C in conditioned media 259
Figure 7.5 Production of tammar and opossum ELP, bovine CTI and Cystatin C in conditioned media 260
Figure 7.6 Western analysis of MBP-tELP fusion protein production in the *E. coli* K12 UT5600 *ompT* protease-deficient bacterial strain 263
Figure 7.7 Western analysis of MBP-tCST3 protein production in the *E. coli* K12 UT5600 cells 264
Figure 7.8 Western analysis of MBP-tELP yield and solubility over an increasing induction period 266
Figure 7.9 Western analysis of the effect of an increasing induction period on MBP-tCST3 yield and solubility 267
Figure 7.10 Western analysis of the effect of induction time and lysosyme and detergent treatment on the yield and solubility of MBP-tELP 268
Figure 7.11 The effect of induction time and lysosyme and detergent treatment on the yield and solubility of MBP-tCST3 270
Figure 7.12 Western analysis of the effect of detergent on the amylose resin binding affinity of MBP-tELP 271
Figure 7.13 Optimisation of Factor Xa cleavage of MBP-tELP 273
Figure 7.14 Optimisation of Factor Xa cleavage of the denatured and refolded MBP-tELP fusion protein 274
Figure 7.15 Amylose resin affinity-purification of MBP-tELP 275
Figure 7.16 Purification of tELP by gel filtration 276
Figure 7.17 Production of MBP-tELP in *E. coli* inclusion bodies 277

Chapter Eight

Figure 8.1 Similarities between products of the secretory epithelium (mucous) and mammary epithelium (colostrum and milk) 302

List of Tables

Chapter One

Table 1.1 Hormones involved in the development and regulation of the female mammary glands during pregnancy, lactation and involution 10

Table 1.2 Milk composition of selected marsupials, eutherians and monotremes 15

Chapter Two

Table 2.1 Stages of post-natal development in the tammar wallaby PY 54

Chapter Three

Table 1 Homology between and within the marsupial ELP and eutherian CTI Peptides 79

Table 2 Average rates of synonymous (dS) and non-synonymous (dN) substitutions occurring in marsupial ELP and eutherian CTI 80

Table 3 Primer sequences and conditions used to amplify *ELP*/*CTI* genes and transcripts 89

Additional Tables

Table S1 Characterisation of the putative functional *ELP*/*CTI* gene, transcript and protein 100

Table S2 Percentage similarity between and within the marsupial *ELP* and eutherian *CTI* transcripts 102

Table S3 Percentage similarity between and within the marsupial ELP and eutherian CTI peptides 103

Table S4 Location, identity and orientation of transposable elements within the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes 110

Table S5 Tammar wallaby mammary gland cDNA microarray data presented in Figure 7 111

Chapter Four

Table 1 Reproductive stages of Cape and subantarctic fur seals used for this study 117

Additional Tables

Supp. File 1 Nomenclature of species listed in this study 125

Supp. File 3 Accession numbers of sequences used in this study 127

Supp. File 4 Structure of the eutherian *CTI*, marsupial *ELP* and bovine *PTI* genes 128

Supp. File 9 Comparison of the putative physical characteristics of the mature BPTI, CTI and ELP peptides 137

Supp. File 10 Percentage similarity and identity between the mature eutherian CTI, marsupial ELP and BPTI peptides 138

Supp. File 12 Comparison of the predicted motifs within the mature BPTI, CTI and ELP peptides 140

Supp. File 13 Similarity and identity and between the active trypsin peptides of selected pinnipeds and other eutherians 141

Chapter Five

Table 1	Summary of tammar mammary gland explant experiments used to investigate hormone(s) required for maintenance and induction of the *ELP*, *WAP* and *LGB* genes at different reproductive stages	147
Table 2	Hormone concentrations used for tammar mammary gland explant culture	148

Chapter Six

Table 1	Primers and conditions used to create the tammar *ELP* and murine *CSN2* promoter constructs	174
Table 2	Genes with a similar expression profile to ELP as identified by cluster analysis of microarray expression data of the tammar wallaby mammary gland during the reproductive cycle	192
Additional Tables		
Supp. File 4	Accession numbers for sequences used in this study	220
Supp. File 5	Primers and conditions used to amplify selected marsupial *ELP* and eutherian *CTI* promoter regions and to screen HC11 cells for the stable integration of the 14.7kb tammar *ELP* gene	221
Supp. File 7	The location of putative transcription factor consensus binding sites within the tammar wallaby *ELP* gene	223

Chapter Seven

Table 7.1	Primers and conditions used to amplify the mature protein coding region and 3' UTR of tammar *ELP* and *CST3* for expression in bacterial cells	251

Chapter Eight

Table 8.1	Summary of marsupial *ELP* and eutherian *CTI* gene structure	290
Table 8.2	Summary of *ELP* response to hormones *in vitro* in the 3D mammary explant culture system	291

List of Abbreviations

°C	Degrees Celcius
~	Approximately
1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
3'	3 prime end of DNA/RNA
3C	Chromosome Conformation Capture
3D	Three-dimensional
4C	Circularised Chromosome Conformation Capture
4-DSC	Four disulphide core
5'	5 prime end of DNA/RNA
5C	Carbon Copy Chromosome Conformation Capture
5-HT	5-hydroxytryptamine (serotonin)
6C	Genome Conformation Capture (GCC) with ChIP
$A_{540\ nm}$	Absorbance at a specified wavelength (nm)
aa	Amino acid
α-CSN	α (alpha)-casein
AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride
$AgNO_3$	Silver nitrate
AGRF	Australian Genome Research Facility
α-LA	α (alpha)-lactalbumin
Ala	Alanine
AMBP	α_1-microglobulin-bikunin
AMP	Ampicillin
ANOVA	Analysis Of Variance
AP	activating protein
A. p. p	*Arctocephalus pusillus pusillus*
APP	amyloid (βA4) precursor protein
APS	Ammonium persulfate
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
A. tropicalis	*Arctocephalus tropicalis*
BGAL	*β-galactosidase*
B3GALT1	*β-1,4-galactosyltransferase*
B4GALT1	*β-1,3-galactosyltransferase*
BAC	Bacterial Artificial Chromosome
BCE1	bovine casein enhancer 1
β-CSN	β-casein
β-LG	β (beta)-lactoglobulin
BLASTn	Nucleotide-nucleotide BLAST
BLASTx	Translated query vs. protein database BLAST
BM	Basement membrane
β-ME	β-Mercaptoethanol

B. taurus	*Bos taurus*
bp	base pairs
BSA	bovine serum albumin
Ca	Calcium
$CaCl_2$	Calcium chloride
CAMP	cathelicidin antimicrobial peptide
CAT	chloramphenicol acetyltransferase
C. familiaris	*Canus familiaris*
CD	Circular dichroism
cDNA	Complementary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
CK2	Casein kinase II
CLD	Cytoplasmic lipid droplet
cm	Centimetre
CMV	Cytomegalovirus
CO_2	Carbon dioxide
Co-IP	Co-immunoprecipitation
Conc.	Concentration
COOH	Carboxy terminus of the protein
Cot-1	Blocking DNA (from the human placenta)
CSN1	*α-casein*
CSN2	*β-casein*
CSN3	*K(kappa)-casein*
CST3	*Cystatin C*
CREB	cAMP response element binding protein
CTI	Colostrum Trypsin Inhibitor
CTF	CCAAT box transcription factor
Cys	Cysteine
d	Day
D	Dexamethasone
d2EGFP	destabilised enhanced green fluorescent protein
ds	Double stranded
Da	Daltons
DamID	DNA adenine methyltransferase identification
dCTP	Deoxycytidine 5'-triphosphate
dd H_2O	Double distilled/deionised H_2O
DEPC	Diethyl pyrocarbonate
DHS	DNase I hypersensitivity site
DRE	Distal regulatory element
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
dNTP	Deoxynucleotide triphosphate
DPM	Disintegrations per minute
ds	Double stranded

DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)
DTT	Dithiothreitol
E	East
E_2	Estrogen
E-64	N-(trans-Epoxysuccinyl)-L-leucine 4-guanidinobutylamide
E. coli	Escherichia coli
ECL	Enhanced chemoluminesence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetra acetic acid
EGF	Epidermal Growth Factor
EGFP	Enhanced Green Fluorescent Protein
EGFR	Epidermal Growth Factor Receptor
EHS	Engelbreth-Holm-Swarm (tumor virus)
ELF	E47-like facto
ELP	Early Lactation Protein
EMSA	Electrophoretic mobility shift assay
ER	Estrogen Receptor
ERE	Estrogen response element
ERRα/β	strogen-related receptor alpha/beta
EST	xpressed sequence ta
EtBr	Ethidium bromide
EtOH	ethanol
ETS	26 transformation-specific
EVI1	Ecotopic virus integration site 1
FATT	Flag-Acidic-Target Tag
FBS/FCS	Fetal Bovine Serum/fetal calf serum
F. catus	*Felis catus*
FIL	Feedback Inhibitor of Lactation
FISH	fluorescent *in situ* hybridisation
g	Gram/s
g	Gravitational force
GAS	Gamma interferon activated sequence
GCC	Genome Conformation Capture
GFP	Green Fluorescent Protein
GH	Growth Hormone
GHR	Growth Hormone Receptor
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GR	Glucocorticoid receptor
½GRE	Glucocorticoid response element
GST	Glutathione S-transferase
Gu-HCl	Guanidine hydrochloride
gDNA	Genomic deoxyribonucleic acid
h/hr	hour
H_2O	Water

H3Ac	Histone H3 acetylation
H3K4me2	Hhistone H3 lysine 4 di-methylation
HBSS (Hanks)	Hank's Balanced Salt Solution
HC	Hydrocortisone
HCl	Hydrochloric acid
HDAC	Histone deacetylase
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
His	Histidine
H. sapiens	*Homo Sapiens*
HL	head length
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
HTRA1	High temperature requirement A1 peptidase
I	Insulin
IB	Inclusion body
ID	Identification
IFNγ	Interferon gamma
Ig	Immunoglobulin
IGFBP	insulin-like growth factor binding protein
IGF	Insulin-like Growth Factor
IGFBP	Insulin-like growth factor binding protein
IGFIR	Insulin-like growth factor 1 receptor
IGF2R	Insulin-like growth factor 2 receptor
IgG	Immunoglobulin G
Ile	Isoleucine
Inr	Initiator element
IPTG	Isopropyl-thio-β-D-galactoside
IRE	Insulin response element
IRF	Interferon regulatory factor
JAK	Janus kinase
K	Potassium
IR	insulin receptor
kb	Kilo bases
KD	Kunitz domain
kDa	Kilo Daltons
kg	Kilogram/s
L	Litre/s
LALBA	*α-lactalbumin*
LB	Luria-Bertani broth
LGB	*β-lactoglobulin*
Leu	Leucine
LF2000	Lipofectamine 2000
Liquid N_2	Liquid nitrogen
LINE	Long interspersed elements
LLP	*Late lactation protein*
LSD	Least Significant Difference

LTR	Long terminal repeat
Lys	Lysine
M	Molar
mA	Milli Amps
MAF	Mammary cell activating facto
MBP	Maltose Binding Protein
M. domestica	*Monodelphis domestic*
MEC	Mammary epithelial cell
MER	Medium Reiterated frequency repeat
Met	Methionine
M. eugenii	*Macropus eugenii*
MeOH	Methanol
mg	Milligram/s
MFG	Milk fat globule
MFGM	Milk fat globule membrane
MG	Mammary gland
MGF	Mammary Gland Factor
$MgSO_4$	Magnesium sulphate
min	Minute/s
miRNA	Micro RNA
mL	Millilitre/s
MLD	Micro-lipid droplet
mm	Millimetre
mM	Milli Molar
MMTV	Mouse mammary tumour virus
MQ H_2O	Milli-Q (ultrapure) water
mRNA	messenger ribonucleic acid
MQ H_2O	Milli-Q purified water
MW	Molecular Weight
MWCO	Molecular Weight Cut-Off
MYA/Myr	Million years ago
MZF	Myeloid zinc finger
Na	Sodium
NaCl	Sodium chloride
$NaHCO_3$	Sodium bicarbonate
$Na_2H_2CO_3$	Sodium bicarbonate
NaH_2PO_4	Sodium orthophosphate
NaOAc	Sodium acetate
NaOH	sodium hydroxide
NaP	Sodium phosphate
NBP	Nuclear Binding Proteins
NBF	Neutral Buffered Formalin
NCBI	National Centre for Biotechnology Information
ncRNA	non-coding RNA
NEB	New England Biolabs
NEM	N-Ethylmaleimide

NFAT	Nuclear activator of T-cells
NFI	Nuclear factor I
NF-IL6	Nuclear factor-interleukin 6
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cell
NFY	Nuclear factor Y
NH_2	Amino terminus of protein
ng	Nano gram
nm	Nano metre
NMR	Nuclear magnetic resonance
NMWL	Nominal molecular weight limit
NP-40	Nonidet P-4
NS	Non-sucked mammary gland
nt	Nucleotide
OCT	Octamer
OD	Optical density
ORF	Open Reading Frame
P	Pregnancy
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate Buffered Saline
P. cinereus	*Phascolarctos cinereus*
PCD	Programmed Cell Death
PCR	Polymerase Chain Reaction
PE	pouch empty
Phe	Phenylalanine
pI	Porcine Insulin/Isoelectric point
PIGT	Phosphatidyl inositol glycan, class T;
PKC	Protein kinase C
PL	Placental lactogen
PMF	Pregnancy-specific mammary nuclear factor
PMGC	Primary mammary gland cells
PMSF	Phenylmethylsulfonyl fluoride
PNGase F	Peptide-*N*-Glycosidase F
pp	Post partum
PRL	Prolactin
PRLR	Prolactin Receptor
$PRLR_S$	Prolactin Receptor short form
$PRLR_L$	Prolactin Receptor long form
$PRLR_I$	Prolactin Receptor intermediate form
Prog	Progesterone
PRSS1	Protease, serine, 1 (trypsin 1)
PS	Prosite
PY	Pouch young
RIPA	Radio immuno precipitation assay buffer
RNA	Ribonucleic acid
RNAi	RNA interference
RPY	Removal of pouch young

RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
s	Second
S	South
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
Ser	Serine
SINE	Short interspersed elements
siRNA	Short interfering RNA
SLPI	Secretory leukocyte protease inhibitor
S. macroura	*Sminthopsis macroura*
SP	Specificity protein
SPINLW1	Serine peptidase inhibitor-like with Kunitz and WAP domains 1
SPINT	Serine protease inhibitor Kunitz type
SSC	Sodium chloride sodium citric acid
S. scrofa	*Sus macroura*
STAT	Signal transducer and activator of transcription
T_3	Tri-iodo thyronine
tammar	tammar wallaby, *Macropus eugenii*
TAE	Tris acetic acid EDTA buffer
TBE	Tris boric acid EDTA buffer
TBP	TATA binding protein
TBS	Tris Buffered Saline
TBST	Tris buffered saline with Tween-20
TE buffer	Tris-EDTA buffer
TEMED	N, N, N', N'-tetramethlyethylenediamine
temp	Temperature
TF	Transcription factor
Tf	Transfected
TFBS	Transcription factor binding site
TFPI	Tissue factor pathway inhibitor
TGF	Transforming Growth Factor
Thr	Threonine
Triton-100X	octylphenol ethylene oxide condensate
tRNA	Transfer RNA
Trp	Tryptophan
TSS	transcription start site
T. truncatus	*Tursiops truncatus*
T. vulpecula	*Trichosurus vulpecula*
Tween-20	polyoxyethylene (20) sorbitan monolaurate
Tyr	Tyrosine
µg	micro gram
U	Units
UTR	Untranslated region
UV	Ultraviolet light

Val	Valine
V	Volts
V/cm	Volts per centimetre
vol	Volume
v/v	Volume per volume
W	Watt
WAP	Whey Acidic Protein
WFDC	WAP four disulphide core
WFIKKN	WAP, follistatin/kazal, immunoglobulin, Kunitz and netrin domain containing protein
WGS	Whole genome sequence
WPPCS	Whey protein promoter core sequence
w/v	Weight per volume
x	times
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactoside
YY1	Yin-Yang 1.

Chapter One:

General Introduction

Introduction

Milk plays an essential role in the nourishment and immunological protection of the young, providing energy, amino acids, vitamins, minerals and immune factors to ensure the growth, development and survival of the newborn (Jenness, 1986; Ward and German, 2004; Watson, 1980). Milk is produced by the mammary gland, an exocrine organ unique to mammals (Blackburn, 1991; Peaker, 2002). The amounts and proportions of the major milk constituents (carbohydrates, fats and proteins) differ between species, with some components synthesised by the mammary gland, whilst others enter the gland via maternal serum (Jenness, 1986).

Lactation plays an important part in the reproductive success of mammals (Blackburn, 1993). It maximizes the survival of the young, providing a buffer against fluctuations in food quality and quantity and the toxicity of adult food (Langer, 2008; Pond, 1977). When resources are scarce, nursing females may use their own body reserves for milk production (Dall and Boyd, 2004; Pond, 1977). Each mammal has developed its own unique reproductive strategy to ensure the nutritional needs of the young are met in a wide variety of terrestrial and aquatic environments (Patton and Neville, 1997; Schulz and Bowen, 2005). Consequently, there is a great diversity in mammary gland anatomy, the nature and number of offspring produced and the duration of gestation and lactation (Hayssen *et al.*, 1985; Long, 1969; Weaver, 1997).

The use of a comparative biology and genomics approach, employing multiple species comparisons, provides a powerful method to better understand mechanisms involved in reproduction and lactation. Much research has focused on the use of eutherian animal models, particularly mice, rats and the dairy cow. Hence, the tammar wallaby (*Macropus eugenii*), an Australian marsupial, provides a novel biomedical model with which to investigate development, the regulation of lactation and the properties of milk constituents (Brennan *et al.*, 2007a; Hickford *et al.*, 2010; Nicholas *et al.*, 1997; Tyndale-Biscoe and Janssens, 1988). A major advantage of the tammar is that the altricial pouch young (PY) is readily accessible via the pouch (marsupium) and its neurological and physiological development and growth is driven entirely by its mother's milk (Nicholas *et al.*, 1997; Shaw and Renfree, 2006).

The major focus of this research was to characterise the tammar *Early Lactation Protein* (*ELP*) gene and to search for an *ELP* homologue/orthologue amongst other marsupials, eutherians and monotremes. Little is known about the function and regulation of *ELP*, yet it is one of the major whey proteins in early lactation tammar

milk (Joss *et al.*, 2009; Simpson *et al.*, 1998a). *ELP* is asynchronously expressed during the marsupial lactation cycle, but the regulation of the gene by endocrine and transcription factors is not well-understood (De Leo *et al.*, 2006; Demmer *et al.*, 1998; Piotte and Grigor, 1996; Simpson *et al.*, 1998a).

The Evolution of Lactation

The ancestral mammary gland most likely evolved ~310-225 million years ago (MYA) from an early therapsid or premammaliaform cynodont which was oviparous and exhibited lactation, like extant monotremes (Figure 1.1) (Blackburn *et al.*, 1989; Oftedal, 2002; Capuco and Akers, 2009; Oftedal, 2012; Oftedal and Dhouailly, 2013). These ancestral mammals did not develop a placenta, but produced parchment-shelled, porous eggs, through which skin or mammary secretions could be absorbed (Oftedal, 2002b; Robinson, 2008). Initial secretions were probably a dilute liquid, which protected the egg(s) against desiccation and microbial attack, with the subsequent addition of a nutritional role (Blackburn *et al.*, 1989; Goldman, 2002; Vorbach *et al.*, 2006). The hypothesis is supported by the presence of milk constituents common to lactation, inflammation and antimicrobial activity present in the mammary glands of extant mammals (Blackburn *et al.*, 1989; Vorbach *et al.*, 2006).

The evolution of the mammary glands and lactation is a contentious issue, with neither fossil evidence, nor a mammary gland homologue in current-day reptiles or birds (Hayssen and Blackburn, 1985; Long, 1969; McClellan *et al.*, 2008). The mammary gland may have evolved from cutaneous glands (Darwin, 1872), either from an ancestral apocrine-like sudiferous gland or from sebaceous glands (Blackburn *et al.*, 1989; Oftedal, 2002a), rather than the eccrine sweat glands (Blackburn, 1991). Alternatively, the mammary gland may have evolved as a neomorphic mosaic, the product of multiple gland types (Blackburn, 1991; Blackburn *et al.*, 1989; Oftedal and Dhouailly, 2013). Oftedal and Dhouailly (2013) proposed that the ancestral structure was an apocrine-pilo-sebaceous unit in the integument, which subsequently evolved into mammolobular-pilo-sebaceous units (MPSU). As the glands became more specialised, additional developmental components were incorporated (e.g., the mammary line, placode, bulb and primary sprout) leading to enlarged structures containing ductal trees and secretory tissue (Oftedal and Dhouailly, 2013). This hypothesis is based on the developmental appearance of MPSUs in monotremes, marsupials and some eutherians.

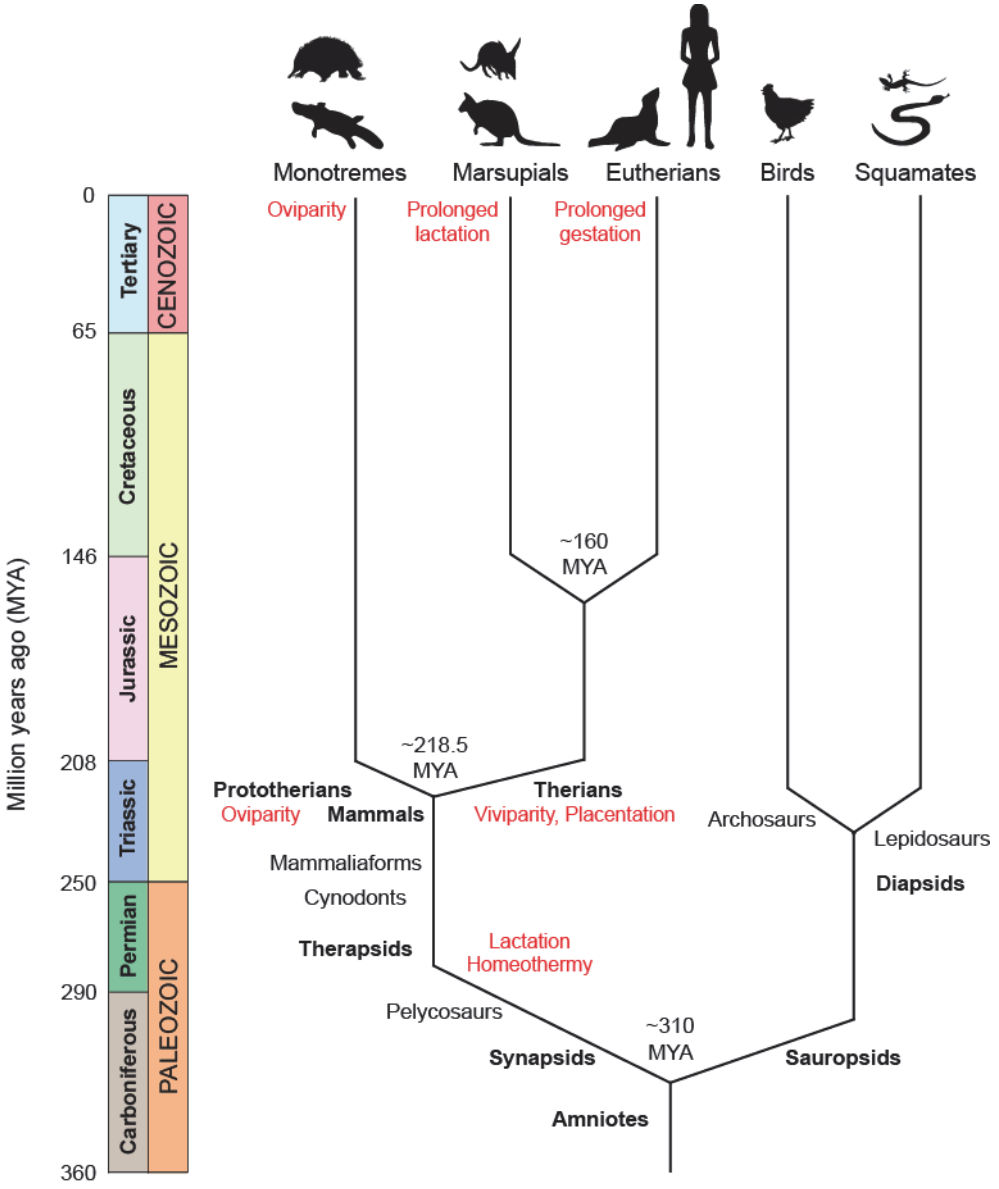


Figure 1.1. The evolution of the extant mammals, birds and squamates

Around 310 MYA, the synapsids and sauropsids diverged from amniotes. While the sauropsids gave rise to the diapsids; from which squamates (lizards and snakes) and birds arose, the synapsids evolved into therapsids and then primitive mammals. These ancestral mammals developed lactation and homeothermy, and ~218.5 MYA in the late Triassic period, gave rise to the ancestral prototherians and therians (ancestral mammals). The marsupials and eutherians then evolved ~182-160 MYA from the therian ancestor, giving rise to the present-day marsupials and eutherians. This figure was adapted from (Warren *et al.*, 2008), with evolutionary divergence times based upon those determined by (Bininda-Emonds *et al.*, 2007; Luo *et al.*, 2011; Yu *et al.*, 2012).

Ancestral therians and prototherians (ancestral monotremes) diverged ~218.5 MYA (Luo *et al.*, 2011; Yu *et al.*, 2012). While the protherians remained oviparous, therians became viviparous i.e., they gave birth to live young (Oftedal, 2002a). Marsupials and eutherians then diverged ~160-182 MYA (Luo *et al.*, 2011; Yu *et al.*, 2012). The more than 5,448 present-day mammalian species (IUCN, 2015) have developed very different reproductive strategies (Lamming, 1994) including a wide variety of placental types. Whilst marsupials provide sustenance to the fetus predominantly via a yolk sac placenta (Freyer *et al.*, 2003; Renfree, 1983), have evolved a complex array of placental structures including yolk sac and chorioallantoic placentas (Mossman, 1987; Wooding and Burton, 2008)

Reproductive Strategies

Monotremes

The five species of extant monotremes, including the platypus and short-beaked echidna lay eggs filled with yolk, from which the young hatch (Grant, 1984; Griffiths, 1968). Nutrients are transferred to the young via the porous egg shell and bilaminar yolk sac egg during the intrauterine period and possibly by mammary secretions during egg incubation (Griffiths, 1978; Oftedal, 2002b). The monotreme has neither nipples, nor teats but has two mammary glands, small 'milk patches', or 'areolae' located on either side of the abdomen (Griffiths, 1968, 1978). The platypus mammary glands have the typical alveolar structure of mammary glands in other taxa. Milk is secreted into the infundibula of specialised mammary hair follicles within the mammary areola, although from there it may "ooze" onto the skin (Griffiths, 1968, 1978). After hatching, the young actively sucks up milk, which is initially dilute, but then becomes more concentrated as lactation progresses (Griffiths, 1965; Oftedal, 2004).

Marsupials

The extant marsupials of today are found in the Americas and Australasia and are likely to have evolved in Gondwana, which encompassed South America, Antarctica and Australasia (Dickman, 2005; Woodburne *et al.*, 2003). The Australian marsupials in particular have had to adapt to an extremely harsh environment of unpredictable conditions, climate, unpalatable plants and low-fertility soils (Tyndale-Biscoe, 2005). This has dramatically influenced their reproductive strategy to ensure the survival of future generations. Hence, marsupials have a relatively long and complex lactation, but a short gestation (Tyndale-Biscoe, 2005), investing maternal resources in lactation, rather than pregnancy (Case, 1978; Hayssen *et al.*, 1985; Lee and Cockburn, 1985;

Renfree, 1982).

Many marsupials are seasonal breeders, with pouch exit and weaning of the young (the most vulnerable period of reproduction) synchronized with spring, the most plentiful period of food supply (Tyndale-Biscoe, 2005). In contrast, mating and birth may occur during mid-summer, the hottest and driest time of the year (Tyndale-Biscoe, 2005). After a short pregnancy, the duration of which is similar to the length of the oestrous cycle, the female gives birth to an altricial young (Tyndale-Biscoe and Janssens, 1988). Gestation length ranges from 10.7 days for the stripe-faced dunnart, the shortest period for any mammal (Selwood and Woolley, 1991), through 26.5 days after diapause for the monovular, polyoestrous tammar (Tyndale-Biscoe and Janssens, 1988) and up to 38 days for the long-nosed potoroo (Hughes, 1962). Some marsupials produce only one offspring, e.g. tammar, koala and kangaroo, whilst others such as the yellow-sided opossum, may produce up to 16 young in a litter (McAllan, 2003). Birth weights also vary, from less than 4 mg for the honey possum (Renfree, 1980) through 400 mg for the tammar wallaby (Renfree *et al.*, 1989; Tyndale-Biscoe and Renfree, 1987), and up to 830 mg for the red and eastern grey kangaroos (Sharman and Pilton, 1964).

Eutherians

Extant eutherians inhabit a wide range of terrestrial and aquatic environments worldwide and have therefore developed many different reproductive strategies to maximise the survival of their young (Derrickson, 1992; Fenelon *et al.*, 2014; Mead, 1993). Eutherians generally have a long pregnancy relative to lactation, during which maternal energy investment is high (Hayssen *et al.*, 1985). This was made possible by the development of an intricate, vascularised placenta (Enders, 2009; Wooding and Burton, 2008). This is true for monotocous eutherians: primates, giraffes, cows, goats, guinea pigs, camels and llamas, that give birth to a precocial young which is mature enough to survive in its environment (Nathanielsz, 1978; Tucker, 1985). These young have sensory functions soon after birth including eyes and ears open and may be furred. During early lactation they may ingest solid food, but remain with their mothers for long periods, for example, chimpanzees and orangutans are weaned at an average of ~5.3 and 7.7 years of age, respectively (Clark, 1977; Galdikas and Wood, 1990; Hayssen, 1993). In contrast, polytocous eutherian mammals including dogs, cats, pigs, rodents and rabbits, give birth to semi-altricial or altricial young, more like the marsupial reproductive strategy (Hayssen, 1993).

The semi-aquatic pinnipeds (fin-footed mammals), which include the Otariidae

(eared seals: fur seals and sea lions), Phocidae (earless seals or 'true seals') and Odobenidae (walruses) families have very different reproductive strategies (Oftedal *et al.*, 1987; Schulz and Bowen, 2005). Otariids, such as the Cape (or South African) fur seal and subantarctic fur seal, have adopted a unique 'foraging-cycle' strategy (Boness and Bowen, 1996; Oftedal *et al.*, 1987). The mother generally spends the first week on shore suckling her pup but then goes to sea to feed. These foraging periods average 23-28 days for the subantarctic fur seal (Georges and Guinet, 2000; Kirkman *et al.*, 2002) and 25 days for the Cape fur seal (Gamel *et al.*, 2005), but periods at sea of up to two months have been recorded (Verrier *et al.*, 2011). Each foraging period is interspersed with the mother spending 2-3 days on shore suckling her young (Oftedal, 2004). Therefore, otariid seals 'suspend' lactation whilst foraging at sea, but resume milk production upon returning to shore, with no evidence of the gland 'involuting' at sea (Arnould and Boyd, 1995b; Cane, 2005; Sharp *et al.*, 2005). Thus, these seals provide a fascinating model for the regulation of lactational persistency and involution (Cane, 2005; Sharp *et al.*, 2005). In contrast, for most mammals, removal of sucking stimulus for several days results in cessation of lactation and involution of the mammary gland (Khalil *et al.*, 2008; Watson, 2006; Wilde *et al.*, 1999).

Development of the Mammary Gland

The mammary gland is unique, as it is the only organ to undergo most of its development postnatally (Andres and Djonov, 2010; Hennighausen and Robinson, 2001; Topper and Freeman, 1980). Mammary gland development occurs in distinct stages closely linked to sexual development and reproduction (Inman *et al.*, 2015; Martinet and Houdebine, 1999; Topper and Freeman, 1980). These include: embryogenesis, puberty, and changes associated with each oestrous cycle in the adult (Ormandy *et al.*, 1997b). However, the final "functional differentiation" stage of mammary development occurs in females who become pregnant, give birth to and secrete milk for a newborn (Anderson *et al.*, 2007; Hennighausen and Robinson, 1998; Shannon and Pitelka, 1981). There are five stages of functional differentiation which are triggered by pregnancy: *proliferation, secretory differentiation, secretory activation*, *lactation* and *involution* (Anderson *et al.*, 2007).

The mammary gland is composed of a mammary fat pad, or 'stroma' of connective tissue and adipocytes into which the mammary epithelium or 'parenchyma' grows (Hovey *et al.*, 1999; Inman *et al.*, 2015; Macias and Hinck, 2012; Neville *et al.*, 1998). The mammary epithelium forms an extensive network of ducts and milk-producing

alveolar cells surrounded by myoepithelial cells within the stroma (Brisken and Ataca, 2015; Findlay and Renfree, 1984; Richert *et al.*, 2000). Signalling between the parenchyma and stroma is essential for the development of the mammary epithelium (Hovey and Trott, 2004; Wiseman and Werb, 2002). In particular, the stroma mediates hormone actions and influences ductal and branching morphogenesis and alveolar differentiation (Brisken and Rajaram, 2006; Fata *et al.*, 2004; Sternlicht, 2006). The mammary gland also contains immune cells of the lymphatic system, vascular endothelial cells which comprise the blood vessels and stem cells (Hassiotou *et al.*, 2012; Shackleton *et al.*, 2006; Watson, 2009).

The endocrine system is the main regulator of the mammary gland (Brisken and Ataca, 2015; Neville *et al.*, 2002; Topper and Freeman, 1980). Reproductive hormones such as estrogen, progesterone, placental lactogen (PL), prolactin (PRL), and oxytocin directly affect the mammary gland (Hovey *et al.*, 2002; Neville *et al.*, 2002). Metabolic hormones including: growth hormone (GH), insulin, glucocorticoids (cortisol and corticosterone) and thyroid hormones: thyroxine (T_4) and the more-active triiodothyronine (T_3), have a direct effect, or play permissive or supportive roles (Table 1.1) (Forsyth, 1991; Neville *et al.*, 2002; Vonderhaar and Greco, 1979). Although traditionally regarded as metabolic hormones, insulin and cortisol are essential for milk protein gene expression (Bolander *et al.*, 1981; Menzies *et al.*, 2009; Nicholas *et al.*, 1983). The mammary gland itself produces hormones such as GH, PRL, parathyroid hormone releasing protein (PTHrP) and leptin (Clevenger and Plank, 1997; Neville *et al.*, 2002; Smith-Kirwin *et al.*, 1998; Steinmetz *et al.*, 1993; Wojcik *et al.*, 1999). Selected hormones that regulate the expression of milk protein genes will be explored later in this review.

Cues from the local tissue microenvironment, i.e. autocrine and paracrine factors, also influence mammary proliferation, differentiation and apoptosis (Hovey and Trott, 2004; Howlett and Bissell, 1993; Lin and Bissell, 1999; Streuli, 1993). These include: the serotonin [5-hydroxytryptamine (5-HT)] neurotransmitter (Horseman and Collier, 2014); the galanin neuropeptide, epidermal growth factor (EGF); fibroblast growth factor; transforming growth factors (TGFα and TGFβ); insulin-like growth factors I and II (IGF-I and IGF-II); IGF binding proteins (IGFBPs); and the IGFBP-related proteins (Blum and Baumrucker, 2008; Marshall *et al.*, 2010; Matsuda *et al.*, 2004; Naylor *et al.*, 2003; Plath *et al.*, 1998). Epigenetic mechanisms including histone methylation and acetylation and DNA methylation also regulate mammary development and function (Rijnkels *et al.*, 2010; Singh *et al.*, 2010).

Table 1.1. Hormones involved in the development and regulation of the female mammary glands during pregnancy, lactation and involution*

Hormone#	Site of Action	Ductal morphogenesis	Functional differentiation				
			Alveolar morphogenesis			(iv) Lactation	(v) Involution
			(i) Proliferation (early-pregnancy)	(ii) Secretory differentiation (mid-pregnancy)	(iii) Secretory activation (parturition)		
Stimulus	-	reproductive state				sucking	no sucking
Estrogen (E_2)	developing follicles of the ovaries, corpus luteum, placenta				withdrawal	inhibitory?	
Progesterone (P)	corpus luteum post-ovulation, adrenal glands, placenta				withdrawal	-	
Prolactin (PRL)	anterior pituitary, mammary gland						
Placental Lactogen (PL)	placenta						
Oxytocin	posterior pituitary gland						
Growth hormone (GH)	anterior pituitary gland, mammary gland		possible				
Corticosteroid (HC)	adrenal cortex of the adrenal glands	unknown					
Thyroid hormones (T_3 and T_4)	thyroid gland	unknown		possible	unknown		
Insulin (I)	pancreas	unknown					

*Table adapted from (Neville *et al*., 2002); #Reproductive hormones (pink), Metabolic hormones (blue); Hormone effects: Required (dark green), required in some species (light green), possible direct action (grey), metabolic effect (orange).

The Secretion of Milk

Milk secretion is a complex and highly-co-ordinated process regulated by endocrine, autocrine and paracrine mechanisms (Burgoyne and Wilde, 1994; Linzell and Peaker, 1971). There are five different pathways by which solutes enter milk: the exocytotic/Golgi (I); lipid secretion (II); membrane transport (III); transcytotic (IV); and paracellular (V) pathways (Figure 1.2) (McManaman and Neville, 2003; Shennan and Peaker, 2000). Pathways I-IV are transcellular, whereby components produced by the mammary epithelial cell (MEC) cross the apical membrane and enter the alveolar lumen (McManaman and Neville, 2003; Shennan and Peaker, 2000). In contrast, the paracellular pathway permits the two-way movement of high and low molecular weight substances such as plasma components and leukocytes between the interstitial space and alveolar lumen via leaky tight junctions (McManaman and Neville, 2003; Shennan and Peaker, 2000). This pathway is active during pregnancy and involution, but inactive during lactation; except during mastitis infections. At parturition, the increase in glucocorticoids, fall in progesterone and activation of the prolactin receptor (PRLR) instigates tight-junction closure between MECs, blocking this pathway (Nguyen *et al.*, 2001; Shennan and Peaker, 2000).

The Composition of Milk

Since the evolution of the mammary glands and lactation ~310 MYA, milk has evolved into a complex, specialised, biological fluid that is remarkably tuned to the specific developmental needs of the developing young, which are often born at very different developmental stages. (Jenness, 1974; Oftedal, 2013). However, due to the diverse range of mammalian reproductive strategies, the proportions of carbohydrates, proteins, fats, minerals and vitamins, i.e. milk composition, varies greatly between species (Jenness, 1974). The eutherian mother secretes colostrum for the first 24-48 h pp (Delouis, 1978; Nakamura *et al.*, 2003), but then produces ‘mature milk’, which generally changes very little throughout lactation (Allen *et al.*, 1991; Picciano, 2001). However, there are some exceptions, e.g. elephant seals, bears (see Oftedal, 1984). Marsupials also secrete colostrum for the first few days pp (Deane and Cooper, 1984; Griffiths and Slater, 1988). However, unlike most eutherians, the marsupial mother progressively alters milk composition and individual milk constituents throughout the long and complex lactation (Green *et al.*, 1980; Joss *et al.*, 2009; Nicholas, 1988a; Nicholas *et al.*, 1997).

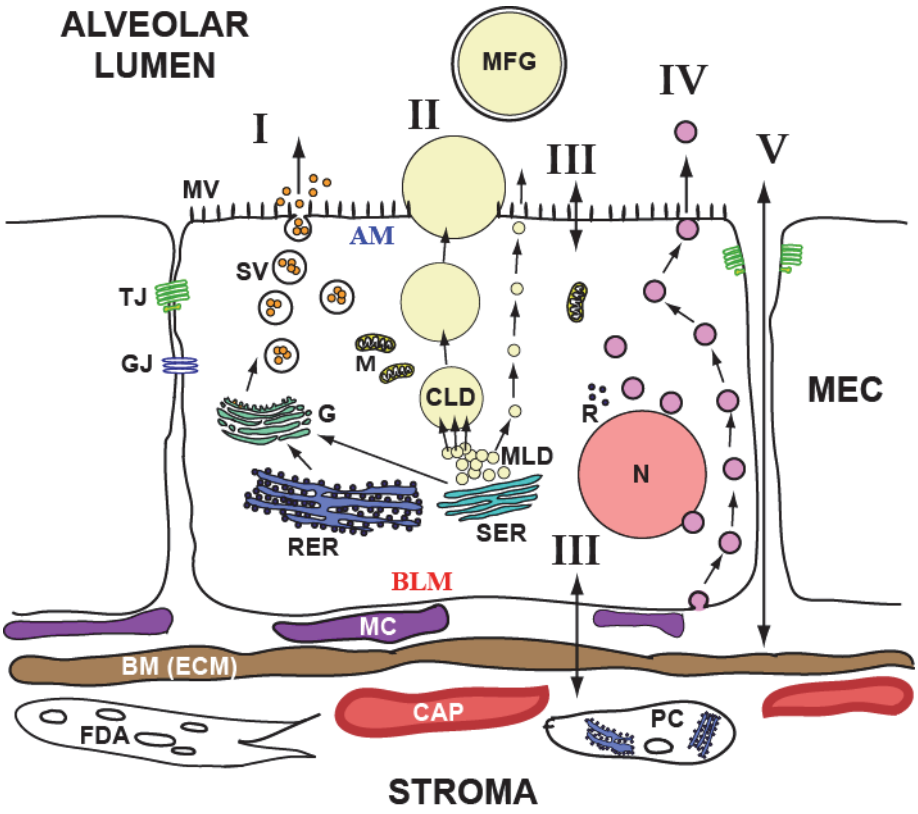


Figure 1.2. The five transport pathways via which constituents enter milk

Solutes enter the alveolar lumen and hence milk by one of four transcellular pathways (I-IV), or a fifth, paracellular pathway (V). The exocytotic/Golgi pathway (I) involves the transport of proteins (which are assembled on the rough endoplasmic reticulum (RER)), plus lactose, complex sugars, calcium, phosphate and citrate ions and water. Components are packaged into membrane-bound secretory vesicles (SV) by 'budding-off' from the *trans*-Golgi and their contents released into the alveolar lumen across the apical membrane (AM). Lipids enter milk via the lipid secretion pathway (II). Triglycerides and phospholipids are assembled on the smooth endoplasmic reticulum (SER) into microlipid droplets (MLDs). MLDs are either secreted into the alveolar lumen, or fuse to form larger cytoplasmic lipid droplets (CLDs), which are secreted as milk fat globules (MFGs). The membrane transport pathway (III) permits the bi-directional movement of ions and small molecules, e.g. amino acids, glucose and water, across both the AM and basolateral membrane (BLM). The transcytotic pathway (IV) involves the transport of maternal serum components such as immunoglobulins, transferrin and serum albumin and hormones (prolactin, estrogen and insulin), plus stromal-derived components (e.g. secretory immunoglobulin A (sIgA), cytokines and lipoprotein lipase). In contrast, for the paracellular pathway (V), constituents travel between, rather than through MECs. However, this route is only available during pregnancy and involution, or during a case of mastitis. G, Golgi; MV, microvilli; M, mitochondria; N, nucleus; R, free ribosomes; TJ, tight junction; GJ, gap junction; PC, plasma cell; FDA, fat-depleted adipocyte; CAP, capilliary; BM (a specialised ECM), basement membrane (extracellular matrix). Redrawn from (McManaman and Neville, 2003).

Colostrum

Colostrum contains bioactives, antimicrobials and growth factors and is vital for the health and gastrointestinal development of the young (Jørgensen *et al.*, 2010; Walker, 2004). Marsupial colostrum is a dilute, translucent secretion that contains a high whey protein content including IgG, IgA, transferrin, protein, lipids and lactose (Deane and Cooper, 1984; Griffiths and Slater, 1988; Joss *et al.*, 2009). Eutherian colostrum is similar to that of marsupials but also has high levels of lactoferrin, lysozyme and ferritin (Brambell, 1969; Hernandez-Castellano *et al.*, 2014; Stelwagen *et al.*, 2009).

All mammalian neonates have an innate immune system at birth and can therefore mount a simple immune response against non-specific pathogens (Newburg and Walker, 2007). However, not all newborns have a mature acquired (adaptive) immune system due to the lack of (or minimal) transfer of passive immunity (predominantly as immunoglobulin G, IgG) via the placenta *in utero* (Hurley and Theil, 2013). They cannot therefore mount an antibody-mediated response to specific pathogens and after birth are reliant upon the transfer of maternal IgG via colostrum (Hurley and Theil, 2011). For mammals such as the cow, horse, pig, goat and North and South American opossums, colostrum provides the sole mechanism for IgG transfer (Baumrucker *et al.*, 2010; Hurley and Theil, 2013; Yadav, 1971). However, in other species, including humans and rabbits, IgG transfer occurs across the placenta via the neonatal Fc receptor (FcRn), or via both the placenta and milk (dog, mouse, rat, tammar) (Brambell, 1970; Larson, 1992; Renfree, 1973). Failure of IgG transfer can be fatal in neonates, and may also cause gastrointestinal, respiratory and urinary infections and impaired growth (Brambell, 1970; Hurley and Theil, 2013; Smith and Little, 1922).

For mammals, there is a limited period in which immunoglobulins and other macromolecules can be absorbed intact across the neonate intestines (Hurley and Theil, 2013; Telemo and Hanson, 1996). After birth there is a progressive loss in intestinal absorptive capacity as gut cells of the young mature and produce a microvillus surface. This signifies ‘gut closure’ (Kruse, 1983; McFadden *et al.*, 1997). Gut maturation varies from ~24 h pp for ruminants, 36 h pp for pigs and horses and up to 21 days pp for rodents (McFadden *et al.*, 1997; Telemo and Hanson, 1996). After gut closure, although immunoglobulins cannot cross the intestinal wall and enter the circulatory system, they may provide protection against enterotoxins and bacterial and viral infections (Belov *et al.*, 2002; Hurley and Theil, 2013; Larson *et al.*, 1980).

At birth, the marsupial young receives only basic protection from innate immune system components, as it lacks circulating lymphocytes or organised lymphoid tissues

(Basden *et al.*, 1997; Borthwick *et al.*, 2014; Siddle *et al.*, 2010). They therefore rely upon maternal immunoglobulins in milk for antibody-mediated immunity (Adamski and Demmer, 2000; Deane *et al.*, 1990; Yadav, 1971). Two periods of elevated immunoglobulin transfer occur in the tammar and brushtail possum coincident with increased exposure to microorganisms. The first occurs immediately postpartum and the second, as the PY exits the pouch and begins to ingest herbage (Adamski and Demmer, 2000; Daly *et al.*, 2007; Deane and Cooper, 1988). As the tammar PY lacks a mature acquired immune system until well after 100 days pp and gut closure does not occur until ~170-180 days pp, the period of immunoglobulin transfer via milk is extended compared to that of the eutherians (Deane *et al.*, 1990; Old and Deane, 2000).

For mammals that transfer immunoglobulin via colostrum alone, IgG is usually the major Ig isotype (Hurley, 2003; Larson *et al.*, 1980). However, for those that transfer IgG *in utero*, the major milk immunoglobulin is secretory IgA, which is produced by mucosal and exocrine glands (Hunziker and Kraehenbuhl, 1998; Lascelles, 1979; McFadden *et al.*, 1997). IgA is also secreted in tammar milk, but does not cross the PY intestinal epithelium and so remains in the gut to provide protection against bacterial and viral pathogens (Adamski and Demmer, 2000; Deane *et al.*, 1990). The marsupial pouch also provides both a physical and chemical barrier to pathogens (Ambatipudi *et al.*, 2007; Deane and Miller, 2000; Wang *et al.*, 2011). Pouch secretions include immunomodulators such as cathelicidin, dermicin, eugenin and cystatin C (Ambatipudi *et al.*, 2008; Baudinette *et al.*, 2005; Daly *et al.*, 2008; Edwards *et al.*, 2012).

Mature Milk

As mentioned previously, the absolute concentrations and proportions of the major milk constituents (carbohydrates, fats (lipids) and proteins) vary between species (Table 1.2). In general, changes in milk composition during the lactation cycle tend to be more dramatic in marsupials than for eutherians. This may be due to the altricial nature of the marsupial young and its changing nutritional needs from development in Phases 2A and 2B to growth in Phase 3 depend on the striking changes in milk composition throughout lactation (Figure 1.3). For the first 200 days pp, tammar milk is dilute, lipid and proteins are low, and carbohydrates are high. In contrast, as the fully developed PY exits the pouch and begins to ingest herbage in Phase 3, there are dramatic changes in milk production, PY growth and milk composition. Milk becomes more concentrated, protein, lipid (and hence energy-content) more than double, while milk carbohydrate drops precipitously (Green *et al.*, 1983; Green *et al.*, 1980; Messer and Green, 1979; Messer *et al.*, 1984).

Table 1.2. Milk composition of selected marsupials, eutherians and monotremes

Milk constituents are shown as percentages of total milk volume.

Species	Lactation stage	Water (%)	Protein (%)	Fat (%)	Carbohydrates (%)
Tammar Wallaby (*Macropus eugenii*)[1]	Early lactation (Phase 2A)	>80	4	2	7
	Mid-lactation (Phase 2B)	>80	4	2	7
	Late lactation (Phase 3)	>60	11	23	1
Brushtail possum (*Trichosurus vulpecula*)[2]	Early lactation	~90	3.6	1.5	3.9
	Mid-lactation	72	7.3	3.8	11.1
	Late lactation	82.2	5.3	6.8	4.1
Platypus (*Ornithorhynchus anatinus*)[3]		60.9	8.2	22.2	3.7
Echidna (*Tachyglossus aculeatus*)[3]		51.1	12.4	31.0	2.3
Human (*Homo sapiens*)[4]		87-89	0.8-0.9	3-5	6.9-7.2
Pig (*Sus scrofa*)[5]		80	5.8	8.2	4.8
Cow (*Bos taurus*)[6]		87.3	3.2	3.9	4.6
Sheep (*Ovis aries*)[5]		83	5.5	5.3	4.6
Horse (*Equus caballus*)[7]	Mid-lactation	89.5	1.93	1.29	6.9
Dog (*Canus familiaris*)[8]		77.3	7.53	9.47	3.81
Antelope (*Antilope cervicapra*)[5]		74.8	6.9	1.3	4.0
Polar bear (*Ursus maritimus*)[9]		57.1	10.2	31.0	0.5
Australian fur seal (*Arctocephalus pusillus doriferus*)[10]		43.6	10.4	42.1	*
Hooded seal (*Cystophora cristata*)[11]	Early lactation	34	6.2	56	0.9
	Late lactation	30	5.1	61	1.0
Common bottlenose dolphin (*Tursiops truncatus*)[12]	Early lactation	75.3	8.5	10.5	1.4
	Late lactation	70.0	10.5	16.6	1.1
Bryde's Whale (*Balaenoptera edeni*)[13]		52.1	14.7	29.6	#

[1](Green, 1984; Green *et al.*, 1980), [2](Cowan, 1989), [3](Griffiths *et al.*, 1984), [4](Jenness, 1979), [5](Hurley, 2004), [6](Oftedal and Iverson, 1995), [7](Oftedal *et al.*, 1983), [8](Oftedal, 1984), [9](Cook *et al.*, 1970), [10](Arnould and Hindell, 1999), [11](Oftedal *et al.*, 1988), [12](West *et al.*, 2007), [13](Oftedal, 1997), ^Data extracted from figures (Cowan, 1989), *Carbohydrate content of seal milk is typically <0.5% of the total milk volume (Arnould and Boyd, 1995a; Oftedal *et al.*, 1987), with minimal CHO content (Urashima *et al.*, 2001), #Cetacean milks also low in sugars (0.7-2.6%) (West *et al.*, 2007).

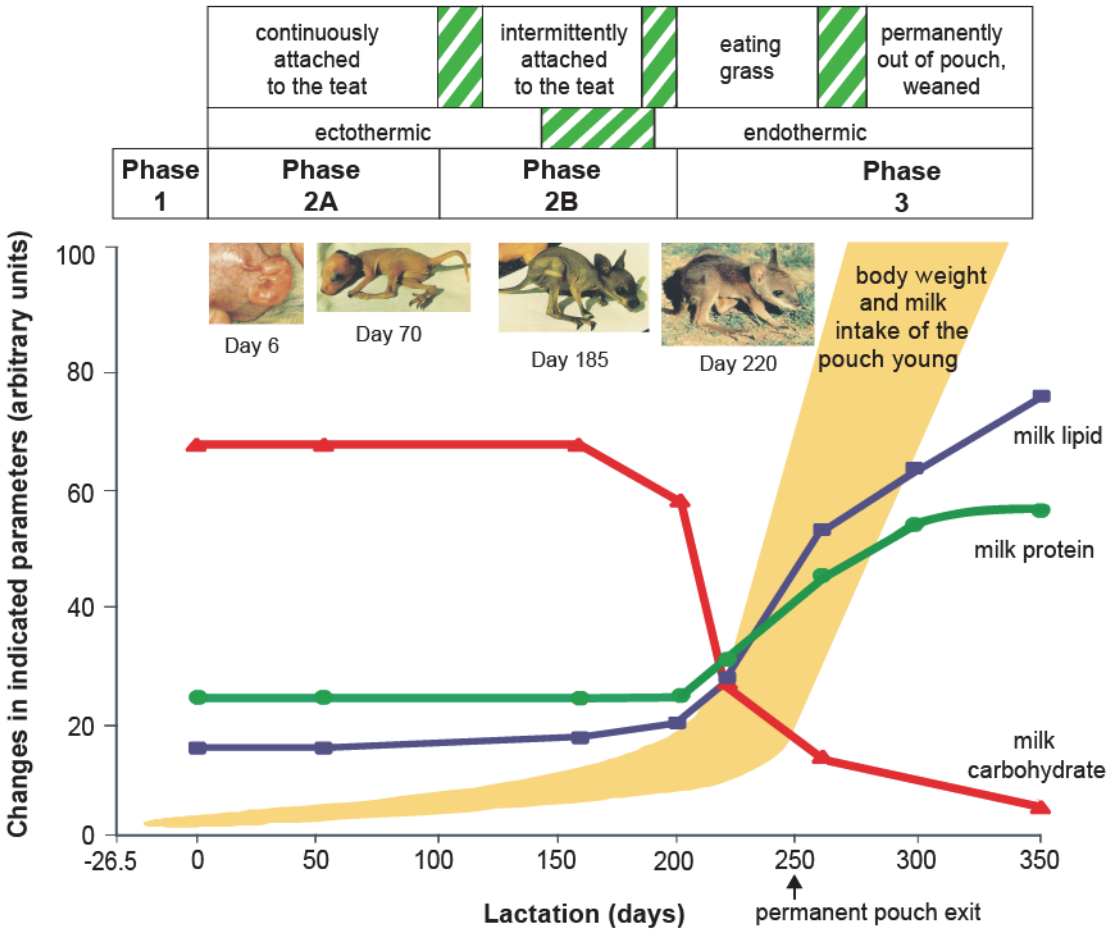


Figure 1.3. Changes in milk composition and the growth and development of the pouch young during the tammar wallaby lactation cycle

The lactation cycle of the tammar comprises a 26.5 day pregnancy after diapause (Phase 1), and a 300-350 day lactation period (Phases 2A, 2B and 3) during which fats, proteins and lipids change dramatically. Each phase is correlated with altered suckling patterns and PY development (green and white diagonal striped boxes) (Tyndale-Biscoe and Janssens, 1988; Tyndale-Biscoe and Renfree, 1987). The PY spends the first 200 days pp in the pouch and ingests dilute milk. In contrast, Phase 3 marks the striking increases in PY body weight and milk intake; and maternal mammary gland size and milk production (Findlay, 1982). Permanent pouch exit occurs around 250 days pp and weaning approximately 50-100 days later. Adapted from (Green and Merchant, 1988; Tyndale-Biscoe and Janssens, 1988).

Like most eutherians, lactose is the major carbohydrate in tammar milk (~90% (w/v)), but only for the first 4 days pp (Messer *et al.*, 1984; Messer and Mossop, 1977; Urashima *et al.*, 1994). Oligosaccharides then comprise the major milk sugar until around 200 days pp. As lactose is the major milk osmole in milk for most mammals and directly affects milk volume (Bleck *et al.*, 2009; Neville *et al.*, 1983; Zhao and Keating, 2007), the secretion of oligosaccharides allows the marsupial mother to secrete low volumes of a high-carbohydrate milk (Green and Merchant, 1988; Messer and Green, 1979). Similarly, the production of little or no milk lactose and minimal carbohydrates also provides evolutionary benefits for seals, whales and dolphins, allowing them to secrete a dense, high-fat, high-energy milk (Arnould and Hindell, 1999; Ashworth *et al.*, 1966; Oftedal, 1997). Some milk oligosaccharides may also promote the growth of beneficial bacteria and prevent the colonisation of the gut by pathogenic microbes (Jost *et al.*, 2015; Kunz and Rudloff, 1993; Martin *et al.*, 2002). They may also enhance gastrointestinal development and promote the gut maturation in young mammals (Kwek *et al.*, 2009b). Hence, whether the minimal oligosaccharide content of most pinniped milks affects the health of the young is unknown.

Like eutherians, triacylglyerols, composed of a glycerol backbone to which three fatty acids are attached, constitute the major lipid of both eutherian and marsupial milks (~75-97%) (Fox and McSweeney, 1998; Green *et al.*, 1983; Green *et al.*, 1980). However, during early lactation, phospholipids comprise up to 25% of tammar milk lipid (Green *et al.*, 1983). Phospholipids are used for the formation of tissues and organs, particularly the brain and nervous system (Fox and McSweeney, 1998; Hanahan and Thompson, 1963). Hence, they are presumably critical for the rapid, disproportionate growth and development of the PY brain during Phase 2A and the completion of the development of most of the major organs PY organs during the first 200 days pp (Renfree *et al.*, 1982).

Milk proteins play important roles in mammalian milk, acting as enzymes, structural proteins (collagen, laminin), peptide hormones (insulin, PL, PRL, GH), growth factors (IGF-I, EGF) and protective agents (protease inhibitors) and bioactives (immunoglobulins, lysozyme and casein phosphopeptides) (Hambraeus and Lonnerdal, 2003; Sharp *et al.*, 2014; Vilotte *et al.*, 2003). Common to all milk, the protein fraction includes the acid-insoluble caseins and the soluble whey proteins, the proportions of which vary between species (Fiat *et al.*, 1993; Fox and McSweeney, 1998). The caseins form ~80% of the protein content of cow's milk (Fox and McSweeney, 1998; Groenen and van der Poel, 1994), but only 50-60% of tammar milk (Green, 1984; Green *et al.*,

1980). Bovine milk protein content ranges from 3-4% w/v during lactation, similar to the first 200 days of tammar lactation (4% w/v, ~25% w/w; Figure 1.3) (Green, 1984; Green *et al.*, 1980). However, unlike the cow, Phase 3 tammar milk protein increases up to (11% w/v, ~35% w/w; Table 1.2) (Green *et al.*, 1980), perhaps due to the increase in maternal plasma PRL concentration in late Phase 2B (Hinds and Tyndale-Biscoe, 1982).

The secretion of milk caseins during late pregnancy and throughout lactation is consistent for both marsupials and eutherians (Bird *et al.*, 1994; Mepham, 1992; Nicholas *et al.*, 1997). However, the temporal, asynchronous secretion of selected whey proteins throughout lactation is quite unique amongst marsupials (Nicholas *et al.*, 1997; Sharp *et al.*, 2009).

Whey Proteins

The whey proteins are cysteine-rich, globular polypeptides (Swaisgood, 1993). Some are produced by the mammary gland, whilst others, such as serum albumin, transferrin, complement, acute phase proteins and immunoglobulins (IgG, IgE and IgM) are derived from maternal serum (Vorbach *et al.*, 2006). The *α-lactalbumin* (*LALBA*), *β-lactoglobulin* (*LGB*) and *β-1,4 galactosyltransferase* (*B4GALT1*) genes are induced at parturition and expressed throughout lactation in marsupials, as for *α-casein* (*CSN1*) and *β-casein* (*CSN2*) (Figure 1.4) (Bird *et al.*, 1994; Nicholas *et al.*, 1997). However, many whey protein genes are expressed asynchronously. These include: *ELP* (Phase 2A-early Phase 2B); whey acidic protein (*WAP*, Phase 2B-early Phase 3) and *late lactation proteins A* and *B* (*LLPA* and *LLPB*, late Phase 2B-Phase 3 and Phase 3, respectively). Selected sugar-modifying enzymes are also temporally expressed: *β-1,3 galactosyltransferase* (*B3GALT1*, early Phase 2A-Phase 2B) and *β-galactosidase* (*BGAL*, Phase 3). The precise function(s) of many of these tammar whey proteins is unknown.

Early Lactation Protein. ELP was first identified in brushtail possum milk (Piotte and Grigor, 1996) in both an 8 kDa non-glycosylated form and a 16 kDa heavily glycosylated form (composed of ~60% sugar based upon SDS-PAGE band size) (Piotte and Grigor, 1996). ELP was also purified from tammar milk by high performance liquid chromatography (HPLC) as an 18 kDa band (SDS-PAGE analysis) (Simpson, 1998; Simpson *et al.*, 1998). Subsequently, seven glycosylated forms of ELP ranging from ~16-20.5 kDa were identified by 2D-gel electrophoresis (Joss *et al.*, 2007). ELP is a major tammar whey protein during the first 73 days of lactation, before declining gradually until barely detectable at day 129 pp (Simpson, 1998; Simpson *et al.*, 1998).

The mature tammar ELP peptide shares 74.5% similarity with possum ELP and 37% with the colostrum-specific bovine colostrum trypsin inhibitor (CTI) (Simpson *et al.*, 1998a), but is less similar to BPTI (bovine pancreatic trypsin inhibitor) and STI (spleen trypsin inhibitor/serum basic protease inhibitor, Figure 1.5). Sequence analysis of marsupial ELP and bovine CTI suggested that these proteins are part of the Kunitz family of serine protease inhibitors (Cechova *et al.*, 1969; Kunitz and Northrop, 1936; Piotte and Grigor, 1996; Simpson *et al.*, 1998a).

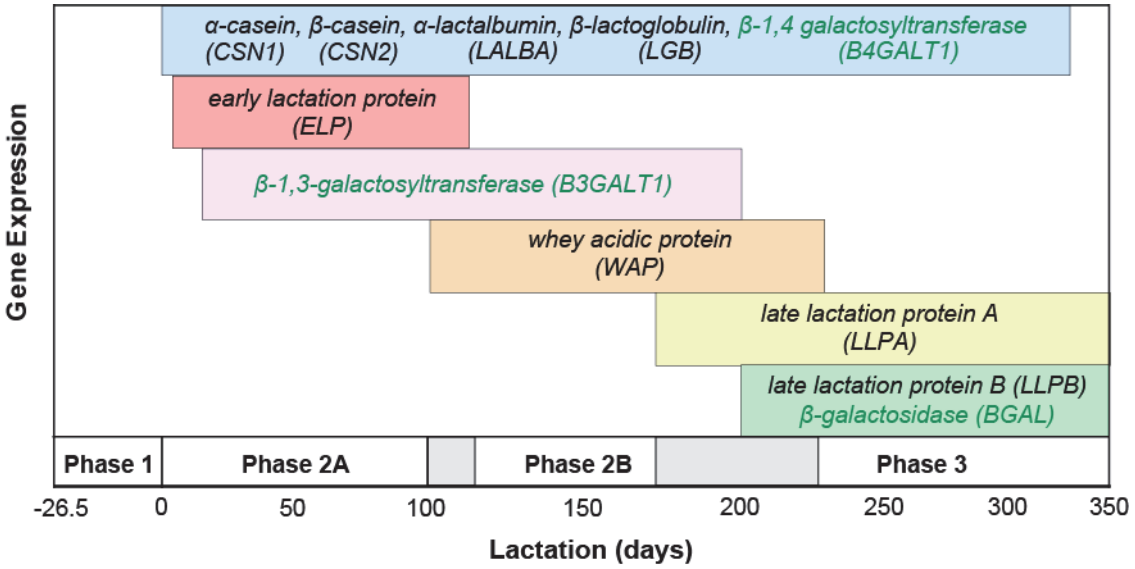


Figure 1.4. Expression of selected tammar milk protein genes and enzymes throughout the lactation cycle

The *α-casein* (*CSN1*), *β-casein* (*CSN2*), *α-lactalbumin* (*LALBA*), *β-lactoglobulin* (*LGB2*) and the *β-1,4-galactosyltransferase* (*B4GALT1*) Golgi transmembrane protein gene are induced at parturition and expressed throughout lactation. However, many whey protein genes are asynchronously expressed in a Phase-specific manner during lactation, though there is some overlap in gene expression between the Phases (grey boxes). While *ELP* is expressed during Phase 2A/early Phase 2B, the precise timing of its induction is unclear. Another Golgi transmembrane protein gene, *β-1,3-galactosyltransferase* (*B3GALT1*) is expressed from ~8-200 days pp and catalyses the formation of complex carbohydrates. *Whey acidic protein* (*WAP*) is expressed during mid-lactation (Phase 2B) to early Phase 3 (~240 days pp). In contrast, *Late lactation proteins A* and *B* (*LLPA*, *LLPB*) and the *β-galactosidase* (*BGAL*) gene, which encodes a glycosidase hydrolase enzyme, are predominantly expressed during late lactation (Phase 3) (Collet *et al.*, 1989; Menzies and Nicholas, 2007; Nicholas, 1988a, b; Nicholas *et al.*, 1987; Nicholas *et al.*, 1997; Nicholas *et al.*, 1994; Nicholas *et al.*, 1995; Simpson *et al.*, 1998a; Simpson *et al.*, 2000; Trott, 1999; Trott *et al.*, 2002).

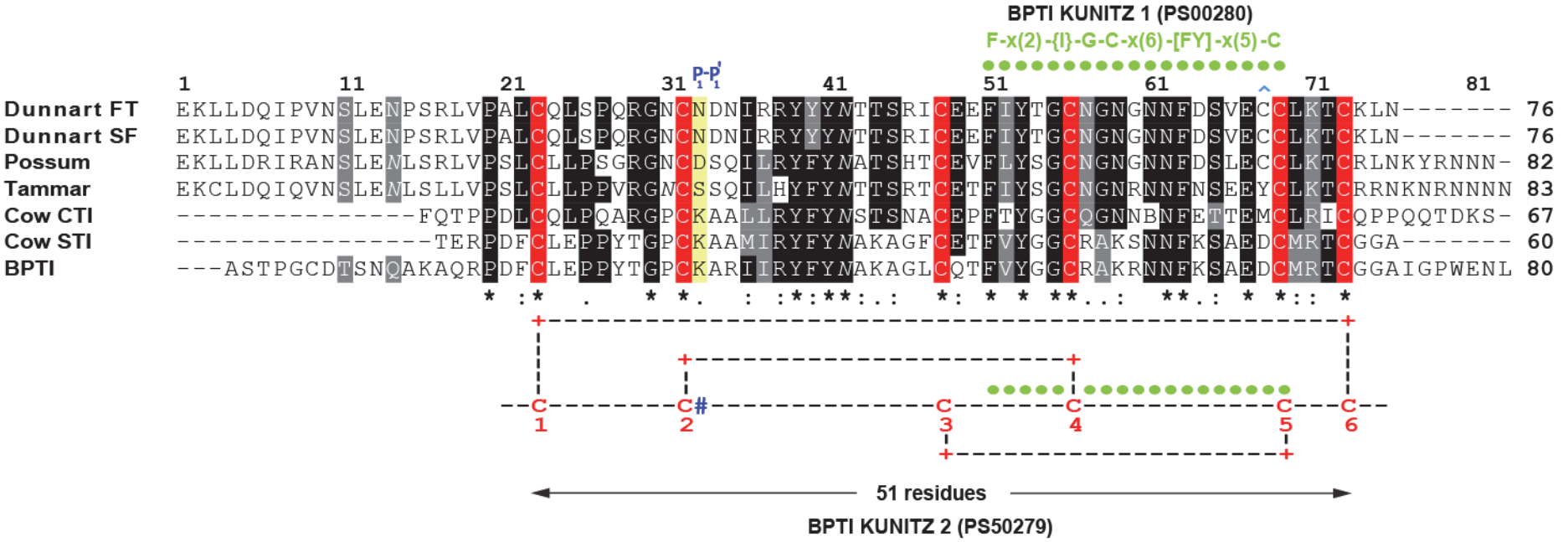


Figure 1.5. Alignment of the mature tammar ELP peptide with selected Kunitz protease inhibitors

Clustal Omega alignment of mature ELP of the tammar [GenBank: Q29143], stripe-faced and fat-tailed dunnarts (De Leo *et al.*, 2006) and brushtail possum [GenBank: Q29143] and bovine CTI [GenBank: P00976], BPTI [GenBank: P00974] and STI [GenBank: P00975]. Amino acid residues common to all sequences are shaded black (*).The BPTI KUNITZ 2 motif [Prosite: PS50279] spans a 51 aa region and has 6 conserved cysteine residues (red shading, red crosses), which form three disulphide bonds (dashed lines) within the domain (C1-C6, C2-C4 and C3-C5). The BPTI KUNITZ 1 motif spans a shorter 19 aa region encompassing C4, C5 and C6 (•). The putative reactive site residue (P_1, #) is shaded yellow. Protease inhibition involves cleavage of the P_1-P_1' peptide bond and the formation of a new bond between the inhibitor and protease. Italicised asparagine residues indicate predicted N-glycosylation sites. ^Unpaired cysteine residue in possum and dunnart ELP. Conservation between groups of amino acids with 'strongly similar' (:) and 'weakly' similar (.) properties, i.e., scoring >0.5 and <0.5 in the Gonnet PAM 250 matrix, respectively, are shown. Gaps within the alignment are indicated (-).

ELP gene expression in the mammary gland is specific to early lactation for the tammar, brushtail possum and fat-tailed dunnart (De Leo *et al.*, 2006; Demmer *et al.*, 1998; Demmer *et al.*, 2001; Nicholas *et al.*, 1997; Piotte and Grigor, 1996; Simpson *et al.*, 1998a) and to the suckled gland (Demmer *et al.*, 1998). Expression of brushtail possum *ELP* increases ~6-fold in the sucked gland from 1.5-6 days pp, but is minimal in the non-sucked gland on day 1.5 pp and undetectable by day 6 pp (Demmer *et al.*, 1998). However, neither the precise timing of tammar *ELP* gene induction, nor whether the gene is expressed in multiple tissue types are known (Simpson, 1998).

The marsupial *ELP* gene comprises three exons and two introns and spans ~4.9 kb and ~3.5 kb for the Striped faced dunnart and South American opossum, respectively (De Leo *et al.*, 2006). The first exon encodes the signal peptide and mature protein N-terminus, the second, the Kunitz domain and the third, the stop codon and/or the ELP C-terminus (De Leo *et al.*, 2006). Whilst the *BPTI* gene has a similar 3-exon gene structure (Kingston and Anderson, 1986), the structure of the eutherian *CTI* gene was not known.

Over 65 years ago, CTI, the putative eutherian orthologue of marsupial ELP was crystallised from bovine and porcine colostrum (Laskowski and Laskowski, 1950, 1951; Laskowski *et al.*, 1957). These proteins were colostrum-specific, inhibitors of trypsin and plasmin (Feeney *et al.*, 1969; Kassell and Laskowski, 1956; Laskowski and Laskowski, 1951; Laskowski *et al.*, 1957) and resistant to pepsin degradation (Kassell and Laskowski, 1956). An acid-soluble protease inhibitor was subsequently identified in sow, cat and ewe colostrum, but was not detected in horse colostrum (Baintner and Csapo, 1996; Baintner, 1973, 1976). Of these proteins, only bovine CTI has been sequenced (Cechova and Ber, 1974; Cechova *et al.*, 1969). Furthermore, while CTI is only produced by bovine MECs during the period of colostrum secretion, it has not been detected in other tissues (Veselsky *et al.*, 1978).

An unidentified trypsin inhibitor is also present in the intestines and urine of young lambs, kittens and piglets (Baintner, 1973, 1976; Carlsson *et al.*, 1974). Its disappearance correlates with the advent of trypsin activity in the small intestines and urine of the young and gut closure. Hence, this protein may delay the development of proteolytic activity in the small intestines, allowing enhanced protein absorbance by the gut (Baintner, 1976). It is unclear whether this inhibitor is CTI, but it may be an orthologue of the ubiquitously expressed *BPTI* (Ascenzi *et al.*, 2003; Veselsky *et al.*, 1985, 1986). Like CTI, BPTI has a single Kunitz domain, an identical P_1 'reactive' site

residue that determines serine protease specificity and similar primary and secondary structures (Cechova *et al.*, 1969). Alternatively, the unknown protein may be urinary trypsin inhibitor, also known as bikunin or plasmin inhibitor, which is present in many tissues, blood and bodily secretions (Chang-Yi *et al.*, 1999; Fries and Blom, 2000; Marlor *et al.*, 1997).

Like marsupial ELP, bovine and porcine CTI are glycosylated, producing isoforms of 10-14 kDa and 18 kDa, respectively (Carlsson *et al.*, 1974; Kress *et al.*, 1971; Yoshimoto and Laskowski, 1982). Bovine CTI is 40% glycosylated at the asparagine residue Asn27 (Asn42, alignment numbering Figure 1.5, Figure 1.6) (Cechova *et al.*, 1969; Klauser *et al.*, 1978; Tschesche *et al.*, 1975). Tammar and possum ELP are also predicted to be N-glycosylated at the equivalent site (Asn42) based upon the Asn-Xaa-Ser/Thr sequons (Nakai and Kanehisa, 1988; Piotte and Grigor, 1996; Simpson *et al.*, 1998a). However, tammar and possum ELP may have additional N-glycosylation sites; at Asn14 and Asn64; and Asn14, respectively (Piotte and Grigor, 1996; Simpson *et al.*, 1998a). It is unknown whether glycosylation affects the function of marsupial ELP. However, the N-glycosidic bond in bovine CTI occurs on the side opposite to the reactive site, at the 'base' of the pear-shaped protein, and is therefore unlikely to interfere directly with the P_1 protease reactive site (Figure 1.6) (Klauser *et al.*, 1978).

During the period when *ELP* is expressed, the marsupial PY is confined to the pouch, permanently attached to the teat, ectothermic, immunologically naïve, and undergoing substantial physiological and neurological development (Holt *et al.*, 1981). Brain growth is accelerated, and the kidneys and lymphatic tissues develop, including the cervical and thoracic thymuses (Deane and Cooper, 1988; Holt *et al.*, 1981). *ELP* expression also coincides with periods of passive immune transfer of maternal IgG and other immunoglobulins to the young when the gut is permeable to macromolecules (Adamski and Demmer, 2000; Kwek *et al.*, 2009a; Old and Deane, 2000). Piotte and Grigor (1996) suggested that ELP may protect the PY against pathogens. The potent trypsin-inhibitory activity of bovine CTI and its temporal expression suggests that it may prevent proteolytic degradation of immunoglobulins transferred to the young via colostrum, allowing them to be absorbed intact through the intestines (Baintner, 2007; Laskowski and Laskowski, 1951; Pineiro *et al.*, 1978)

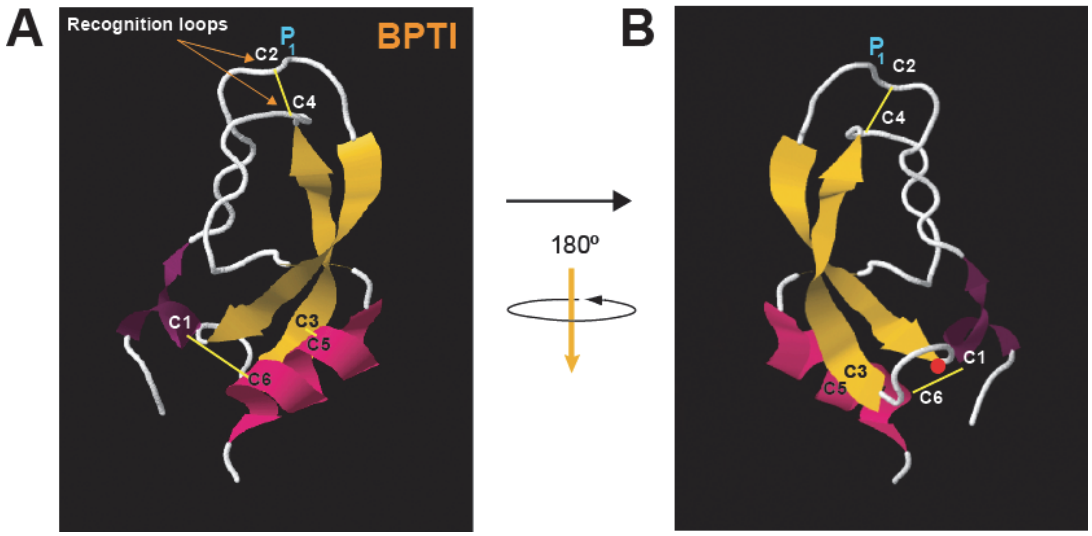


Figure 1.6. Ribbon representation of the 3D structure of BPTI

A. BPTI contains two α-helices (purple and pink), two anti-parallel β-sheets (yellow arrows) and non-structured coils (white lines). The three disulphide bonds which stabilize the globular protein (C1-C6, C2-C4 and C3-C5) are indicated (yellow lines). The P_1 reactive site residue (Lys) is adjacent to C2 at the 'top' of the protein. [Protein Data Bank (PDB) model 1PIT] (Berndt *et al.*, 1992). **B.** A 180º rotation of BPTI. While BPTI is not glycosylated, the analogous location of the bovine CTI N-glycosylation site (Asn42) at the base of the protein is noted (red dot).

The Kunitz-BPTI (I2) Family of Proteases Inhibitors: The Kunitz domain evolved from a single ancestral gene over 500 MYA (Ikeo *et al.*, 1992) and is now ubiquitous in mammals, reptiles, birds, plants, insects, viruses and bacteria (Rawlings *et al.*, 2004). The Kunitz domain is characteristic to inhibitors of the S1 Chymotrypsin superfamily of serine endopeptidases (Laskowski and Kato, 1980; Rawlings *et al.*, 2014). Kunitz-BPTI family members may have one or more Kunitz serine protease inhibitor domains that are characterized by two different motifs (Figure 1.5). The BPTI KUNITZ 2 motif (Prosite: PS50279), spans the entire 51 amino acid domain and has six conserved cysteine residues which form three disulphide bonds: C1-C6, C2-C4 and C3-C5 (Kataoka *et al.*, 2002). For ELP and CTI the three disulphide bonds occur between Cys23-Cys73, Cys32-Cys48 and Cys40-Cys69, respectively (alignment numbering, Figure 1.5). The disulphide bonds produce a stable, globular protein as for the archetypal BPTI Kunitz domain (Figure 1.6) (Wagner *et al.*, 1978). The second,

shorter BPTI KUNITZ 1 motif: F-x(2)-{I}-G-C-x(6)-[FY]-x(5)-C; residues within square brackets are permitted, but those within curly brackets are not and x represents any residue (Prosite: PS00280); is located within the BPTI KUNITZ 2 motif. Interestingly, dunnart and possum ELP have one unpaired cysteine residue (Cys68), adjacent to the paired Cys69 (alignment numbering, Figure 1.5). An unpaired cysteine is generally rare in extracellular proteins (Carter and Ho, 1994), but does occur in some proteases, including tissue-type plasminogen activator (Sehl *et al.*, 1996). Notably, functional studies showed no difference in protease activity of the wild-type protein (unpaired sulphydryl) versus a mutant protein (unpaired cysteine residue mutated to alanine) (Sehl *et al.*, 1996).

While some Kunitz inhibitors have only one domain, e.g. BPTI, CTI, ELP (Kunitz and Northrop, 1936), hookworm *Ancylostoma caninum*-Kunitz protease inhibitor-1 (Ac-KPI-1) has 12 inhibitor domains (Hawdon *et al.*, 2003). Two domain Kunitz inhibitors include serine peptidase inhibitor, Kunitz type 1 and 2 (SPINT1 and SPINT2), also known as hepatocyte growth factor activator inhibitor 1 and 2 (HAI1 and HAI2), as well as bikunin. In contrast, tissue factor pathway inhibitor 1 and 2 (TFPI1 and TFPI2) have three inhibitor domains, most likely due to domain and then gene duplication (Ikeo *et al.*, 1992; Rawlings *et al.*, 2004). The Kunitz domain has also been co-opted into many proteins to form multi-type inhibitors. Serine peptidase inhibitor-like, with Kunitz and WAP domains 1 (SPINLW1), also known as eppin, contains a WAP and Kunitz domain. WAP, Kazal, immunoglobulin, Kunitz and netrin domain-containing protein 1 and 2 (WFIKKN1 and WFIKKN2) also contains multiple domains, the possible effect of exon shuffling (Ikeo *et al.*, 1992).

$$\mathbf{E + I \rightleftharpoons L \rightleftharpoons C \rightleftharpoons X \rightleftharpoons L^* \rightleftharpoons E + I^*}$$

Figure 1.7. The 'standard' (or 'Laskowski') reaction mechanism characteristic to the Kunitz inhibitor family

E is the enzyme; I and I* are the virgin and modified inhibitors, respectively; L and L* are loose, non-covalent (rapidly dissociable) complexes of E with I and I*, respectively; X is the relatively long-lived intermediate in the E + I* reaction and C is the stable enzyme-inhibitor complex (Laskowski and Kato, 1980).

Kunitz inhibitors prevent proteolysis by forming reversible, 1:1 tight-binding complexes with a serine protease using the 'standard' (or 'Laskowski') mechanism (Figure 1.7) (Laskowski and Kato, 1980). Within the BPTI-Kunitz domain there are two putative recognition loops which interact with ~10-15 residues of the serine protease (Laskowski, 1986; Xu *et al.*, 1998). The convex 'binding' loop of the inhibitor docks with the catalytic cleft of the serine protease in a classical 'lock and key' interaction (Grzesiak *et al.*, 2000). The P_1 reactive site within the inhibitor-binding loop (Figure 1.6) is the main site of recognition and interaction between enzyme and inhibitor and determines enzyme-inhibitor specificity (Berger and Schechter, 1970; Jering and Tschesche, 1976; Wenzel and Tschesche, 1981).

The inhibition of trypsin by BPTI involves the nucleophilic attack of Ser195 of trypsin on the carbonyl carbon of the P_1 residue of the inhibitor (Figure 1.8) (Zakharova *et al.*, 2009). This produces a covalent acyl-enzyme intermediate and an amino terminus on the P_1' residue. The peptide bond formed between P_1(inhibitor)-Ser(enzyme) produces a stable complex which inactivates the protease (Kataoka *et al.*, 2002; Laskowski, 1986; Laskowski and Qasim, 2000; Zakharova *et al.*, 2009). The reaction is reversed by the nucleophilic attack of a water molecule on the carbonyl group of the acyl-enzyme (Zakharova *et al.*, 2009). Hence, a carboxyl group is attached to the P_1 inhibitor residue and Ser195 (enzyme) dissociates from the complex. The P_1-P_1' peptide bond in the inhibitor is then reformed.

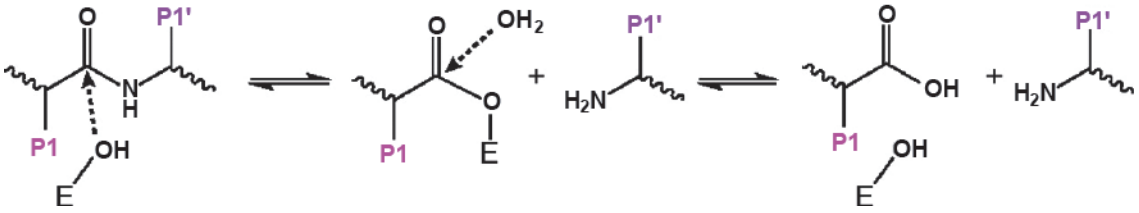


Figure 1.8. The two-step mechanism of serine protease inhibition

The carbonyl carbon of the P_1 residue of the Kunitz inhibitor undergoes nucleophilic attack by Ser195 of trypsin. The P_1-P_1' peptide bond is cleaved. A covalent acyl-enzyme intermediate is formed and an amino terminus added to the P_1' residue. Nucleophilic attack of a water molecule on the carbonyl group of the acyl-enzyme complex reverses protease inhibition. The Ser195 residue of trypsin dissociates from the inhibitor, with a carboxyl group attached to the P_1 residue of the inhibitor. The P_1-P_1' peptide bond in the inhibitor is then reformed. From (Zakharova *et al.*, 2009).

The formation of the Kunitz inhibitor-S1 protease complex also involves interactions between protease and inhibitor (Zakharova *et al.*, 2009). For the protease, these include a 'catalytic triad' of residues: histidine, aspartate and serine, e.g. His57 and Asp102 and Ser195 for trypsin (Figure 1.9). The strength of protease 'inhibition' is determined by van der Waals forces and hydrogen bond interactions between 10-17 residues of the inhibitor and 17-29 residues of the protease (Laskowski, 1986; Laskowski and Kato, 1980; Laskowski and Qasim, 2000). Reactions kinetics are also important. Neutral pH conditions favour enzyme-inhibitor complex formation, but acidic conditions cause the dissociation of the inhibitor and protease (Laskowski and Kato, 1980; Laskowski and Qasim, 2000).

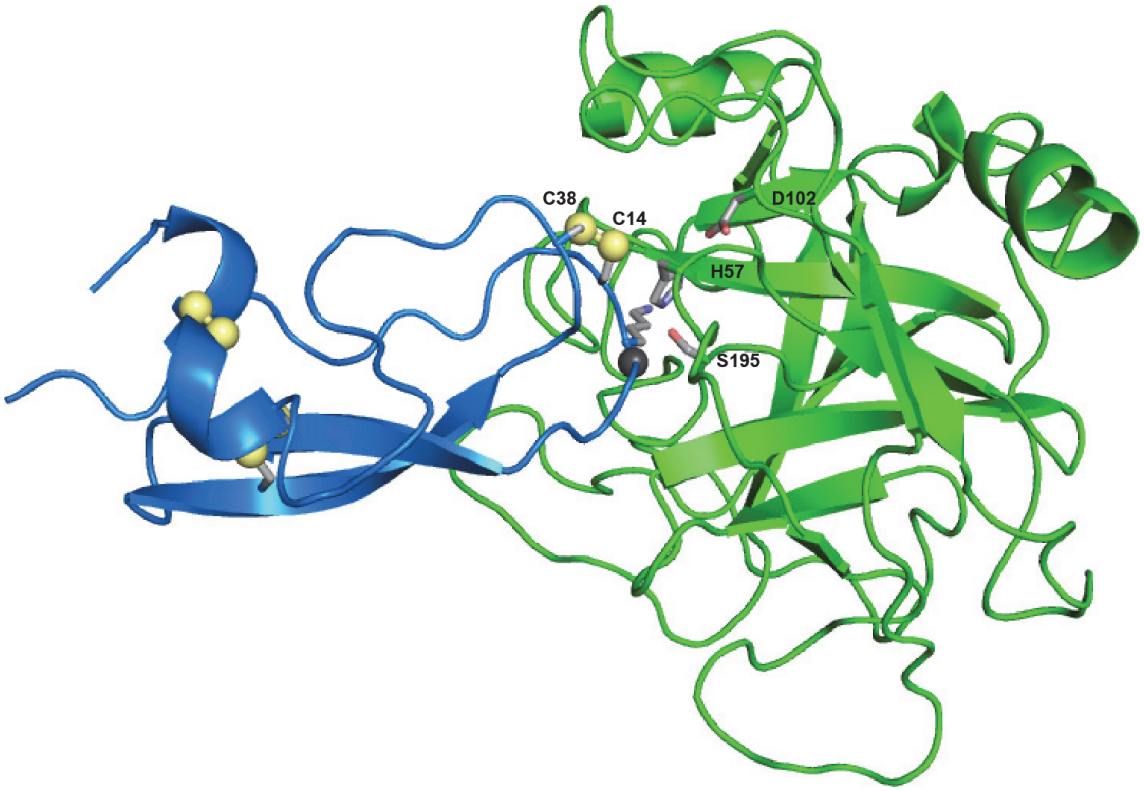


Figure 1.9. Structure of the BPTI-trypsin complex

Ribbon representation of the BPTI-rat anionic trypsin complex [PDB: 3FP6]. BPTI is shown in blue and trypsin in green. The three disulphide bonds within BPTI are shown (yellow balls and sticks). The catalytic triad of trypsin (D102, H57 and S195) are shown as sticks and the carbonyl carbon of the BPTI scissile bond as a black ball. From (Zakharova *et al.*, 2009).

There is little conservation of the P_1 reactive site residue within the ELP and CTI proteins (Figure 1.5). Like BPTI, the reactive site of bovine CTI is a lysine residue (Lys33) (Cechova and Muszynska, 1970; Muszynska and Cechova, 1970). This is not so for marsupial ELP. The P_1 residue for dunnart ELP is an asparagine residue (Asn33), while brushtail possum and opossum ELP P_1 is aspartic acid (Asp33), but the tammar P_1 residue is serine (Ser33). Whilst Kunitz inhibitors with Lys or Arg at P_1 inhibit trypsin or trypsin-like enzymes (Ikeo *et al.*, 1992), those with alanine (Ala) or Ser in this position inhibit elastase-like enzymes (Laskowski and Kato, 1980). However, P_1 may not be the biologically active site of ELP and CTI. Like possum and opossum ELP, tick anti-coagulant peptide (TAP) has a P_1 Asp residue is a Kunitz inhibitor and prevents Factor Xa activity (Jordan *et al*., 1992). However, the TAP reactive site occurs at the N-terminal end of the protein. Whether ELP and/or CTI act as serine protease inhibitors *in vivo* is not known. Furthermore, neither the reactive site(s) of ELP or CTI, nor the ligand(s) they may interact with, have been identified.

Whey Acidic Protein. The evolutionary diversity in mammalian whey proteins is exemplified by WAP, a cysteine-rich putative protease inhibitor with a variable number of WAP-type 'four-disulfide core' (WFDC/4-DSC) domains: C-{G}-{C}-[DN]-x(2)-C-x(3)-{G}-x-C–C (Prosite: PS00317) (Hennighausen *et al.*, 1982). Each WAP domain spans ~50 amino acids and is stabilised by eight cysteine residues which form four disulphide bonds (McKnight *et al.*, 1991; Simpson and Nicholas, 2002). While the eutherian and echidna WAPs (~14 kDa) have two WFDC domains (Hennighausen and Sippel, 1982; Sharp *et al.*, 2007; Simpson *et al.*, 2000), marsupial and platypus WAP have a third WFDC domain and are therefore larger (~24-30 kDa) (Simpson *et al.*, 2000). Furthermore, not all mammals have WAP, with a *WAP* pseudogene only in ruminants and humans (Hajjoubi *et al.*, 2006; Rival-Gervier *et al.*, 2003).

The function of milk WAP is unknown (Groenen and van der Poel, 1994; Ranganathan *et al.*, 1999; Vilotte *et al.*, 2003). However, eutherian WAP has anti-proliferative and anti-tumourigenic activity (Ikeda *et al.*, 2004; Nukumi *et al.*, 2007a, b). In contrast, marsupial milk WAP has the opposite effect (Topcic *et al.*, 2009). Tammar WAP, specifically WFDC domain III stimulates MEC proliferation *in vitro*. While this domain has been lost to eutherians, it has been retained by marsupials and monotremes (Sharp *et al.*, 2007; Topcic *et al.*, 2009). This difference in mammalian WAP activity may explain the differences in WAP expression patterns. While eutherian *WAP* is induced in late-pregnancy and expressed throughout lactation (Devinoy *et al.*, 1988; Hennighausen and Sippel, 1982; Piletz *et al.*, 1981; Simpson *et al.*, 1998b),

marsupial *WAP* is only expressed during mid-lactation (Phase 2B) (De Leo *et al.*, 2006; Demmer *et al.*, 2001; Nicholas *et al.*, 2001; Simpson *et al.*, 2000). Hence, tammar WAP may facilitate the increased Phase 2B mammary development and milk production (Topcic *et al.*, 2009). Alternatively, WAP may prevent the proteolytic degradation of immunoglobulins in milk (Demmer *et al.*, 2001), or the high cysteine and hence sulphur content of WAP may contribute to the Phase 2B development of PY fur (Simpson *et al.*, 2000).

α-Lactalbumin and Lysozyme. Although they evolved from a common ancestral gene, α-LA, the c-type lysozyme (LYZ) and calcium-binding lysozymes (LYSC1) have different functions (Irwin *et al.*, 2011; Messer *et al.*, 1998; Qasba and Kumar, 1997). α-LA and LYSC1 bind calcium (Brew, 2013; Sordillo and Streicher, 2002), but unlike LYZ, have lost their antibacterial function, i.e. the ability to breakdown bacterial cell walls (Jolles *et al.*, 1979; Jolles and Jolles, 1984). Furthermore, while α-LA is a mammary gland-specific milk protein (Nicholas *et al.*, 1989; Shaw *et al.*, 1993; Shewale *et al.*, 1984), LYZs are present in numerous tissues and secretions: saliva, tears, mucous, milk, egg white, etc (Irwin *et al.*, 2011; Jolles *et al.*, 1979; Piotte *et al.*, 1997; Teahan *et al.*, 1991). In contrast, LYSC1 is only present in selected species including the echidna, horse, donkey and dog and pigeon egg white (Irwin *et al.*, 2011).

α-LA is best known for its essential role in lactose synthesis (Groenen and van der Poel, 1994; Stuart *et al.*, 1986). It forms the lactose synthase complex which catalyses the transfer of galactose from UDP-galactose to glucose to form lactose in the MEC Golgi (Brew, 2003; Brodbeck *et al.*, 1967; Qasba and Kumar, 1997). However, in the absence of α-LA, β4GALT1 transfers galactose to N-acetylglucosaminyl (GlcNAc) producing long-chain complex sugars (Brew, 2003). As mentioned previously, otariid and odobeniid milk contains no lactose (Oftedal *et al.*, 1987) due to a mutations within the *LALBA* gene (Reich, 2009; Reich and Arnould, 2007; Sharp *et al.*, 2008).

The Milk Lipocalins: β-Lactoglobulin and Late Lactation Proteins A and B. While *LGB* is present in most mammalian milks (Joss *et al.*, 2009; Sawyer, 2013), additional lipocalin-encoding milk protein genes such *LLPA*, *LLPB* and *trichosurin* have evolved in the tammar (Collet and Joseph, 1993; Collet *et al.*, 1991; Trott *et al.*, 2002; Woodlee *et al.*, 1993). The lipocalins are a large group of small, secreted proteins with diverse functions (Flower, 1996). Lipocalins have a highly-conserved central barrel-shaped hydrophobic pocket, inside which small hydrophobic molecules are bound (Åkerström *et al.*, 2000; Flower *et al.*, 2000; Kontopidis *et al.*, 2004; Pervaiz and Brew, 1987).

Hence, they may be involved in the transport of milk fatty acids, retinol (vitamin A) and pheromones as well as the regulation of immune and inflammatory responses, olfaction, prostaglandin synthesis, invertebrate cryptic coloration and cell homeostasis (Flo *et al.*, 2004; Flower *et al.*, 2000; Godovac-Zimmermann *et al.*, 1985; Logdberg and Wester, 2000).

Although β-LG, is the major whey protein of ruminant, marsupial and monotreme milk (Davies *et al.*, 1983; Hambling *et al.*, 1992; Joss *et al.*, 2009; Sawyer, 2013), it is absent from human and rodent milk (Hambraeus and Lonnerdal, 2003; Kontopidis *et al.*, 2004). Marsupials have only one copy of *LGB*, but gene duplication has occurred in eutherians such as the dog, cow, donkey and seal, which have both *LGB1* and *LGB2* (Cane *et al.*, 2005; Godovac-Zimmermann *et al.*, 1988; Godovac-Zimmermann *et al.*, 1990), while the cat has three *LGB* genes (Halliday *et al.*, 1990; Halliday *et al.*, 1991; Pena *et al.*, 1999). As for eutherians, tammar *LGB* is induced at parturition and expressed throughout lactation (Figure 1.4). However, there is a Phase 3 increase in tammar *LGB* expression (Nicholas *et al.*, 1997; Nicholas *et al.*, 1994), which coincides with the expression of the marsupial-specific *LLPA* and *LLPB* genes (the latter of which is only present in the tammar) (Piotte *et al.*, 1998; Topcic, 2010; Trott, 1999; Trott *et al.*, 2002; Woodlee *et al.*, 1993). β-LG can bind retinol, retinoic acid and cationic lipids (Belatik *et al.*, 2012; Bourassa *et al.*, 2014; Hasni *et al.*, 2011; Khorsand Ahmadi *et al.*, 2015). Hence, the late lactation increase in milk lipocalins may facilitate the uptake of retinol released by lipolysis in the PY intestines (Collet *et al.*, 1991).

Caseins

The evolution of the calcium-sensitive caseins ~310-220 MYA most likely led to the evolution of lactation in ancestral mammals (Brawand *et al.*, 2008; Kawasaki and Weiss, 2003; Smyth *et al.*, 2004). These caseins are members of the secretory *calcium-binding phosphoprotein* (*SCPP*) gene family. The milk caseins are one of the most rapidly diverging protein families (Bawden and Nicholas, 1999; Vilotte *et al.*, 2013). Most mammals have at least three casein genes: the calcium-sensitive *α*-and *β-casein* (*CSN1, CSN2*) plus the calcium-insensitive *κ-casein* (*CSN3*) (Martin *et al.*, 2013; Mercier and Grosclaude, 1999; Rijnkels, 2002). *CSN1* and *CSN2* presumably evolved from a common ancestral gene (*SCPP-Pro-Gln-rich 1* (*SCPPPQ1*)) (Kawasaki *et al.*, 2011), due to tandem gene duplication, exon shuffling and sequence divergence (Bonsing and Mackinlay, 1987; Mercier and Grosclaude, 1999; Mercier and Vilotte, 1993; Rijnkels, 2002). For many years, *CSN3* was thought to have evolved from the fibrinogen (blood clotting) gene family (Fiat and Jolles, 1989; Jolles *et al.*, 1978;

Thompson *et al.*, 1985). However, Kawasaki and colleagues (2011) showed that *CSN3* probably originated from the *follicular dendritic cell secreted peptide* (*FDCSP*) gene, another member of the *SCPP* gene family. Moreover, they found that *odontogenic ameloblast–associated* (*ODAM*) gene is probably the ancestral casein gene (Kawasaki *et al.*, 2011).

In milk, the calcium-sensitive caseins form colloidal aggregates known as micelles (Dalgleish, 1998; Holt, 1998; Horne *et al.*, 2007; Smyth *et al.*, 2004). These random coil structures contain water-filled cavities and channels and several hundred high-density, colloidal calcium phosphate nanoclusters (Trejo *et al.*, 2011), which consist mainly of calcium, magnesium, phosphate and citrate (Fox *et al.*, 2015) and are stabilised by κ-casein (Creamer *et al.*, 1998; Hambraeus and Lonnerdal, 2003). Most importantly, the nature of the micelle structure allows milk to be super-saturated with calcium phosphate (Trejo *et al.*, 2011). They also regulate milk viscosity by preventing the precipitation of α- and β-casein (Davies *et al.*, 1983; Hambraeus and Lonnerdal, 2003; Lenton *et al.*, 2015; Sawyer and Holt, 1993). Interestingly, milk casein phosphopeptides have antimicrobial, anticariogenic, and immunomodulatory properties (Cochrane and Reynolds, 2012; Naqvi *et al.*, 2015).

Milk casein genes are developmentally-regulated in mammals. They are expressed during the secretory differentiation phase from mid-pregnancy, stimulated at parturition and expressed at high levels throughout lactation (Bird *et al.*, 1994; Demmer *et al.*, 1998; Nicholas *et al.*, 1997; Rosen *et al.*, 1980; Rosen *et al.*, 1999).

Other Milk Constituents

Milk Fat Globules

Milk fat globules (MFGs) also play a vital role in milk. MFGs are formed from triacylglycerides (TAG) and phospholipids that are released into the MEC cytoplasm as micro-lipid droplets (MLD). These fuse to form cytoplasmic lipid droplets (CLDs), which are expelled into the alveolar lumen as milk fat globules (MFGs) (Mather and Keenan, 1998; Rudolph *et al.*, 2007). MFGs are surrounded by a lipid bilayer known as the milk-fat globule membrane (MFGM). The outer layer of the MFGM is derived from the apical (luminal) membrane of the mammary epithelial cell (MEC) and contains numerous attached proteins and lipids (Mather and Keenan, 1975; Davies *et al.*, 1983; Jenness, 1986; Mather and Keenan, 1998; McManaman *et al.*, 2006; Reinhardt and Lippolis, 2006). Most importantly, MFGs allow the fat to remain in milk without coalescing, and also provide a high-energy source for the neonate (Evers, 2004; Aoki,

2006; Cavaletto *et al.*, 2008).

Like the major milk constituents, the minor milk components also vary greatly between species. These include: salts (ions and ion complexes), e.g., K^+, Na^+, $Ca2^+$, magnesium, zinc, copper and iron, inorganic phosphate, citrate, biocarbonate and sulphate (Davies *et al.*, 1983; Green, 1984; Jenness, 1974). Milk also contains water soluble vitamins, e.g. B6 and B12 (Jenness, 1986), plus the fat-soluble vitamins A, D, E and K, as well as trace elements, alcohols, aldehydes, gases, etc (see reviews) (Davies *et al.*, 1983; Jenness, 1974).

The Regulation of Milk Protein Gene Expression

Milk protein genes are expressed in MECs. Their regulation is complex and influenced by hormones, transcription factors, signals from the extracellular matrix (ECM), local and global epigenetic mechanisms and non-coding RNAs (Modepalli *et al.*, 2014; Qian and Zhao, 2014; Rijnkels *et al.*, 2013; Rosen *et al.*, 1999; Topper and Freeman, 1980). As for eutherians, tammar *LALBA*, *LGB*, *CSN1* and *CSN2* are up-regulated at parturition and expressed for the remainder of lactation (Figure 1.4). However, the asynchronous expression of many of the major tammar whey proteins such as *ELP*, *WAP*, *LLPA* and *LLPB*, provides intriguing questions as to their regulation.

Endocrine Regulation of Mammary Gland-Specific Milk Protein Gene Expression

The hormone requirements for milk protein gene transcription varies between genes (Forsyth, 1986; Rosen *et al.*, 1999) and may differ for the same gene in different species (Houdebine, 1999; Topper and Freeman, 1980; Vonderhaar and Ziska, 1989). During lactation, the maximal expression of *CSN1*, *CSN2* and *WAP* requires the synergistic interaction of the "lactogenic" hormones: insulin, glucocorticoids and prolactin (Devinoy *et al.*, 1988; Puissant and Houdebine, 1991; Rosen *et al.*, 1999). For many mammals, the surge of prolactin, increase in glucocorticoids and the withdrawal of progesterone at parturition promotes milk protein gene expression and milk secretion (Houdebine, 1999; Shamay *et al.*, 1987; Topper and Freeman, 1980). In ruminants, GH acts synergistically with lactogenic hormones to maximise milk production (Neville *et al.*, 2002; Sejrsen *et al.*, 1999; Tucker, 1985). Thyroid hormones are also essential for lactation, but like GH, their precise effect(s) on milk protein gene transcription are unknown (Burgoyne and Wilde, 1994; Linzell *et al.*, 1975; Neville *et al.*, 2002).

Prolactin. Prolactin is a pleiotropic, polypetide hormone best known for its role in the mammary gland (Horseman, 1999; Kelly *et al.*, 2002). Prolactin induces growth and differentiation during puberty, pregnancy and lactation and most importantly, milk protein synthesis and milk secretion (Groner and Gouilleux, 1995; Oakes *et al.*, 2008a; Rosen *et al.*, 1999). Prolactin is also essential for the production of milk lipids and lactose and increases the half-life of mRNA transcripts (Guyette *et al.*, 1979; Kwek *et al.*, 2007; Palmiter, 1969). Prolactin is secreted primarily by the anterior pituitary gland but also by the skin, sweat and lacrimal glands, brain, uterus and the mammary gland (Chen *et al.*, 2012; Freeman *et al.*, 2000; Goffin *et al.*, 2002). Hence it acts both as a "classic" hormone, where it is secreted and transported to its target organ(s) and locally, in an autocrine and/or paracrine manner (Brisken *et al.*, 1999; Manhes *et al.*, 2005; Shiu and Friesen, 1980).

For eutherians, prolactin levels are generally low during pregnancy, but highest at parturition and in colostrum due to a transient prolactin pulse (Neville *et al.*, 2002), promoting milk production. Prolactin then declines throughout lactation (Houdebine, 1999; Neville *et al.*, 2002). The profiles of maternal plasma prolactin are generally similar in eutherians and marsupials until late lactation when they change dramatically (Hinds, 1988; Neville *et al.*, 2002). During early and mid-lactation, tammar milk production is low, PY growth is slow and maternal prolactin levels are similar to those in the pregnant tammar (Hinds and Tyndale-Biscoe, 1982, 1985; Renfree, 1994). However, during Phase 3, maternal plasma prolactin increase significantly from 20-40 ng/mL to more than 100 ng/mL (Findlay and Renfree, 1984; Hinds, 1988; Hinds and Tyndale-Biscoe, 1982; Nicholas, 1988a). Prolactin levels remain high until ~day 250-270 of lactation when the PY permanently exits the pouch (Hinds, 1988; Renfree, 1994). However, the sucking stimulus is necessary to maintain high prolactin levels (Hinds, 1988; Renfree, 1994) and so when the young is weaned, prolactin levels decline precipitously (Hinds, 1988).

Interestingly, although there is a transient prolactin pulse at parturition (~130 ng/mL) in the maternal tammar (Hinds and Tyndale-Biscoe, 1985; Renfree, 1994; Tyndale-Biscoe *et al.*, 1984) prolactin is not essential for lactogenesis (Renfree, 1994; Tyndale-Biscoe *et al.*, 1984). Milk secretion can be induced in non-pregnant tammars by sucking stimulus alone (Renfree, 1994; Tyndale-Biscoe *et al.*, 1984).

Prolactin Receptor. In the mammary gland, prolactin acts via the PRLR, a single-chain transmembrane protein of the Class 1 hematopoetin/cytokine receptor superfamily

(Bole-Feysot *et al.*, 1998; Goffin *et al.*, 2002). PRLR signalling occurs via the Janus Kinase 2-signal transducer and activator of transcription (JAK-STAT) pathway; the prototype signalling pathway used by all cytokine/hematopoietin receptors (Darnell, 1996; Levy and Darnell, 2002; O'Shea *et al.*, 2015; Yu-Lee, 1997). There are four JAKS (JAK1, JAK2, JAK3, TYK2) and seven mammalian STATs (STAT1, STAT2, STAT3, STAT3, STAT4, STAT5A, STAT5B, STAT6) which activate diverse cytokines, growth factors and interferons, highlighting the interactive promiscuity of JAK-STAT signalling (Murray, 2007; O'Shea *et al.*, 2015; Villarino *et al.*, 2015).

PRL signalling via the PRLR is crucial for mammary development, transcription of milk protein genes and the inhibition of involution (Binart *et al.*, 2003; Gouilleux *et al.*, 1994; Hynes *et al.*, 1997; Yu-Lee *et al.*, 1998). The PRLR has three domains: the transmembrane, intracellular, and extracellular ligand-binding domains (Kelly *et al.*, 1989). A soluble prolactin-binding protein identical to the extracellular domain of the receptor also exists (Goffin and Kelly, 1996). PRLR activation (dimerisation) occurs via the binding of PRL to the extracellular receptor of the PRLR, which are located in the MEC extracellular membrane (Figure 1.10) (Jahn *et al.*, 1997; Rui *et al.*, 1992). One molecule of JAK2 binds to each PRLR intracellular domain (Rui *et al.*, 1994). Tyrosine residues within both JAK2 and the PRLR cytoplasmic domain are phosphorylated (Clevenger, 2004; Lebrun *et al.*, 1994; Rui *et al.*, 1992). This creates docking sites on the receptor to which proteins with phosphotyrosine-binding Src Homology 2 (SH2) domains (for example, activated STAT5A transcription factors) can bind (Leonard and O'Shea, 1998).

Marsupials such as the tammar and brushtail possum have one PRLR (Demmer, 1999; Gehring, 2005), but many eutherians have multiple PRLR isoforms due to alternative splicing of the *PRLR* gene (Binart *et al.*, 2010; Davis and Linzer, 1989; Hu *et al.*, 1998; Trott *et al.*, 2004). This may provide an additional level of control of cell proliferation and differentiation in the eutherian mammary gland. Both rodents and ruminants have two PRLR isoforms; the short $PRLR_S$, and long $PRLR_L$ (Bignon *et al.*, 1997; Edery *et al.*, 1989; Kelly *et al.*, 1991). An intermediate $PRLR_I$ isoform is also present in rat T lymphoma Nb2 cells, (Ali *et al.*, 1992; Ali *et al.*, 1991).

While the extracellular and transmembrane domains are conserved, the cytoplasmic tail differs between isoforms (Goffin *et al.*, 2002; Yu-Lee, 1997). As the PRLR cytoplasmic domain includes binding sites for STATs 1, 3, 5A and 5B, insulin receptor substrates and adapter proteins, any changes in the cytoplasmic domain affect signal

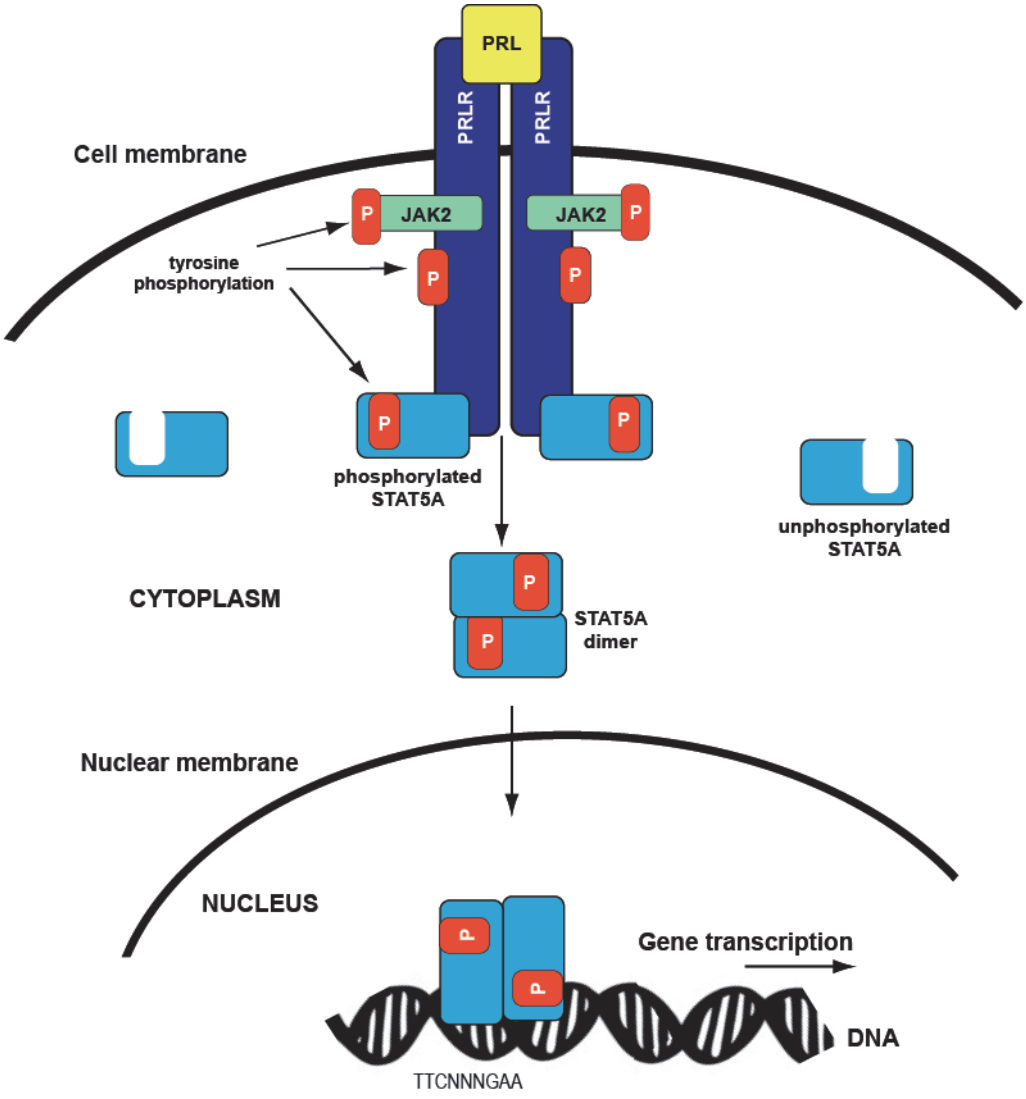


Figure 1.10. Prolactin receptor signaling in the mammary gland

PRLR signaling in the mammary gland occurs via the JAK/STAT pathway (Jahn *et al.*, 1997; Yu-Lee, 1997). PRLR activation is triggered by the binding of the prolactin ligand to two PRLR molecules, inducing receptor dimerisation. Two molecules of JAK2 then bind to the PRLR dimer. This results in the rapid phosphorylation of selected tyrosine resides in JAK2, the cytoplasmic domains of the PRLRs and STAT5A (Lebrun *et al.*, 1994) (Clevenger, 2004; Goffin and Kelly, 1996 Rui, 1992 #4282; Lebrun *et al.*, 1994; Rui *et al.*, 1994). STAT5A homodimers are translocated to the nucleus, where they bind to their target sequences in milk protein gene promoters (GAS elements: TTCNNNGAA), thereby activating gene transcription (Yu-Lee, 1997).

transduction and function (Bole-Feysot *et al.*, 1998; Goffin and Kelly, 1996; Gouilleux *et al.*, 1994; Hynes *et al.*, 1997). While the PRLR membrane-proximal cytoplasmic domain affects proliferation via box 1 (the proline-rich motif (PRM)) and box 2, the PRLR membrane-distal cytoplasmic region affects differentiation (O'Neal and Yu-Lee, 1993; Yu-Lee, 1997).

The rodent $PRLR_L$ has both the membrane proximal (proliferation) and distal regions (differentiation), but the $PRLR_S$ has box 1 only (proliferation) (Hynes *et al.*, 1997). Therefore, only $PRLR_L$ and $PRLR_I$ can induce the transcription of milk protein genes including *LGB* and *CSN2* (Ali *et al.*, 1992; Berlanga *et al.*, 1997; Edery *et al.*, 1989; Lesueur *et al.*, 1991). The types and ratio of PRLR isoforms present regulates the shift between growth and differentiation in the rodent mammary gland (Cassy *et al.*, 1998; Gallego *et al.*, 2001; Goffin *et al.*, 2002; Ormandy *et al.*, 1997a). If $PRLR_S$ is most abundant, milk protein gene expression is inhibited by 'inactive' $PRLR_S$-$PRLR_S$ homodimers and $PRLR_S$-$PRLR_L$ heterodimers (Berlanga *et al.*, 1997; Perrot-Applanat *et al.*, 1997; Saunier *et al.*, 2003). However, if $PRLR_L$ is dominant, milk protein production is favoured (Edery *et al.*, 1989; Lesueur *et al.*, 1991; Perrot-Applanat *et al.*, 1997).

In contrast, for marsupials such as the tammar, the control of mammary proliferation and differentiation via the PRLR may be dependent upon changes in receptor concentration and binding-affinity throughout the reproductive cycle (Hinds, 1988; Hinds and Tyndale-Biscoe, 1985; Sernia and Tyndale-Biscoe, 1979; Tyndale-Biscoe and Hinds, 1984). This may facilitate the concurrent asynchronous lactation (CAL) phenomenon (refer to the 'Oxytocin' section below).

Notably, apart from the JAK-STAT pathway, prolactin also activates the phosphatidylinositol 3-kinase/protein kinase B (PI3K/AKT) and the RAS/RAF/mitogen activated kinase (MAPK) pathways through its synergistic interaction with progesterone. For the PI3K/AKT pathway, PRL activates receptor activator of nuclear factor-κB ligand (RANKL, a member of the tumour necrosis factor (TNF) family) through STAT5A (Chen *et al.*, 2012; Das and Vonderhaar, 1995; Molina and Adjei, 2006; Song *et al.*, 2005). The binding of RANKL to the RANK receptor triggers a signalling cascade which activates NF-κB, MAPKs and AKT, thus promoting MEC proliferation (Chen *et al.*, 2012; Horseman and Gregerson, 2014). Notably, this pathway only occurs in 'hormone-sensing' MECs, and not in 'secretory' MECs (Horseman and Gregerson, 2014).

Growth Hormone. The evolution of PRL, GH, the PRLR and the growth hormone receptor (GHR) from a common ancestral gene over 400 MYA may explain the cross-talk between these signalling pathways (Gallego *et al.*, 2001; Niall *et al.*, 1971) (Goffin *et al.*, 1996). While PRL acts only via the mammary epithelium (Brisken *et al.*, 1999), GH acts on both mammary parenchyma and stroma, but preferentially in the stroma (Gallego *et al.*, 2001). Interestingly, eutherians have an additional PRL/GH orthologue, placental lactogen (PL) which may bind to the PRLR and/or GHR (Goffin *et al.*, 1996).

Insulin. Insulin is essential for mammary differentiation, lactation and the expression of milk protein genes (Bolander *et al.*, 1981; Menzies *et al.*, 2010; Sheehy *et al.*, 2000; Topper and Freeman, 1980). Insulin may act through the insulin, IGF-I and/or IGF-II ligands, by binding to the insulin, IGF-I, or IGF-II receptors, respectively. However, cross-talk between insulin ligands and receptors occurs (Baumrucker and Erondu, 2000; Withers and White, 2000). Like prolactin, insulin ligands may activate the PI3K/AKT and/or RAS/RAF/MAPK pathways (Bevan, 2001; LeRoith, 2000; Lizcano and Alessi, 2002). However, further investigation is required to determine the precise actions of insulin in the mammary gland and their effects on milk protein gene expression.

Glucocorticoids. The adrenal cortex secretes glucocorticoids in response to adrenocorticotrophic hormone release from the anterior pituitary (Tucker, 1985). Glucocorticoid signalling occurs through the binding of the ligand to its receptor (Schoneveld *et al.*, 2004). For eutherians, plasma glucocorticoids have a similar profile to prolactin and change little throughout lactation apart from a transient spike at parturition (Houdebine, 1999; Schwalm and Tucker, 1978; Tucker, 1985). In contrast, there is minimal change in the maternal tammar plasma profile during the reproductive cycle (Janssens and Hinds, 1981). Although cortisol is essential for the birth of the tammar PY (Shaw *et al.*, 1996), an increase in PY rather than maternal cortisol induces birth by promoting prostaglandin secretion (Renfree *et al.*, 1994). The main effects of glucocorticoids on the mammary gland during lactation include the enhanced transcription of both casein and whey protein genes, partly due to the stabilisation of mRNA transcripts (Doppler *et al.*, 1990; Nagaiah *et al.*, 1981; Rosen *et al.*, 1998). They also stimulate anti-inflammatory responses and suppress both cell-mediated and humoral immunity (Amsterdam *et al.*, 2002; Waller, 2000). This may prevent inappropriate immune responses in the neonate during the periods of colostrum production and early lactation.

Thyroid Hormones. Thyroid hormones are essential for maximal milk protein

secretion and milk production (Houdebine *et al.*, 1978; Tucker, 1985). During lactation, thyroid hormones regulate the partitioning of metabolic activity by controlling the conversion of T_4 to the more-active T_3. Activity in peripheral tissues is reduced, while activity in the mammary gland is increased, thereby directing maternal resources; i.e. milk, to the young (Jack *et al.*, 1994; Song and Oka, 2003). In ruminants, thyroid hormones enhance lobuloalveolar growth and the galactopoietic response of the mammary gland to GH and PRL (Capuco *et al.*, 1999; Cowie, 1969; Forsyth, 1986). The combination of T_3 and lactogenic hormones more than doubles *LALBA* expression in the murine mammary gland (Ziska *et al.*, 1988); while the IHCPRLT_3E_2-treatment of Phase 1 tammar mammary gland explants induces expression of the Phase 2B-specific *WAP* gene (Simpson *et al.*, 1998a; Simpson *et al.*, 2000). Furthermore, thyroid hormones in milk are critical to the survival of the tammar PY which lacks an active thyroid gland and is therefore incapable of thermoregulation until ~ 180 days pp (Gemmell and Sernia, 1992; Richardson *et al.*, 2002).

Oxytocin. For marsupials and eutherians, oxytocin (mesotocin is the tammar equivalent) is released from the maternal posterior pituitary gland at parturition and is essential for milk 'let-down' (Lincoln and Renfree, 1981; Tucker, 1985) and indeed for the birth of the tammar PY (Parry *et al.*, 1996; Sebastian *et al.*, 1998; Shaw and Renfree, 2006). Oxytocin does not directly affect milk protein gene transcription and is responsible only for milk let-down, milk secretion and removal by the sucking young promotes further milk synthesis under the influence of prolactin. A decrease in oxytocin-sensitivity of the mammary gland as lactation proceeds (Bathgate *et al.*, 1995; Chauvet *et al.*, 1981; Lincoln and Renfree, 1981; Sebastian *et al.*, 1998) may facilitate the CAL phenomenon (Griffiths *et al.*, 1972; Lincoln and Renfree, 1981; Nicholas, 1988a; Sharman and Calaby, 1964). This allows the marsupial mother to secrete two different types of milk at the same time from different glands; for a young at foot (Phase 3) and for an altricial newborn (Phase 2A). Hence, sucking stimulus of the neonate allows smaller volumes of milk to be ejected from the 'young' oxytocin-sensitive gland for the PY (Lincoln and Renfree, 1981). However, the young at foot has to provide greater sucking stimulus to stimulate milk let-down in the 'older' oxytocin-insensitive gland (Lincoln and Renfree, 1981).

It was originally thought that the tammar wallaby PY controlled the progression of the lactation cycle due to the correlation between changes in the sucking patterns of the PY and lactation phase changes (Simpson *et al.*, 1998a). However, macropodid marsupial foster experiments, both within the same species and between different

species have shown that the lactation cycle is developmentally-regulated and controlled by the mother (Findlay, 1982; Kwek *et al.*, 2009b; Menzies *et al.*, 2007; Trott *et al.*, 2003). For tammar PYs fostered onto 'older' glands (~60 days advanced in lactation stage), neither milk protein gene expression, nor milk protein content are altered. However, rates of both development and growth of fostered young were accelerated compared to control animals (Kwek *et al.*, 2009b; Trott *et al.*, 2003). In contrast, the fostering of older PYs onto 'younger' glands caused severe emaciation and the young were euthanised (Findlay, 1982).

Transcriptional Regulation of Milk Protein Gene Expression

Transcription of milk protein genes is a highly regulated process mediated by RNA polymerase II, transcription factors and epigenetic mechanisms and non-coding RNAs (Lemay *et al.*, 2013; Mercier and Grosclaude, 1999; Rijnkels *et al.*, 2013). However, the asynchronous, temporal expression of many marsupial milk protein genes in the absence of major hormone changes suggest that their regulation is complex. Most research has focussed on the regulation of eutherian milk protein genes (Qian and Zhao, 2014), with few studies of marsupial milk protein gene transcription (Ginger, 1998; Topcic, 2010; Trott, 1999; Trott *et al.*, 2005). Milk protein gene expression is regulated by many different transcription factors, only some of which will be explored in this review.

Signal Transducers and Activation of Transcription. STATs 1, 3, 5A and 5B are critical for mammary gland function (Hennighausen *et al.*, 1997). However, while STAT1 is involved in immune-modulation and is activated by factors in the IFN-γ signalling pathway, STAT3 is activated during mammary gland involution (Stein *et al.*, 2007; Watson and Neoh, 2008). In contrast, STAT5A is essential to the PRL-induction of milk protein genes including *LALBA*, *BLG*, *CSN1*, *CSN2* and *WAP* (Doppler *et al.*, 1991; Pierre *et al.*, 1994; Wakao *et al.*, 1992; Watson *et al.*, 1991).

STAT5A, originally known as mammary gland factor (MGF), or milk protein binding factor (MPBF), is the major transcription factor involved in PRL-PRLR signalling during lactation (Schmitt-Ney *et al.*, 1992; Wakao *et al.*, 1992; Watson *et al.*, 1991; Welte *et al.*, 1994a). Latent, cytoplasmic STAT5A docks with the tyrosine phosphorylated PRLR and is tyrosine phosphorylated (Goffin and Kelly, 1996; Gouilleux *et al.*, 1994; Hynes *et al.*, 1997; Shuai *et al.*, 1993). Activated STAT5A homodimers are rapidly translocated to the nucleus, bind to interferon gamma (IFNγ) activated sequences (GAS), i.e., TTNCNNNGAA and TTCNNNGAA motifs within

milk protein gene promoters and induce gene transcription (Darnell *et al.*, 1994; Gouilleux *et al.*, 1994; Groner *et al.*, 2000; Yu-Lee, 1997). Furthermore, histone acetylation can also regulate STAT activity (Zhuang, 2013).

Although STAT5A and STAT5B share ~96% similarity at the protein level, they have different roles in the mammary gland (Hennighausen and Robinson, 2008; Liu *et al.*, 1995; Teglund *et al.*, 1998). STAT5A is involved in PRL/PRLR-signalling and milk production and STAT5B is associated with GH/GHR-signalling and mammary gland growth and development (Akira, 1999; Davey *et al.*, 1999; Grimley *et al.*, 1999; Woelfle *et al.*, 2003). However, the presence of both activated STAT5B homodimers and STAT5A-STAT5B heterodimers in the MEC cytoplasm and the mammary phenotype of STAT5A and/or STAT5B homo- and heterozygous knockout mice indicates that cooperation and redundancy exists between STAT5A and STAT5B (Liu *et al.*, 1996; Teglund *et al.*, 1998).

Glucocorticoid Receptor (GR) and the Glucocorticoid Response Element (GRE). Glucocorticoids are involved in the transcription of the CSN1, CSN2 and WAP genes (Ganguly *et al.*, 1979; Lechner *et al.*, 1997; Welte *et al.*, 1993). The activated GR acts synergistically with STAT5A to mediate lactogenic hormone signal transduction by binding to either full GR (GGTACA$(N)_3$TGTTCT) or GRE half-sites (½GRE, TGTTCT) in milk protein gene promoters (Doppler *et al.*, 1989; Li and Rosen, 1994b; Vonderhaar and Ziska, 1989).

CCAAT Box-Binding Transcription Factor/Nuclear Factor I (CTF/NFI). NFI is implicated in the expression of *WAP*, *BLG* and *B4GALT1* (Li and Rosen, 1994b; Mukhopadhyay *et al.*, 2001; Watson *et al.*, 1991). NFI binds to the CCAAT box (TTGGC(N_{2-5})GCCAA), with DNA-binding activity in MECs influenced by MEC-ECM interactions, but independent of lactogenic hormones (Rosen *et al.*, 1998; Santoro *et al.*, 1988). In the mammary gland, the NFI consensus half-site (TTGGC) is generally used in preference to the complete consensus site (Mink *et al.*, 1992; Watson *et al.*, 1991).

Mammary Cell-Activating Factor (MAF). MAF is a member of the ETS (E26 Transformation-specific Sequence) family of transcription factors which have a GGAA/T DNA-binding core (Kolb *et al.*, 1994; Mink *et al.*, 1992; Welte *et al.*, 1994b). Although MAF regulates the mammary gland specific transcription of *WAP* and mouse mammary tumor virus (MMTV) (Kolb *et al.*, 1994; Mink *et al.*, 1992; Welte *et al.*, 1994b), its precise identity is not known. Many ETS factors, including ETS1, ETS2, ETS translocation variants 1, 4 and 5 (ETV1, 4 and 5), ETS-like factors 3 and 5 (ELF3

and ELF5) are expressed in a stage-specific manner in the mammary gland, during development, pregnancy and lactation (Galang *et al.*, 2004; Oakes *et al.*, 2008b; Zhou *et al.*, 2005). Interestingly, ELF5 is a master regulator of alveolar fate and mammary gland function (Choi *et al.*, 2009; Lee and Ormandy, 2012). It is secreted by milk-producing, but not hormone-sensing MECs (see review) (Horseman and Gregerson, 2014). Furthermore, PRL-induced ELF5 forms a positive feedback loop with STAT5A (Lee *et al.*, 2011), promoting the explosive differentiation of the mammary epithelium before lactation (Horseman and Gregerson, 2014). In contrast, the activation of ELF5 by RANKL promotes MEC proliferation rather than differentiation (Horseman and Gregerson, 2014).

ETS family members also interact with multiple binding partners to regulate genes in different types of tissues and cells (Li *et al.*, 2000; Verger and Duterque-Coquillaud, 2002). MAF and NFI act cooperatively to induce murine *WAP* expression, but only if the binding sites are within close proximity (Mink *et al.*, 1992; Welte *et al.*, 1994b). ETS family members also interact with activating protein 1 (AP1), NF-κB, nuclear factor of activated T-cells (NFAT), STAT5A and SP1 (Bassuk *et al.*, 1997; Li *et al.*, 2000). Furthermore, ETS factors can act as both transcriptional activators and repressors (Gutierrez-Hartmann *et al.*, 2007; Wasylyk *et al.*, 1993; Yordy and Muise-Helmericks, 2000) and therefore have important roles in development, differentiation, cell proliferation, cell cycle control, transformation, T-cell activation and tumour formation (Maroulakou and Bowe, 2000; Sharrocks, 2001; Shepherd and Hassell, 2001).

CCAAT/Enhancer Binding Proteins (C/EBPs). The C/EBPs are members of the leucine zipper family of sequence-specific bZIP DNA binding proteins and are associated with tissue-specific gene expression (Birkenmeier *et al.*, 1989; Gigliotti and DeWille, 1998; O'Rourke *et al.*, 1997). There are six highly conserved C/EBPs that are encoded by separate intronless genes: C/EBPα, β, δ, ε, γ and ζ (Grimm and Rosen, 2003; Landschulz *et al.*, 1988). C/EBPβ and C/EBPδ, also known as nuclear factor-interleukin-6 and -6β (NF-IL6 and NF-IL6β, respectively) and C/EBPα are involved in mammary gland proliferation and functional differentiation (Doppler *et al.*, 1995; Raught *et al.*, 1995; Robinson *et al.*, 1998; Seagroves *et al.*, 1998). All C/EBP except C/EBPζ, form homo- and hetero-dimers (Ramji and Foka, 2002). Dimerisation is essential for DNA-binding activity. However, gene transcription is dependent upon the relative proportions of hetero- and homo-dimers that bind to a promoter (Davies *et al.*, 2000; Legraverend *et al.*, 1993), as demonstrated for C/EBPβ in the rat mammary gland (Descombes and Schibler, 1991; Zahnow *et al.*, 2001).

Three different C/EBPβ isoforms interact to direct either proliferation or differentiation of the rat mammary gland (Descombes and Schibler, 1991; Zahnow *et al.*, 2001). C/EBPβ liver inhibitory protein (LIP), the smaller dominant-negative isoform, lacks the N-terminal trans-activation domain and is associated with proliferation (Zahnow *et al.*, 2001). In contrast, C/EBPβ liver-enriched activator protein-1 (LAP1) and C/EBPβ LAP2 are associated with differentiation. Hence, the C/EBPβ LAP/LIP ratio acts as a cell cycle switch, resulting in either MEC proliferation or differentiation (Descombes and Schibler, 1991; Zahnow *et al.*, 2001). During mammary development and pregnancy C/EBPβ LIP is expressed (Rosen *et al.*, 1998), inhibiting MEC differentiation and the expression of milk protein genes (Dearth *et al.*, 2001). At parturition, C/EBP LIP is down-regulated, allowing C/EBPα and C/EBPβ LAP to activate mammary differentiation and milk protein gene transcription (Rosen *et al.*, 1999). Generally a C/EBPβ LAP/LIP ratio less than five suppresses *CSN2* expression during pregnancy, while the surge of prolactin and glucocorticoids at parturition promotes more than a 100-fold increase in C/EBPβ LAP/LIP ratio during lactation, favouring milk protein gene expression (Dearth *et al.*, 2001; Raught *et al.*, 1995; Rosen *et al.*, 1999).

Other Transcription Factors. The octamer binding proteins (OCT1 and OCT2) (Dong and Zhao, 2007a, b; Groenen *et al.*, 1992) and insulin response element (IRE) (Locker, 1993) also promote milk protein gene transcription. In contrast, Yin-Yang-1 (YY1) (Meier and Groner, 1994; Raught *et al.*, 1994), pregnancy-specific mammary nuclear factor (PMF) (Lee and Oka, 1992) and single stranded DNA-binding transcriptional regulator (STR) (Altiok and Groner, 1994) suppress *CSN2* transcription.

CoREs (Composite Response Elements). Milk protein gene activation and repression often depends upon the synergistic interaction of multiple transcription factors (Qian and Zhao, 2014; Rosen *et al.*, 1998). Each distinct cluster of elements, a 'composite response element' (CoRE), may include both positive and negative regulatory factors (Rosen *et al.*, 1989; Schmidhauser *et al.*, 1992). As transcription factors usually ubiquitously expressed, the mammary gland-specific expression of milk protein genes often relies upon the interaction of factors within a CoRE (Miner and Yamamoto, 1991; Rosen *et al.*, 1998).

Rat *CSN2*, one of the most-studied milk protein genes, has a proximal CoRE of STAT5, ½GRE, C/EBP and YY1 sites, part of which includes the 'milk-box', or 'lactation-associated complex (LAC)' (-140/-110) (Doppler *et al.*, 1989; Raught *et al.*,

1994; Rosen *et al.*, 1998; Schmitt-Ney *et al.*, 1991). In the presence of lactogenic hormones, YY1 has no effect on *CSN2* gene transcription. However, in their absence, YY1 binds preferentially to the LAC site and represses *CSN2* transcription (Groner *et al.*, 1994; Raught *et al.*, 1994; Schmitt-Ney *et al.*, 1991). The conservation of the 'milk box' in other milk protein gene promoters (rat *CSN1*, bovine *CSN1S1*, guinea-pig *CSN1S2* and rat, human and guinea-pig *LALBA*) suggests that this may be an evolutionarily conserved regulatory mechanism specific to the mammary gland (Hall *et al.*, 1987; Laird *et al.*, 1988; Yu-Lee *et al.*, 1986). Furthermore, milk box-like CoREs containing STAT5, ½GRE and C/EBP sites are also present in the distal promoters of bovine *CSN2* and rabbit *CSN1S1* (Raught *et al.*, 1995; Wyszomierski and Rosen, 2001).

Other CoREs are also located in the *WAP*, *BLG* and *LALBA* promoters. *WAP* has both a hormone-dependent STAT5, NFI and ½GRE CoRE (Li and Rosen, 1994a, b, 1995; Mukhopadhyay *et al.*, 2001; Rosen *et al.*, 1998) and a hormone-independent proximal promoter (-195/+28) CoRE which includes ½GRE, MAF and NFI sites, plus an F11 element (ACAAAG) (Doppler *et al.*, 1991; Lubon and Hennighausen, 1987; Mink *et al.*, 1992). The transcriptional regulation of human and rat *LALBA* may be similar to *WAP* due to the conservation of putative MAF, F11 and NFI motifs (Mink *et al.*, 1992). Furthermore, lactogenic hormones synergistically activate NFI and STAT5 to induce *LGB* expression (Watson *et al.*, 1991; Whitelaw *et al.*, 1991).

Extracellular Matrix and Milk Protein Gene Expression – Lessons from in vitro Culture Systems

The mammary gland is an extremely complex organ, governed by both systemic interactions and local factors intrinsic to the gland. This has presented challenges for the identification of regulatory mechanisms *in vivo* (Dimri *et al.*, 2005; Ip and Darcy, 1996; Weaver and Bissell, 1999). Therefore, both 2D and 3D *in vitro* culture systems have provided excellent models with which to investigate the regulation of the mammary gland. Many different models have been used; each of which has both advantages and disadvantages. Some of these include: whole animal models, i.e. *in vivo* hormone treatments of animals, transgenic animals and gene knockout models (not discussed in this review); mammary tissue explants (endocrine and growth factor regulation of genes) and MEC primary and immortalised cell culture systems (induction of endogenous and/or inserted foreign milk protein genes/promoters). *In vitro* mammary culture experiments have highlighted the importance of interactions between the ECM, MECs and myoepithelial cell and in particular, the 3D architecture of tissue for both mammary gland development and milk protein gene expression (Aggeler *et al.*, 1988;

Dimri *et al.*, 2005; Lin and Bissell, 1999; Streuli and Bissell, 1991).

Mammary Gland Explant Culture. Mammary gland tissue explants generally retain their 3D-glandular structure, biochemical functions and hormonal responses *in vitro* and so provide a close approximation of the gland *in vivo* (Brennan *et al.*, 2008a; Dils and Forsyth, 1981; Enami and Nandi, 1977; Matusik and Rosen, 1978). They therefore provide excellent models with which to investigate the regulation of milk protein genes. Mammary explants from numerous species including rodents, rabbits, pigs, ruminants and marsupials, have been used for *in vitro* gene regulation studies (Brennan *et al.*, 2007b; Collier *et al.*, 1977; Forsyth, 1971; Nicholas and Tyndale-Biscoe, 1985; Voytovich and Topper, 1967).

The tammar mammary explant culture system has provided valuable insights into the endocrine regulation of milk protein genes (Brennan *et al.*, 2008a; Brennan *et al.*, 2007b). While the *LALBA* and *BLG* genes are responsive to prolactin alone, *CSN2* requires insulin and prolactin for induction (Collet *et al.*, 1991, 1992; Maher and Nicholas, 1987; Nicholas and Tyndale-Biscoe, 1985). In contrast, all three lactogenic hormones are essential for the maximal *in vitro* expression of *CSN1*, *CSN2*, *WAP*, *LLPA* and *LLPB* (Collet *et al.*, 1992; Simpson, 1998; Trott *et al.*, 2002). However, the inability to induce *WAP*, *LLPA* and *LLPB in vitro* with lactogenic hormones unless these genes are expressed *in vivo* suggests that their regulation is complex (Simpson, 1998; Trott *et al.*, 2002). The induction of *WAP* with IHCPRLT$_3$E$_2$ in Phase 1 (day 24 of pregnancy) explants is a notable exception (Simpson, 1998). Interestingly, preliminary experiments showed that tammar *ELP* appears to be responsive to treatments in the absence of PRL (Simpson, 1998), uncommon for milk protein genes (Nicholas *et al.*, 1981; Nicholas and Topper, 1980).

Although explant culture systems provide excellent models, there are several disadvantages. Explants have limited tissue viability (from five days to ~2-3 weeks), contain multiple cell types, lack growth, and have a carryover of latent endogenous steroids, growth factors, inhibitors and hormones (Bolander *et al.*, 1979; Dils and Forsyth, 1981; Ip and Darcy, 1996; Resau *et al.*, 1991). Hence, the interpretation of results and in particular, the determination of MEC-specific effects can be problematic.

Explants must be cut small enough to allow for the absorption and even dispersal of exogenous hormones within the tissue. For the tammar, mammary explants must be cut into 1-2 mg pieces. Additionally, mammary explants often lack the feedback mechanisms present *in vivo*. Hence, although lactogenic hormones mimic lactogenesis

in explants, the lack of systemic signalling uncouples apoptosis in tammar, murine and bovine explants (Brennan *et al.*, 2008a; Brennan *et al.*, 2008b). Furthermore, explants must be used immediately. Therefore, tissue culture experiments cannot be repeated for the same animal (Keys *et al.*, 1997).

Primary Cells and Immortalised Cell Lines. Both primary and immortalised MECs have been used for cell culture studies. However, a significant challenge for researchers has been to create the appropriate culture environment for MECs to express milk protein genes. Primary MECs freshly isolated from the mammary gland provide an ideal resource, as they have the same physiological, biochemical and metabolic characteristics as those of the gland *in vivo* (Matitashvili *et al.*, 1997; Stampfer and Bartley, 1988). Unlike explants, stocks of the original cells can be stored and used for future experiments (Matitashvili *et al.*, 1997; Stampfer and Bartley, 1988). Primary cells isolated from the mammary may either be a mixed population of cells: MECs, myoepithelial cells, adipocytes, fibroblasts, immune and stem cells (Dimri *et al.*, 2005; Shackleton *et al.*, 2006). Alternatively, they can be fractionated, sorted or passaged to obtain a purified MEC population (Cane, 2005; Hu *et al.*, 2009; McGrath, 1987). Primary MECs have been isolated from different species including the tammar, cow, goat, sheep, seal, human and rodents (Cane, 2005; Ethier *et al.*, 1991; Finch *et al.*, 1996; Matitashvili *et al.*, 1997).

When cultured under appropriate conditions *in vitro*, primary MECs form 3D-alveoli-like 'mammospheres' (Huynh *et al.*, 1991; Johnson *et al.*, 2007; Plachot *et al.*, 2009; Stampfer and Bartley, 1988). Most importantly, mammospheres contain polarised MECs which vectorially secrete milk constituents into the central lumen, as occurs *in vivo* (Blatchford *et al.*, 1995; Burgoyne and Duncan, 1998; Schweisguth, 2004). The outer layer of a mammosphere consists of the basolateral membranes of MECs and hence contains cell-surface receptors for growth factors and hormones and transporters for ions and nutrients. Mammospheres also form an attachment to and secrete basement membrane components including laminins, type IV collagen, heparin sulphate proteoglycan, entactin and fibronectin (Handler, 1989; Leblond and Inoue, 1989; Rodriguez-Boulan and Salas, 1989; Strange *et al.*, 1991).

The functional differentiation of MECs *in vitro* requires cell-cell contact between MECs, adipocytes and fibroblasts; cell-ECM interactions; lactogenic hormones and growth factors (Bissell and Ram, 1989; Levine and Stockdale, 1985; Wicha *et al.*, 1982). Signalling from the tissue microenvironment, particularly the ECM, is essential

for the expression of the *CSN1S1*, *CSN2*, *BLG* and *WAP* genes (Chen and Bissell, 1989; Dale *et al.*, 1992; Jolivet *et al.*, 2001; Li *et al.*, 1987; Schmidhauser *et al.*, 1992). The two-way signalling between the external environment and the cell is mediated by integrin adhesion receptors in the cell membrane (Katz and Streuli, 2007). The contact between MECs and the ECM via laminin-111 ligation of integrins is essential for *CSN2* expression (Streuli *et al.*, 1991; Streuli *et al.*, 1995). Hence, the transmission of the correct external signal via integrin receptors to the intracellular cytoskeleton, alters cell shape and the 3D nuclear organisation of the chromosome(s) change (Ballester *et al.*, 2008; Nelson *et al.*, 1990; Parry *et al.*, 1990; Shannon and Pitelka, 1981). Critically, chromatin within and flanking milk protein genes becomes 'open' allowing transcription factors to bind the milk protein gene promoter, thereby inducing transcription (Kress *et al.*, 2010; Lemay *et al.*, 2013; Rijnkels *et al.*, 2013).

Primary murine, bovine, porcine and human MECs cultured without ECM on tissue culture plastic form a 2D-sheet of cells and lose their lactogenic phenotype (Ceriani, 1976; Li *et al.*, 1987; Riley *et al.*, 2010; Sun *et al.*, 2006; Yang *et al.*, 1979). The use of synthetic ECM substrates has alleviated these problems by providing a substratum for cell attachment and signalling. Collagen I gels (Lee *et al.*, 1984; Yang *et al.*, 1979), floating collagen membranes (Emerman *et al.*, 1977) and Matrigel, a reconstituted basement membrane derived from Engelbreth-Holm-Swarm (EHS) mouse sarcoma (Kleinman *et al.*, 1987; Kleinman *et al.*, 1986) have been used for this purpose. However, while bovine, porcine, human and rodent primary MECs require a matrix substratum in order to express milk protein genes, this is not true for Cape fur seal and tammar MECs which secrete their own matrix (personal communication, Sonia Mailer); (Cane, 2005; Sharp *et al.*, 2006).

One disadvantage of using primary cells is that they have a finite life-span, with a limited capacity to divide (Hayflick, 1965; Hayflick and Moorhead, 1961). The life-span for human MECs is 10-15 population doublings, or three to four passages (Ethier *et al.*, 1991; Yaswen and Stampfer, 2002). Therefore, once cells have reached their 'Hayflick limit', i.e., the number of times the cells can divide before the telomere length reaches a critical limit; they stop dividing and enter replicative senescence (Hayflick, 1998; Linskens *et al.*, 1995; Marx, 1994; Shay and Wright, 2000). Therefore, for animals in which the mammary glands are small and animal numbers are limited, it can be difficult to generate sufficient purified primary MECs for culture experiments.

Immortalised cell lines provide the same advantages as primary cells. However, they

have a stable telomere length which allows them to keep dividing indefinitely, while generally maintaining their phenotype (Dimri *et al.*, 2005; Harley *et al.*, 1990; Stampfer and Yaswen, 1993). A number of different immortalised MEC lines have been developed. HC11s, a clonal line derived from COMMA-1D cells from the mammary gland of a mid-pregnant mouse (Ball *et al.*, 1988; Danielson *et al.*, 1984) are commonly used for milk protein gene studies. However, the conditionally-immortalised murine KIM-2 cells (Gordon *et al.*, 2000), the human MCF-7 line derived from a breast carcinoma (Soule *et al.*, 1973) and bovine mammary alveolar cell-T (MAC-T) MECs isolated from the bovine mammary gland (Huynh *et al.*, 1991) have also formed the basis for *in vitro* regulatory studies.

Like primary MECs, immortalised MECs generally require lactogenic hormones and an ECM substrate to activate functional differentiation and the expression of milk protein genes. This is true for KIM-2 cells (Gordon *et al.*, 2000) and the bovine PS-BME-L6 and PS-BME-L7 lines (Gibson *et al.*, 1991). In contrast, ECM is not essential for the *in vitro* expression of *CSN2* and *WAP* in HC11s (Ball *et al.*, 1988; Ballester *et al.*, 2008; Doppler *et al.*, 1989; Goodman and Rosen, 1990). However, HC11s can lose their ability to respond to lactogenic hormones (personal communications, Jerome Demmer and Peter Williamson) (Trott, 1999).

A common problem for many cell-based studies has been the absence of a consistently reliable, lactogenic hormone-responsive MEC line (Groenen and van der Poel, 1994). An alternative approach has been to create the JAK2-STAT5 signalling pathway in non-MECs by the co-transfection of the plasmids ubiquitously expressing the PRLR and STAT5A (Groenen and van der Poel, 1994). For some cell lines, the GR is also required to ensure the maximal response of milk protein gene promoter constructs to lactogenic hormones *in vitro* (Buser *et al.*, 2007; Stocklin *et al.*, 1996; Yamashita *et al.*, 2001). Commonly used non-MEC lines include: human embryonic kidney cells (HEK293s); Chinese hamster ovary cells; and the African green monkey COS-1 and COS-7 kidney cell lines (CV-1 in Origin with SV40) (Ali *et al.*, 1992; Cereghini and Yaniv, 1984; Demmer *et al.*, 1995; Doppler *et al.*, 2001; Jolivet *et al.*, 1996; Trott *et al.*, 2005; Wyszomierski *et al.*, 1999).

Scope of the thesis

The specific control of protein synthesis during the first critical period of early lactation in the tammar wallaby is as yet unknown. This thesis therefore focuses on the

characterisation of the *Early Lactation Protein* gene in the tammar wallaby that changes dynamically between birth and ~day 100 pp. It will also investigate the evolution of the *ELP* gene, its possible function in mammals and its relationship to the *Colostrum Trypsin Inhibitor* gene in certain eutherian mammals that appear to be related to marsupial *ELP*.

The sequencing of the complete genomes of selected vertebrates, including the tammar wallaby (Renfree *et al.*, 2011) as well as PCR-based sequencing approaches provide an excellent method to characterise and compare the marsupial *ELP* gene and to identify the orthologous gene in eutherians such as the Cape fur seal. They also facilitate the determination of whether *ELP* is specific to mammals and therefore when the *ELP* gene most likely evolved.

The temporal, developmentally- and maternally-regulated expression of tammar *ELP* suggests that its regulation is complex. To understand the mechanisms regulating *ELP*, a tammar mammary gland explant culture system will be used, as well as primary and immortalised mammary epithelial cells. The tammar *ELP* promoter will be investigated and an *in silico* comparative genomics approach will be used to identify transcription factor binding sites within the *ELP* promoter region.

The presence of a single Kunitz-BPTI protease inhibitor domain within the mature ELP and CTI proteins suggests that ELP, like bovine CTI, inhibits serine endopeptidases such as trypsin. Structural homology modelling will be used to explore this hypothesis. Furthermore, mammalian and bacterial expression systems will be evaluated for the production of recombinant tammar ELP *in vitro*.

The marsupial mode of reproduction has elaborated lactation over placentation, and understanding its complex lactational physiology will provide novel information on the control of lactation in mammals in general. This thesis will address these questions.

Chapter Two:

General Materials and Methods

Introduction

The tammar is a highly synchronized seasonal breeder and the normal breeding season commences after the summer solstice in late-December (Tyndale-Biscoe, 2005). The quiescent corpus luteum, or the dormant blastocyst is reactivated by 2-4-fold increases plasma progesterone levels (Berger and Sharman, 1969). This is followed by oestrus, or the birth of a PY in mid-late January respectively (McConnell and Hinds, 1985). After emerging from the urogenital sinus, the tammar neonate uses its well-developed forearms and claws to climb unassisted up its mother's fur and into her pouch, guided by odours emanating from the pouch (Gemmell *et al.*, 2002; Schneider *et al.*, 2009). The PY then attaches to one of the four teats (Findlay, 1982a; Renfree *et al.*, 1989). However, while lactogenesis occurs in all four mammary glands, lactation continues only in the gland to which the neonate attaches. The non-suckled glands regress, returning to a quiescent state in readiness for the next pregnancy (Findlay and Renfree, 1984; Sharman, 1962). Neither pregnancy, nor progesterone withdrawal are prerequisites for lactation (Hinds, 1988; Maher and Nicholas, 1987; Nicholas and Tyndale-Biscoe, 1985), unlike in eutherians (Brisken and Ataca, 2015; Topper and Freeman, 1980). This was demonstrated by the induction of lactation in a non-pregnant female tammar by fostering a PY on to the teat of an adult female at the correct stage of the oestrous cycle (Sharman and Calaby, 1964).

Within 24 h of giving birth, the female tammar undergoes a postpartum (pp) oestrus and mating (Shaw and Renfree, 2006). By day 8 pp, the embryo has developed to the 80-100 cell (blastocyst) stage. However, embryonic development is halted by lactation, i.e., 'lactational quiescence' and the embryo enters 'diapause' (Figure 2.1) (Berger, 1966; Tyndale-Biscoe *et al.*, 1974). Furthermore, if the PY sucking stimulus is removed for ~72 h (removal of the pouch young; RPY), the corpus luteum and blastocyst reactivate and a new PY is born 26-27 days later (Gordon *et al.*, 1988; Tyndale-Biscoe and Renfree, 1987). This feature has been exploited by researchers to set the precise timing of tammar PY birth (Renfree and Tyndale-Biscoe, 1978; Shaw and Renfree, 2006). However, 'reactivation' can only occur prior to the onset of 'seasonal quiescence' at the winter solstice (early-June), after which photoperiodic inhibition of the corpus luteum occurs (Tyndale-Biscoe, 2005; Tyndale-Biscoe and Renfree, 1987). In contrast, marsupials such as the red kangaroo are 'opportunistic breeders' and can rapidly respond to their environment (Low, 1978). In poor seasons with limited food availability, the mother may terminate the PY and lactation (Hayssen *et al.*, 1985;

Tyndale-Biscoe, 1973), but when conditions become favourable, the dormant blastocyst is reactivated and the development of the embryo resumes (Renfree, 1983).

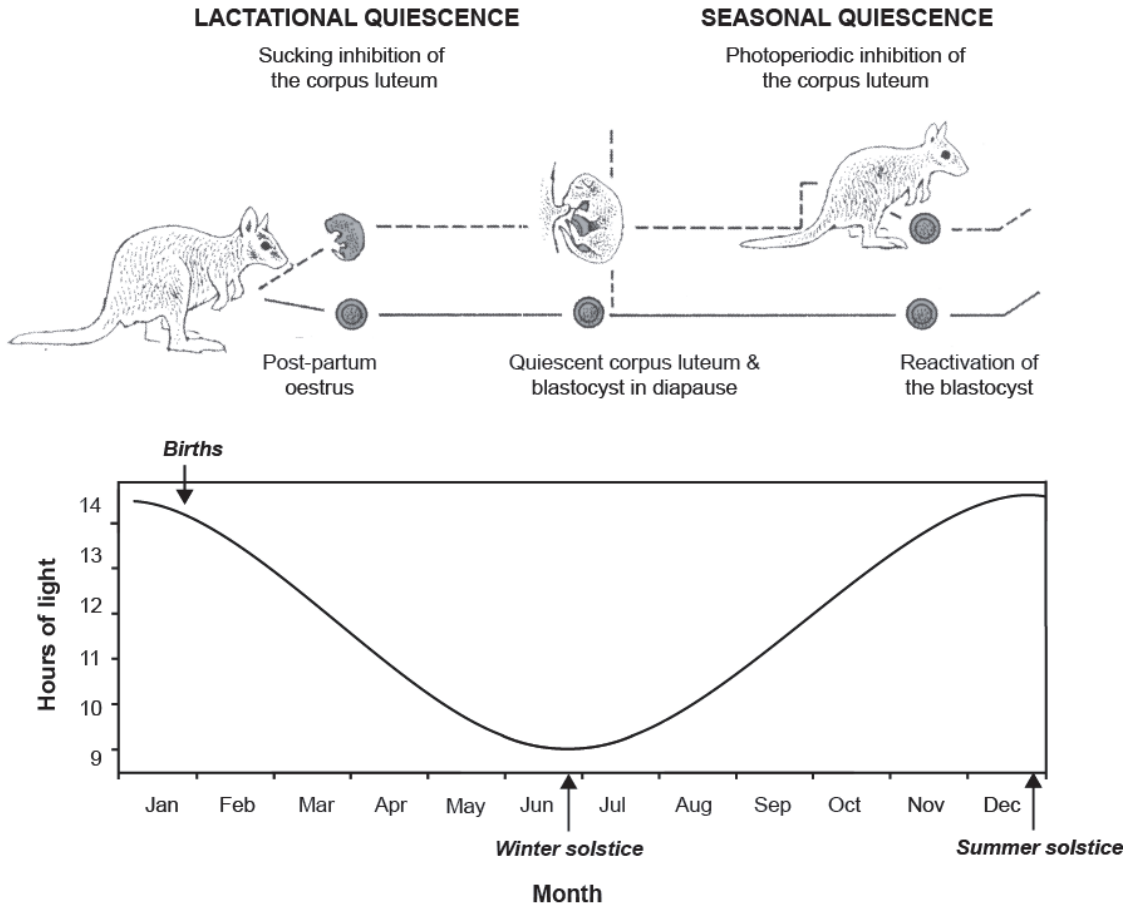


Figure 2.1. The annual reproductive cycle of the tammar wallaby, *Macropus eugenii*

The reproductive cycle of the tammar wallaby on Kangaroo Island, South Australia is governed by changes in photoperiod (day length) throughout the year. Six weeks after the summer solstice*, the majority of new births occur. Within 24 hours after birth tammars enter post-partum oestrus and mate. The blastocyst develops to the 100-cell stage and then enters diapause with its development suspended due to inhibition of the corpus luteum by the sucking stimulus of the pouch young (lactational quiescence). After the winter solstice* (June 20th or 21st) the corpus luteum remains quiescent and the blastocyst dormant (seasonal quiescence) until the summer solstice (December 21st or 22nd) due to the reduction in day length. (*exact date is dependent upon the year). Redrawn from (Tyndale-Biscoe, 2005).

In some cases, the blastocyst reactivates towards the end of the breeding season, so the mother may give birth to another PY at the start of the next season (Sharman and Calaby, 1964; Sharman and Pilton, 1964). Hence, the mother produces two different types of milk at the same time from different glands, for a young at foot and for an altricial newborn. This unique phenomenon is known as concurrent asynchronous lactation (CAL) (Lincoln and Renfree, 1981; Messer and Green, 1979). CAL was first observed in the red kangaroo and also occurs in other macropods such as the tammar and agile wallabies (Sharman and Calaby, 1964; Griffiths *et al.*, 1972; Lincoln and Renfree, 1981; Nicholas, 1988). This suggests that marsupial mammary gland regulation is complex and includes a component of regulation intrinsic to the mammary gland.

The tammar reproductive cycle includes a 26.5 day pregnancy (Phase 1) after diapause and a 300-350 day lactation period, which is divided into early- (Phase 2A), mid- (Phase 2B) and late- (Phase 3) lactation (Green *et al.*, 1980; Tyndale-Biscoe and Janssens, 1988). During the long and complex lactation, many developmental milestones occur in the tammar PY (Table 2.1).

The tammar PY spends the first 200 days pp in the pouch (Phases 2A and 2B), somewhat like an 'in pouch' gestation, before emerging fully developed in Phase 3 to ingest herbage (Tyndale-Biscoe and Renfree, 1987). For the first 100-110 days pp (Phase 2A) the PY's large mouth, lips and tongue grow around and become firmly attached to the teat (Gemmell *et al.*, 1988; Tyndale-Biscoe and Renfree, 1987). The ectothermic PY breathes and ingests milk at the same time and the milk travels to the relatively well-developed stomach and duodenum (Tyndale-Biscoe, 2005). Until around 90-110 days pp, the PY has only an innate immune system (non-specific immune response) but cannot mount a specific immune response (acquired or adaptive immune system) itself (Basden *et al.*, 1997; Edwards *et al.*, 2012). During Phase 2A, PY brain growth is maximal and its kidneys are immature (Renfree *et al.*, 1982; Tyndale-Biscoe and Renfree, 1987). During this period, which has been equated with the chorioallantoic phase of eutherian gestation (Hayssen, 1993; Piotte and Grigor, 1996), *ELP* is expressed and the protein secreted into milk (Simpson, 1998; Simpson *et al.*, 1998). However, neither the precise timing of *ELP* gene induction nor its role in the mammary gland and/or PY development have been determined.

Table 2.1. Stages of post-natal development in the tammar wallaby PY*

PY Age (days)	Characteristic
0 (birth)	Birth weight ~370-400 mg Eyes closed, sense of smell, touch and taste Disproportionately developed forearms Lungs immature Brain growth is disproportionate to the rest of the body from days 0-100 Low metabolic rate, PY is ectothermic and immuno-incompetent
5	Appearance of lymphocytes in thymus and lymph nodes
12	Appearance of lymphocytes in spleen
23	Hassall's Corpuscles in cervical thymus
39	Hassall's Corpuscles in thoracic thymus
42-56	Significant increase in metabolic rate
65	Formation of first alveoli in lungs
68	White pulp in spleen and Follicles in nodes
70	Six layers in the cerebral cortex
90	Maturation of the lymphoid tissues
95-105	First voluntary release of the teat
97	Serum IgG concentration increases
100	Brain growth slows dramatically
111	PY metabolic rate reaches adult levels
120	Fore-stomach contents acidic until ~day 200
125	Mature GALT (gut-associated lymphoid tissues)
140	Eyes open, ears reflect from the head, underfur is visible
142	Formation of respiratory bronchioles, alveolar ducts and alveolar sacs in the lungs
150-170	Light velvet fur appears
160	The PY can stand unaided and the pelage thickens
170-180	Maturation of the PY gut Differentiation of the fore- and hind stomach apparent
170-200	Full velvet fur develops
175	Intra-lumenal pepsin, tryspin and chymotrypsin activities remain low until this stage
180	Thyroid gland is fully developed and the young is homeothermic The young pokes its head out of the pouch and nibbles grass
190	Nephrogenesis (kidney development) complete PY makes its first exit from the pouch Herbage present in the fore-stomach
200-250	Pelage complete – mature fur
230	Fore-stomach contents approach neutral pH
240	Peak milk intake
250	Permanent pouch exit
300-350	Weaning complete - PY is independent
9 months	Females reach sexual maturity
24 months	Males reach sexual maturity

*Data was compiled from (Menzies, 2008; Setchell, 1974; Holt *et al.*, 1981; Renfree *et al.*, 1982; Cavanagh *et al.*, 1985; Reynolds *et al.*, 1985; Tyndale-Biscoe and Renfree, 1987; Hughes and Hall, 1988; Janssens and Messer, 1988; Reynolds and Saunders, 1988; Tyndale-Biscoe and Janssens, 1988; Basden *et al.*, 1997; Renfree *et al.*, 2001; Tyndale-Biscoe, 2005; Waite *et al.*, 2005; Szdzuy *et al.*, 2008; Joss *et al.*, 2009; Kwek *et al.*, 2009)

From days ~100-200 pp (Phase 2B), the young remains in the pouch, but is only intermittently attached to the teat (Tyndale-Biscoe and Janssens, 1988; Tyndale-Biscoe and Renfree, 1987). Towards the end of Phase 2B, ~170-180 days pp, the cells lining the PY gut mature and the foregut and hindgut become differentiated (Kwek *et al.*, 2009a; Kwek *et al.*, 2009b; Waite *et al.*, 2005). In eutherians, gut cell differentiation marks 'closure', as mucosal enterocytes lose the capacity to absorb macronutrients and intact immunoglobulins (Kruse, 1983; McFadden *et al.*, 1997). However, whether gut maturation is correlated with the loss of Ig absorption in the tammar is not known. By the end of Phase 2B, PY is development is complete, with the maturation of the small intestines and colon, the completion of the pelage and the development of endothermy (Tyndale-Biscoe and Renfree, 1987).

Phase 3, the final stage of tammar lactation, is similar to eutherian lactation (Dove and Cork, 1989). The mature PY emerges from the pouch, increases its milk consumption and begins to ingest herbage and so its growth rate is exponential (Tyndale-Biscoe and Janssens, 1988). Consequently, maternal mammary gland size increases as the mother boosts milk production to meet the increased needs of the PY for energy and growth (Findlay, 1982b). Permanent pouch exit occurs around day 250 pp, with weaning generally complete by ~270-350 days pp (Janssens *et al.*, 1997).

Animals

Tammar Wallabies

Tammar wallabies (*Macropus eugenii*) used for this research (Appendix I) were sourced from two different captive colonies. Firstly, from the Victorian Institute of Animal Science (VIAS), Department of Primary Industries, Attwood, Victoria and the remainder, from the University of Melbourne Macropod Research Facility, Wantirna, Victoria. Animals were kept in open grass yards with *ad libitum* access to food, water and shelter in accordance with the National Health and Medical Research Council guidelines (NH&MRC, 2013).

Cape and Subantarctic Fur Seal Samples

Total RNA extracted from Cape (South African) fur seal mammary glands was provided by Dr. Kylie Cane. Mammary gland tissue was obtained from six Cape fur seal females and included: foraging at-sea (early lactation), on-shore lactating (early lactation) and from pregnant females (Refer to Chapter 4). Tissue samples were collected by Dr. John Arnould and Dr. Kylie Cane in February 2003. At-sea samples

were collected from two females on a foraging trip west of Cape Town, South Africa; past the continental shelf and 100 km from the nearest colony. Both at-sea females were lactating, with no signs of involution in their mammary glands, as determined by histological analyses (Cane, 2005). Mammary tissue was collected from two females on-shore during early-lactation, in a colony 40 nautical miles from Cape Town at Kleinsee (29˚40'0 S, 17˚4'60 E). Tissue from two Cape fur seals during late gestation was collected from a breeding colony at Robesteen (33˚37'60 S, 18˚23'60 E). One of these females was determined to be in its first pregnancy (Cane, 2005).

Subantarctic fur seal mammary gland RNA was also provided by Dr. Kylie Cane. The mammary tissue was originally collected by Dr. John Arnould at La Mare aux Elephants (46˚22' S, 51˚40' E), Possession Island (Iles Crozet), southern Indian Ocean. One of the females was two months post-partum and had been on-shore for more than 12 hours (Cane, 2005).

Miscellaneous Samples

Southern Koala tissue samples, from which genomic DNA was isolated, were sourced from the ear-tagging of selected animals in a Victorian population of koalas on French Island. Samples were generously supplied by Dr. Kath Handasyde and Emily Hynes; as was a small quantity (~100 μL) of koala milk from late-lactation. A small quantity of dog colostrum (~ 20 μL) was also kindly donated by Cate Pooley from Guylar, a late-pregnant (2 weeks pre-partum) Labrador in her first pregnancy.

Mammary gland total RNA extracted from a South American opossum mammary gland at day 15 of lactation (early-lactation) was provided by Denjal Topcic (The University of Melbourne). Alison de Leo (The University of Melbourne) provided total RNA from the mammary gland of the fat-tailed dunnart from day 37 of lactation (early-lactation). Genomic DNA from the brushtail possum) was provided by Kim Huyhn (The University of Melbourne).

Ethics Protocols

All experiments conducted using tammar wallabies were approved by the Animal Experimentation Ethics Committees from the Victorian Institute of Animal Science, Department of Primary Industries, or by The University of Melbourne Animal Ethics Committee. Wallabies were euthanised as described in the animal ethics permits and tissues were dissected from animals using sterile techniques, frozen in either liquid nitrogen (liquid N_2) or on dry ice and stored at -80°C. Mammary tissue for culture experiments was placed in Medium 199 (Figure 2.2; Chapter 5) prior to cutting of

explants and tissue for histology was collected in 10% neutral buffered formalin (NBF) prior to processing. Approval for Cape fur seal research was obtained from the South African, French and Australian Governments (Chapter 4).

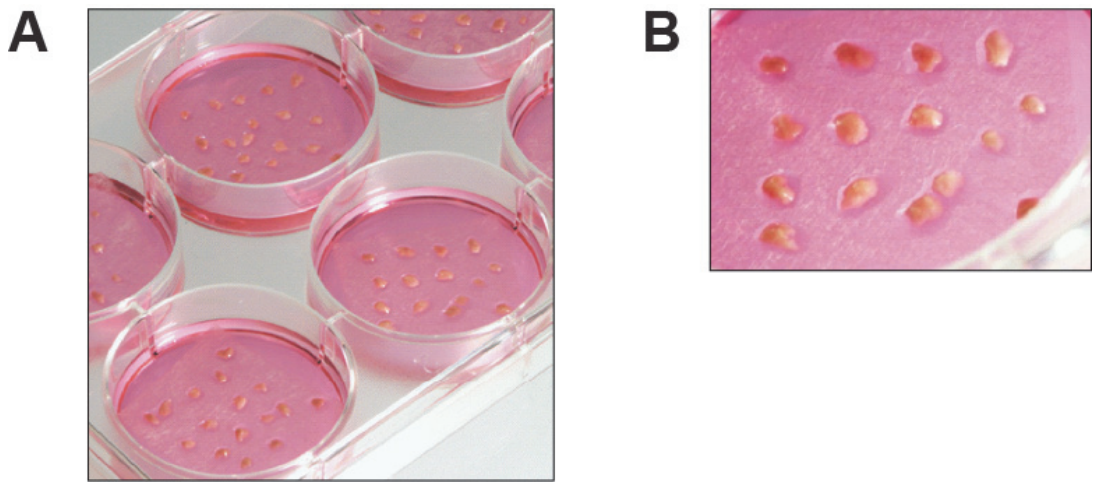


Figure 2.2. Tammar wallaby mammary gland explant culture system

A. Explants were cultured on siliconised lens paper on culture medium (2 mL) in 6-well plates. **B**. The porous siliconised lens paper, allows the transfer of nutrients from the media to the tissue.

Histology

Fresh mammary gland tissue or mammary explants were placed in greater than 10-fold excess of 10% NBF. Frozen mammary gland tissue (-80°C) was thawed in 10% NBF, processed using an LX120 Tissue Processor and embedded in Paraplast tissue embedding medium by Bruce Abaloz, Department of Zoology, University of Melbourne. Tissue sections (8 µm) were cut on a MICRON/Zeiss microtome using Feather disposable blades. Sections were deparaffinised in histolene/solvent (2 x 10 min), followed by washes in ethanol (EtOH, 100%: 20 dips, 100%: 12 dips, 80%: 12 dips, 70%: 12 dips), rinsed in MQ H_2O until clear (~20 dips) and dried overnight (45°C). Sections were either stained with Mallory's phosphotungstic acid hematoxylin (PTAH) stain, or Hematoxylin and Eosin, a cover slip applied and then examined under an Olympus BX50 system microscope and photographed using a Nikon Digital Sight DS-5M camera and imaging system (DS-L1).

Milk Collection

Tammar PYs were removed from their mothers for ~3 h and placed in calico bags in a warm box to allow milk to accumulate in the mother's mammary gland. Tammar mothers were anaesthetised with an intramuscular injection of Zoletil 100 (0.15 mL, Virbac, Melbourne, Victoria). Once sedated, wallabies were injected intramuscularly with oxytocin (0.1 mL, 10 IU/mL, Heriot Agvet, Rowville, Victoria). Milk was manually expelled from the mammary gland by applying gentle pressure to the base of the gland (Figure 2.3), collected in 1.5 mL centrifuge tubes, snap frozen on dry ice and stored at -80°C.

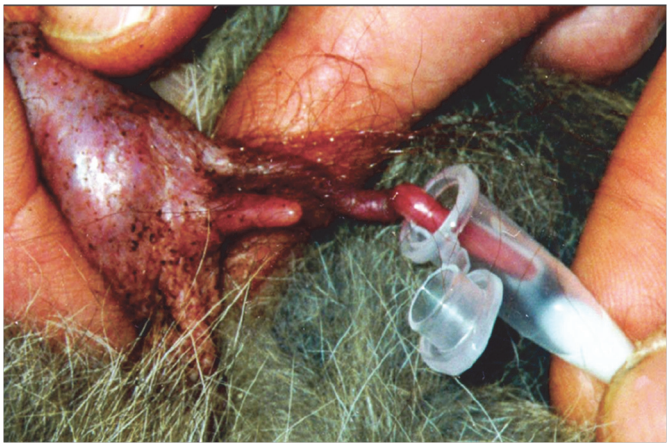


Figure 2.3. Collection of milk from the mammary gland of a tammar wallaby

Tammar PYs were sexed and their head length measured with calipers, to determine their age and hence stage of lactation of the mothers (Poole *et al.*, 1991) (Figure 2.4).



Figure 2.4. Measurement of tammar PY head length with calipers

Molecular Biology

DNA Oligonucleotide Design

DNA oligonucleotides were designed by manual assessment or with Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Rozen and Skaletsky, 2000). Oligo Calculator (http://www.pitt.edu/~rsup/OligoCalc.html) was used to assess primer characteristics, with the exception of Phusion High-Fidelity DNA Polymerase (F-530S, Finnzymes). Phusion primer annealing temperatures were determined using the Phusion-specific calculator on the Finnzymes website (http://www.finnzymes.fi/Java/tm_determination.htm).

Isolation of Genomic DNA

Genomic DNA was isolated from tissues using a modified version of the protocol described by (Sambrook and Russell, 2001) and from cells using TriPure Isolation Reagent (1667 165, Roche Diagnostics). Tissue was homogenized in DNA lysis buffer (10 mM Tris-Cl pH 8.0, 0.1 M EDTA, pH 8.0 and 0.5% (w/v) SDS; 1 mg tissue/10 mL buffer). RNaseA (19101, Qiagen) was added (final concentration: 20 µg/mL) and the homogenate incubated at 37 ˚C for 1 h. Proteinase K (V302B, Promega) was added (final concentration: 100 µg/mL) and the tube incubated at 50 ˚C for 3 h. DNA was extracted 3x with an equal volume of PCI (phenol: chloroform: isoamyl alcohol, 25:24:1). Briefly, PCI was added, the tube inverted several times to mix the contents and then centrifuged at 3220 *g* for 15 min. The supernatant was removed and extracted twice more with an equal volume of PCI and then extracted with an equal volume of chloroform: isoamyl alcohol (49:1). Tubes were centrifuged at 3220 *g* for 15 min and the supernatant transferred to a new tube with a wide-bore pipette tip. The DNA was precipitated by the addition of 0.1 vol 3 M NaOAc (sodium acetate, pH 4.5) and isopropanol (0.54 vol) and the tube spun briefly. The supernatant was discarded, the pellet dried and then resuspended in TE, pH 8.0.

Genomic DNA was extracted from adherent cultured cells using TriPure (1 mL per ~10^7 cells). Total RNA was extracted first (Refer to 'Isolation of Total RNA from Tissue, Milk and Cells' section) and DNA then extracted from the remaining inter and organic phases. EtOH (100%, 300 µL) was added to the remaining inter- and organic phases, tubes inverted several times, incubated for 2-3 min at room temperature (RT) and centrifuged at 2,000 *g* for 5 min at 4 ˚C. The supernatant was removed for isolation of proteins. The remaining DNA was then washed 3x with 0.1 M sodium citrate in 10% EtOH. Sodium citrate/EtOH (1 mL) was added to the DNA, which was then incubated

at RT for 30 min with occasional mixing. The sample was centrifuged (2,000 *g* for 5 min at 4 ˚C) and the supernatant removed and washed twice more. The DNA was then washed with 75% EtOH (2 mL) and incubated at RT for 15 min with occasional mixing, centrifuged at 2,000 *g* for 5 min at 4 ˚C. The supernatant was removed and the pellet resuspended with 8 mM NaOH to ~ 0.3 µg/mL. Genomic DNA was quantified with the Bio-Rad SmartSpec 3000 Spectrophotometer or Nanodrop.

Reverse Transcriptase-PCR

First strand complimentary DNA (cDNA) was generated using Superscript III Reverse Transcriptase (18080-044, Invitrogen Corporation), oligo(dT)$_{20}$ primer (50 µM) from Sigma-Proligo and 5 µg of total RNA isolated from mammary tissue, milk, or cells. A minus Reverse Transcriptase control (no RNA) was also used. PCR was performed using 2 µL (10%) of the first strand reaction as a template, with the appropriate forward and reverse primers.

A number of different *Taq* DNA polymerases were used depending upon the application of the PCR reaction. For long-range and high-fidelity PCR reactions, either the Invitrogen Platinum *Taq* DNA Polymerase High Fidelity (11304-011) or Phusion High-Fidelity DNA Polymerase (F-530S, Finnzymes) with proof-reading ability were used. Invitrogen Taq DNA Polymerase, recombinant (10342-053) was used for medium range PCRs and *Taq* DNA Polymerase in Storage Buffer B (M1661, Promega Corporation) was used for diagnostic PCRs, i.e. screening for the presence or absence of a transcript.

Electrophoresis of DNA

PCR products were analysed by electrophoresis through agarose gels (0.7-2.0%) containing 1 µg/mL ethidium bromide (EtBr) in 1x TAE (Tris-acetate-EDTA) buffer (Appendix II). The 1kb Plus DNA Ladder (0.1 µg, 10787-018, Invitrogen) was run concurrently with samples to determine fragment size(s) of dsDNA band(s). Electrophoresis was carried out for ~1 h in 1x TAE at 100 V, gels visualized using the UVP BioDoc-It 2UV transilluminator (UVP Inc., Upland, CA, USA) and the images captured using WinFastPVR (Leadtek Research Inc., Chung Ho City, Taipei County, Taiwan).

Gel Extraction

DNA fragments less than 10 kb were gel extracted using the QIAquick Gel Extraction Kit (28704, Qiagen). All centrifugation steps were performed at 10,000 *g*. Briefly, fragments were excised from the gel and weighed, 3 gel volumes of Buffer QC

added and the sample incubated for 10 min at 50°C, with vortexing at 2 min intervals. One gel vol of 100% isopropanol was added, the sample mixed, applied to a spin column and centrifuged for 1 min. The flow-through was discarded, as for the following two steps. Buffer QC (500 µL) was applied to the column and the sample centrifuged for 1 min. The column was washed with Buffer PE (750 µL), spun for 1 min, then dried by spinning for a further minute. Gel-extracted DNA was eluted with Buffer EB (50µL) after a 1 min spin.

DNA fragments greater than 10 kb were extracted using the QIAEX® II Gel Extraction Kit (20021, Qiagen) according to the manufacturer's protocol. Briefly, the gel slice was weighed, 3 vol of Buffer QX1 and 2 vol MQ H_2O were added, plus 30 µL QIAEX II beads, which had been vortexed for 30 seconds. The sample was incubated at for 50°C for 10 min, gently mixed every 2 min, then centrifuged at 10,000 *g* for 30 s. The supernatant was removed, the pellet resuspended with Buffer QX1 (500 µL) and washed 2x with Buffer PE (500 µL). The pellet was air-dried for 10 min, resuspended in 10 mM Tris-Cl, pH 8.5 (20 µL), centrifuged for 30 s and the supernatant containing the DNA removed.

Ligations

Gel-purified PCR products were cloned using the pGEM-T Easy Vector System I (A1360, Promega). Briefly, the dsDNA fragment was combined with pGEM-T Easy Vector (1 µL), 10x ligase buffer and T4 DNA ligase and incubated overnight at 4 ˚C. Gel-purified, restriction enzyme-digested dsDNA was either ligated into pGEM-T easy for sequencing, or into an alternate vector depending upon the downstream use (fragments were ligated overnight: sticky-ends at 4°C, blunt-ended dsDNA at 16°C).

Electroporation or Transformation of DNA into *E. coli*

Electrocompetent cells (50 µL), XL10-Gold Ultracompetent Cells, (200314, Stratagene; refer to Appendix III for the method used to prepare of electrocompetent cells) were thawed on ice, either 2 µL (20% of ligated DNA) or plasmid DNA (0.5 µL) added and transferred to an ice-cold Gene Pulser/MicroPulser Cuvette 0.2 cm gap (165-2082EDU, Bio-Rad). The bacteria were electroporated (Bio-Rad MicroPulser set at Ec2 to deliver a 2.5 kV pulse). The cells were placed on ice, Luria-Bertani (LB broth, 950 µL) added and the tubes shaken at 200 rpm, 37°C for 1 h. Appropriate amounts of the bacterial culture (generally 50 µL and 200 µL) were spread onto LB agar plates containing the relevant antibiotic and incubated overnight with shaking at 37°C. Single colonies were inoculated into LB broth (5 mL) and grown overnight (250 rpm, 37°C).

Glycerol stocks were prepared using 500 µL of bacterial growth, as described (Sambrook and Russell, 2001).

JM109 chemically competent cells 100 µL, (L2001, Promega, see Appendix III for the method used to prepare chemically competent cells) were thawed on ice for 20 min, 5 µL of ligated DNA, or 1 µL plasmid DNA added, the cells heat-shocked at 42°C for 45 s and placed on ice for 2 min. LB broth (900 µL) was added and the cells grown and plated, as previously described.

Preparation of Plasmid DNA

Three different types of plasmid preparation methods were used during the course of this work, dependent upon the quantity, type and purity of plasmid DNA required. For small-scale, diagnostic purposes, miniprep DNA was extracted from a 5 mL overnight culture using the QIAprep Spin Miniprep Kit (27104, Qiagen). Transformed cells were pelleted at 2000 *g* for 7 min, resuspended with Buffer P1 (250 µL), Buffer P2 (250 µL) added and the tube inverted several times. Buffer N3 (350 µL) was added, the tube inverted several times and then centrifuged at 16,000 *g* for 10 min. The supernatant was transferred to a spin column and centrifuged at 16,000 *g* for 30 s. The column was washed with Buffer PB (500 µL), then Buffer PE (750µL), with the flow-through discarded after each step. The column was centrifuged for a further minute at 16,000 *g*, 50 µL Buffer EB (10 mM Tris-Cl, pH 8.5) added, allowed to stand for 1 min and the DNA eluted by centrifugation at 16,000 *g* for 1 min.

Bacterial Artificial Chromosome (BAC) DNA was isolated using the Qiagen Midi Kit, but using the "Protocol for Very Low Copy Plasmid/Cosmid Purification", (12145, Qiagen). An LB broth culture (5 mL) with the appropriate selective antibiotic was inoculated with a single colony and grown for ~8 h at 37°C with vigorous shaking (~300 rpm). This starter culture was diluted 1:500 into 500 mL of 2YT media (Appendix II) and grown overnight (37°C, 300 rpm). Bacterial cells were harvested by centrifuging at 6,000 *g* for 15 min at 4°C (Beckman J2-21M/E Centrifuge, JA20 rotor). The bacterial pellet was resuspended in Buffer P1 (20 mL), Buffer P2 added (20 mL), mixed gently by inverting 4-6 times and incubated at RT for 5 min. Chilled Buffer P3 (20 mL) was added and mixed immediately by inverting 4-6 times, then incubated on ice for 30 min. The tubes were centrifuged at 20,000 *g* for 30 min at 4°C, the supernatant transferred promptly to new tubes and centrifuged again at 20,000 *g* for 15 min at 4°C. The supernatant was transferred into 50 mL tubes and the DNA precipitated by adding 42 mL (0.7 vol) of RT isopropanol and stored overnight at -20°C.

The following day, the tubes were centrifuged at 3,220 *g* (Eppendorf 5810R centrifuge rotor A-4-62) for 1 h at 4°C. The pellet was redissolved in TE buffer (500 µL), pH 8.0 and Buffer QBT added to a volume of 5 mL. A QIAGEN-tip 100 column was equilibrated with Buffer QBT (4 mL) and the flow-through discarded. The dissolved DNA was bound to the column and washed with 2x 10 mL Buffer QC. The DNA was eluted with Buffer QF (5 mL), precipitated with 3.5 mL (0.7 vol) RT isopropanol, mixed and then centrifuged at 3,220 *g* for 1 h at 4°C. The pellet was washed with RT 70% EtOH (2 mL) and centrifuged for another hour as previously described. The supernatant was removed and the pellet air-dried for 5-10 min, then redissolved in TE buffer, pH 8.0.

The Qiagen Maxi Plasmid Prep Kit (12163, Qiagen) was used to prepare large volumes of DNA for transfection into cells. The maxi prep method was very similar to the midi prep. A starter culture was prepared and diluted 1/500 dilution into 100 mL LB broth plus antibiotic, grown overnight and pelleted, as previously described. The pellet was resuspended in Buffer P1 (10 mL), then Buffer P2 (10 mL) added, followed by chilled Buffer P3 (10 mL) and incubated on ice for 20 min. The sample was centrifuged at 20,000 *g* as described, but after the second spin the supernatant was applied to a QIAGEN-tip 500, which had been equilibrated with Buffer QBT (10 mL). The column was washed with 2x 30 mL Buffer QC and eluted with Buffer QF (15 mL). Plasmid DNA was precipitated with RT isopropanol (10.5 mL) and centrifuged at 3,220 g for 1 h at 4°C, the DNA pellet washed with RT 70% EtOH (5 mL), centrifuged for another hour as previously described, the supernatant discarded, the pellet air-dried and resuspended in TE buffer.

For transfection of plasmid DNA into primary cells, endotoxin-free DNA was prepared using the QIAfilter Plasmid Maxi Kit (12263, Qiagen). A bacterial culture (100 mL) was prepared, as described for the non-filter maxi prep. Bacterial cells were harvested by centrifugation at 6,000 *g* for 15 min at 4°C, the supernatant discarded and the pellet resuspended in 10 mL Buffer P1 containing LyseBlue reagent, and then Buffer P2 (10 mL) added. The sample was thoroughly mixed by inverting the tube vigorously 4-6 times and then incubated at RT for 5 min. Chilled Buffer P3 (10 mL) was added to the lysate and immediately and thoroughly mixed by vigorously inverting 4-6 times. The lysate was added to the QIAfilter cartridge, incubated at RT for 10 min and a QIAGEN-tip 500 equilibrated as previously described. The lysate was filtered onto the QIAGEN-tip 500 column using the cartridge plunger and the column washed as described for the maxi kit, but with endotoxin-free reagents used.

Restriction Enzyme Digestion

For diagnostic digests, generally 2 µL (~0.5 µg) of plasmid DNA was digested in a 20 µL reaction, with 0.5 µL enzyme, plus 2 µL of the corresponding 10x enzyme buffer, as per the manufacturers' instructions (Promega, NEB, Roche). For double digests, ~1 µL of each enzyme was used. For excision of dsDNA fragments for subcloning, reactions were generally scaled up and performed in multiple tubes to ensure efficient DNA digestion.

Sequencing of Plasmid DNA

The nucleotide sequence of a cloned PCR product was determined by sequencing in either one or both directions using Sp6 and/or T7 oligonucleotide primers. Alternatively, a gene-specific primer (Appendix IV), using the BigDye Terminator v3.1 Cycle Sequencing Kit (4337456, Applied Biosystems) was used. The following components were assembled in a 20 µL reaction: 5x buffer (4 µL), BigDye 1.5 µL, 5 µM primer (1 µL), DNA (the amount used dependent upon the template size), plus MQ H_2O. Cycle sequencing was performed using an Eppendorf Mastercycler.

Sequencing reactions were then added to 80 µL of 75% isopropanol and the sample vortexed for 10 s, allowed to stand at RT for 20 min, centrifuged at 18,000 *g* for 20 min and the supernatant removed. The sample was pulse centrifuged and the remainder of the supernatant removed. Isopropanol (200 µL, 75%) was added, the sample vortexed for 5 s, spun at 18,000 *g* for 10 min and the supernatant removed as previously described. The sequencing reaction was dried down for 10 min at 50°C in a CHRIST RVC 2-25 speedy-vac (Martin Christ Gefriertrocknungsanlagen GmbH, Germany). Reactions were analysed by an ABI Prism 3700 Genetic Analyser at Applied Genetic Diagnostics, Pathology Department, University of Melbourne and chromatograms verified using Chromas Lite 2.0 (Technelysium Pty Ltd).

Isolation of Total RNA from Tissue, Milk and Cells

Total RNA was extracted from tissues using either the Qiagen RNeasy Mini Kit (74106) for up to 30 mg of tissue or the Qiagen RNeasy Midi Kit (74144) for up to 250 mg of tissue. The protocol for Isolation of Total RNA from Animal Tissues was used for both extractions. For microarray analysis of tammar mammary gland tissue, TriPure Isolation Reagent (1667 165, Roche Diagnostics) was used, but with an extra extraction step with Tripure performed on the aqueous phase containing the RNA. Total RNA was extracted from adherent, cultured cells with Tripure and from cells in milk using RNAwiz (9736, Ambion, Inc.).

Frozen tammar tissues and mammary explants were mechanically disrupted with either a hammer or mortar and pestle. Tissue (30 mg) was weighed and immediately homogenized Ultra-Turrax T8, probe S8N-5G (5 mm diameter, IKA-WERKE, Germany) for ~2 min at high speed in 600 µL Buffer RLT, to which β-Mercaptoethanol (β-ME) had been added (10 µL β-ME per 1 mL Buffer RLT). The lysate was centrifuged for 3 min at ~16,100 *g*, the supernatant transferred to a new tube and 1 vol 70% EtOH added and mixed thoroughly by pipetting. Up to 700 µL was applied to an RNeasy spin column and centrifuged at 8,000 *g* for 15 s, the flow-through discarded and this step repeated for the remaining sample. The column was then washed with Buffer RW1 (700 µL) by centrifugation at 8,000 *g* for 15 s, the flow-through discarded and washed an additional two times with Buffer RPE (500 µL). An optional step using the column placed in a new collection tube and spun at max. speed for 1 min was used to eliminate Buffer RPE carry-over. The column was then placed in a new collection tube, RNase-free water (30-50 µL) added, allowed to stand for 1 min, then centrifuged at 8,000 *g* for 1 min to elute the RNA.

Total RNA was extracted from 30 mg-250 mg tissue with the Qiagen RNeasy Midi Kit, with the method similar to that of the Mini Kit. Tissue (250 mg) was homogenised in 4 mL Buffer RLT, centrifuged at 3,000-5,000 *g* for 10 min, the supernatant transferred to a new tube, an equal volume of 70% EtOH added and the sample shaken vigourously. The sample (~8 mL) was applied to the column in 2 steps, with the flow-through discarded after each centrifugation step (3,000-5,000 *g* for 5 min). Buffer RW1 (4 mL) was added to the column, which was centrifuged again (3,000-5,000 *g* for 5 min), and the flow-through discarded. The column was washed by two sequential additions of Buffer RPE (2.5 mL), but with the first centrifugation step carried out at 3,000-5,000 *g* for 5 min and the second for 2 min. The flow-through was discarded after each wash. The column was transferred to a new collection tube and total RNA eluted by a 1 minute equilibration with RNase-free water (250 µL), followed by a final centrifugation step at 3,000-5,000 *g* for 3 min.

For microarray analysis of the tammar wallaby mammary gland, total RNA was isolated from mammary gland tissue with TriPure (1 mL per 50-100 mg tissue). The tissue was homogenised and extracted as described for cultured cells, but with the additional TriPure extraction step. Total RNA was extracted from adherent cultured cells using TriPure (1 mL/10^7 cells). Briefly, cell culture media was removed from wells and TriPure added. Cells were disrupted by pipetting, transferred to a tube, incubated for 5 min at RT to dissociate nucleoprotein complexes, centrifuged at 12,000 *g* for 15

min and the supernatant transferred to a new tube. Chloroform (0.2 mL per 1 mL Tripure) was added, the tube shaken vigorously for 15 s, incubated for 15 min at RT, centrifuged at 12,000 g for 15 min at 4 ˚C. The upper, aqueous phase was transferred to a new tube, the RNA precipitated by the addition of cold isopropanal (0.5 mL per 1 mL Tripure), mixed by inversion, incubated at RT for 10 min, then centrifuged at 12,000 *g* for 10 min at 4°C and the supernatant discarded. The cell pellet was washed with 75% EtOH (1 mL per 1 mL Tripure), centrifuged at 7,500 *g* for 5 min at 4 ˚C and the supernatant discarded. The RNA pellet was air-dried, resuspended in DEPC-treated H_20, incubated for 15 min at 55-60 ˚C and stored at -80°C.

Total RNA was extracted from milk using RNAwiz (9736, Ambion, Inc.), using a protocol very similar to that for Tripure. Milk was thawed on ice and centrifuged at 300 *g* for 15 min at 15°C. The fat and whey were removed and the cell pellet disrupted and resuspended in RNAwiz (1 mL per 10^7 cells) by vigorous pipetting. The homogenate was incubated at RT for 5 min, 0.2 vol chloroform added, shaken vigorously for 20 s and incubated at RT for 10 min. The mixture was centrifuged at 10,000 *g* for 15 min at 4°C, with the upper, aqueous phase transferred to a new tube, 0.5x the starting vol of RNase-free water added and mixed well. Isopropanol (1 starting vol) was added, the sample mixed well, incubated at RT for 10 min and centrifuged at 10,000 *g* for 15 min at 4°C to pellet the RNA. The supernatant was discarded, the pellet washed with at least 1 starting vol of cold 75% EtOH by vortexing and then centrifuged at 10,000 *g* for 5 min at 4°C. The supernatant was discarded and the RNA pellet air-dried for ~10 min and dissolved in RNase-free water to ~1 µg/µL. Total RNA was quantified using the Bio-Rad SmartSpec 3000 Spectrophotometer or Nanodrop.

Electrophoresis of Total RNA

RNA samples were precipitated if too dilute. Briefly, 2.5 vol 100% EtOH, 0.1 vol 3 M sodium acetate (NaOAc, pH 5.2) and 1 µL GlycoBlue 15 mg/mL (9515, Ambion) were added to the RNA, which had been thawed at 55°C for 10 min and kept on ice. Samples were mixed, centrifuged at 18,000 *g* for 15 min at 4°C, the RNA pellet washed with RT EtOH (300 µL, 70%), the sample spun and the ethanol removed. The pellet was air-dried for 10 min then dissolved in an appropriate volume of DEPC-treated H_2O. An equal volume of RNA loading dye (Appendix II) was added and samples heated at 65°C for 15 min, placed on ice for 2 min and centrifuged briefly prior to gel electrophoresis. RNA markers 2.5 µL (2.5 µg) 0.28-6.58 kb, (G3191, Promega) were thawed on ice and 7.5 µL DEPC-H_2O plus an equal volume of RNA loading dye added (10 µL). Markers were prepared as for the RNA samples, but heated at 65°C for 10 min prior to

electrophoresis. Generally, 10 µg total RNA was electrophoresed through a 1% agarose, low-formaldehyde (1.1%) gel, with 1x 3N-(Morpholino) Propane Sulfonic Acid (MOPS) buffer, Appendix II) at ≥ 5 V/cm i.e. 100 V for 90 min (Bio-Rad wide mini Sub-Cell GT) at 4°C.

Northern Blotting

RNA was transferred to Zeta-Probe Genomic Tested (GT) Blotting Membrane (162-0196) in 20x SSC overnight. The membrane was then rinsed in 2xSSC, dried between Chromatography Paper (3030917, Whatman), UV crosslinked (120 mJ) (Stratagene UV Stratalinker1800) and stored in plastic wrap at 4°C.

[α-^{32}P] dCTP Labelling of cDNA Probes

Linearised cDNA was labelled using the DECAprime II Random Priming DNA Labeling Kit (1456, Ambion). Briefly, template cDNA (25ng) was denatured at 100°C for 3-5 min in the presence of 10x decamer solution (2.5 µL), then snap frozen on dry ice. The denatured DNA was thawed, then labelled with 5 µL [α-^{32}P] dCTP using 1 µL Exo-Klenow (5 U/µL) for 10 min at 37°C. Unincorporated radionucleotide was removed using a ProbeQuant G-50 MicroColumn (27-5335-01, GE Healthcare). Before use, the MicroColumn containing Sephadex G-50 DNA Grade F resin, was resuspended by vortexing the inverted column. The cap was loosened a quarter turn, the column placed in a 2 mL microfuge tube, centrifuged for 1.5 min at 735 *g* and the buffer discarded. The column was placed in a new microfuge tube, the labelled cDNA probe applied drop-wise to the top of the resin column, spun for 3 min at 735*g* and the eluate containing purified-probe collected. Percentage incorporation of radionucleotide wascrudely measured by comparing the counts of the eluate to those of the used column.

Membrane Hybridisation

Membranes were pre-hybridised for 1-4 hrs at 42°C in deionised formamide-based cDNA hybridisation buffer (25mL, Appendix II). Purified, radiolabelled probe was diluted 1:40 in (1 mL) of hybridisation buffer, carefully added to the hybridisation bottle and incubated overnight at 42°C. Membranes were washed using high stringency conditions (0.1x SSC, 0.1% SDS, Appendix II). Briefly, membranes were rinsed twice at RT, then washed twice at 60°C for 15 min, rinsed in RT 0.1x SSC, 0.1% SDS, wrapped in cling film and sealed into plastic pockets. For stripping of probes, membranes were incubated for two 15 min periods with boiling 1x SSC, 0.1% SDS on a shaking platform, then rinsed with RT 1x SSC, 0.1% SDS.

Quantitation of mRNA Expression

Hybridised membranes were exposed to a General Purpose Storage Phosphor screen 20 x 25 cm (63-0034-86, Molecular Dynamics/GE Healthcare) and then scanned on a Typhoon 8600 Scanner (Molecular Dynamics/GE Healthcare). The phosphorimager screen provides several advantages over x-ray film, as it provides around ten times greater sensitivity and readings can be obtained over several orders of magnitude (Johnston *et al.*, 1990). In contrast, x-ray film provides only a linear range of detection. Scans were acquired using the "Storage Phosphor" option at a pixel resolution of 200 μm using the phosphorimager filter 390BP filter/Red 633 nm. ImageQuant Version 5.2 (GE Healthcare) software was used to quantitate relative gene expression levels.

Microarray Analysis of the Tammar Wallaby Mammary Gland

The experiment design for the microarray analysis of the tammar wallaby mammary gland throughout pregnancy, lactation and involution, plus the virgin gland (Figure 2.5) was conceived by Dr. Matthew Digby, a member of the lab group. Microarray slides were printed by the Peter MacCallum Cancer Institute, Melbourne, Australia. Complimentary DNA probes (10,368 spots, 9984 genes) spotted on to the microarray slides were generated from a normalized mammary gland cDNA library commercially made by Invitrogen Corporation, USA.

The library was prepared by pooling tammar mammary gland total RNA from various time points in pregnancy (P), lactation (L) and involution (I): d26P, d55L, d87L, d130L, d180L, d220L, d260L and d5I. Briefly, first-strand cDNA was carried out with Superscript II and oligodT with a *Not*I adapter (Figure 2.6 A). The resulting cDNA fragments were blunt-ended at the opposite end with T4 DNA polymerase and directionally-cloned into an acceptor vector with a 5' blunt-ended *Eco*RV site and a 3' *Not*I site (Figure 2.6 B). The cDNA inserts, with an average size of ~1.2 kb, were transferred into the pCMVsport6.ccdb mammalian expression vector, by 'swapping out' the region encompassing the *att*B1 and *att*B2 sites and the ccdb gene, which is obscured (Figure 2.6 B and C). Clones containing cDNA inserts were selected for, with those retaining the ccdb gene discarded. The library was normalized against α-casein and β-lactoglobulin, two highly expressed milk protein genes.

The tammar library was plated, with individual clones picked, grown and plasmid DNA isolated. Clones were sequenced with the SP6 primer i.e. sequenced from 5' end of the clone at VIAS, DPI, Attwood, Victoria. The chromatograms were processed with Phred/Phrap (Ewing and Green, 1998; Ewing *et al.*, 1998), subjected to Basic Local

Alignment Search Tool (BLAST) searches; blastn (nucleotide), blastx (protein) and tblastn (translated nucleotide) searches of the Genbank http://www.ncbi.nlm.nih.gov/BLAST/), Unigene and Swiss-Prot databases, the best hit identified and compiled into the CRC-IDP Database. An e-value of 1x e^{-8} was used as a cut-off for alignment and clones with best hit values above this range were designated "unknown".

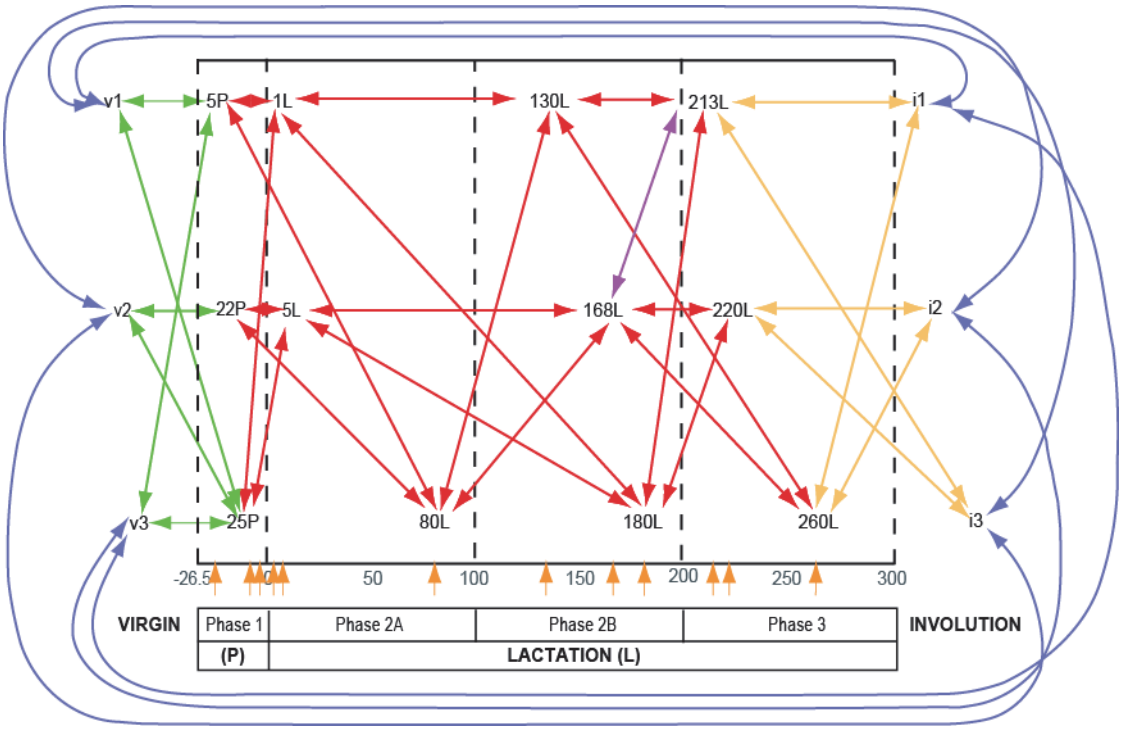


Figure 2.5. cDNA microarray analysis of the tammar wallaby mammary gland from the virgin gland (V), through to pregnancy (P), lactation (L) and involution (I)

Custom-made tammar cDNA microarray slides, each with 10,368 spots encoding 9,984 genes, were used to analyse changes in mammary gene expression during the different reproductive states of the mammary gland. A total of 72 slides, including the red/green dye swaps were used for the analysis, with each comparison made indicated (arrows).

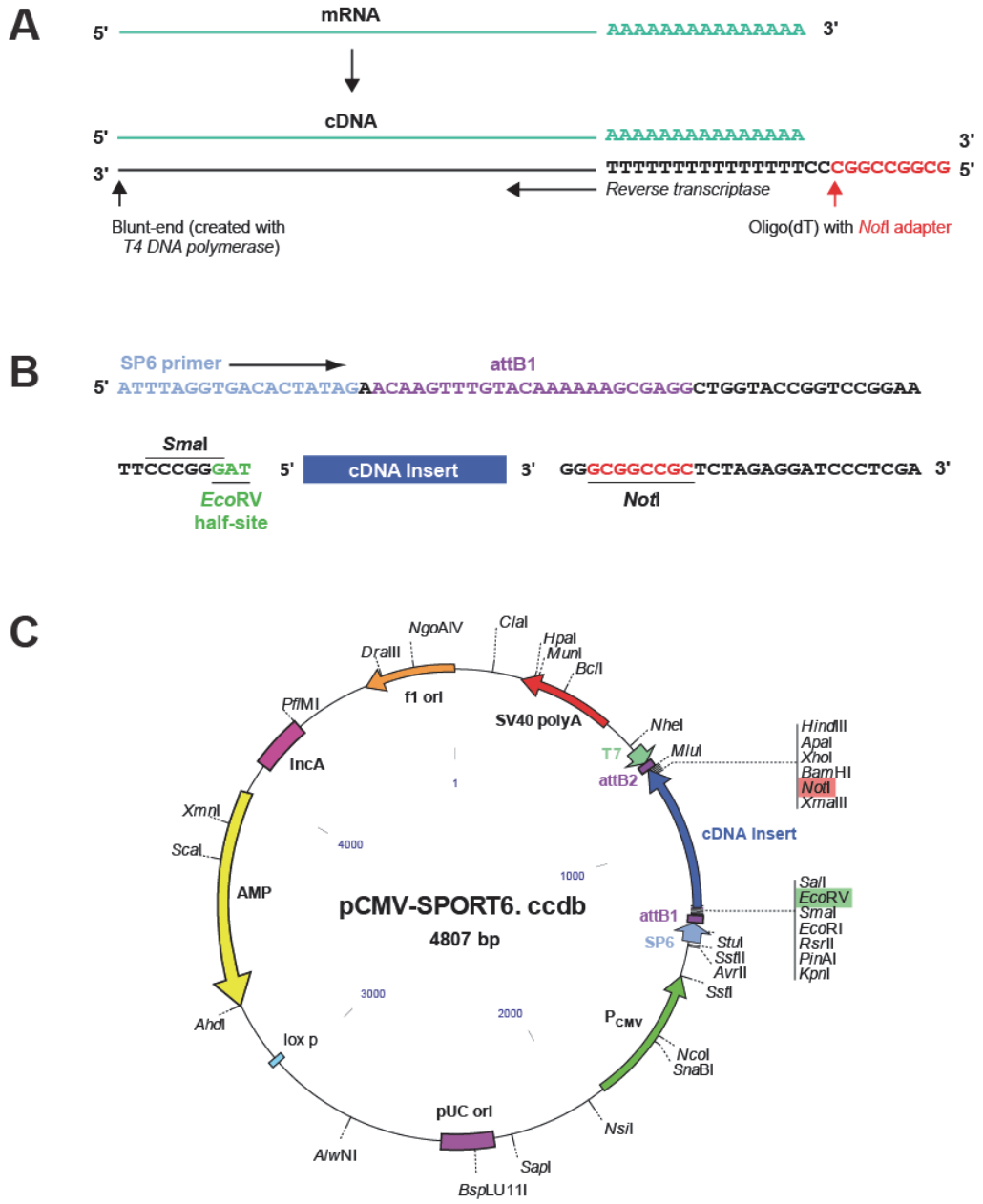


Figure 2.6. Creation of the commercially prepared tammar wallaby mammary gland cDNA library in the pCMV-SPORT6.ccdb mammalian expression vector (Invitrogen)

A. Mammary transcripts from different stages of lactation were reverse transcribed using an oligo(dT) primer with a *Not*I adapter at the 3'-end, with the opposite ends of the cDNA blunt-ended with T4 DNA polymerase. **B.** cDNA inserts were directionally-cloned into the EcoRV-cut/NotI-digested acceptor vector. Inserts were 'swapped into' the pCMV-Sport6.ccdb vector via the attB1 and attB2 sites and cDNA inserts sequenced from the 5'-end with the SP6 primer. **C.** Transcription of mammary cDNA inserts within the ampicillin-resistant pCMV-SPORT6.ccdb vector is under the control of the Cytomegalovirus (CMV) promoter.

Labelling of Microarray Probes

Total RNA (5 µg), anchor primer (2µL) and water were combined to a volume 20.9 µL and incubated at 70°C for 10 min. M-MLV Reverse Transcriptase, RNase H Minus, Point Mutant (M3682, Promega) and Aminoallyl-dUTP sodium salt (0.6µL, A0410 , Sigma) were added. The sample was incubated at 42°C for 2.5 hr, hydrolysed by the addition of 0.25 M NaOH (10 µL), 0.5M EDTA (5 µL) and 0.2 M acetic acid (15 µL) and incubation at 65°C for 15 min. The samples were purified (QIAquick PCR Purification Kit; 28104, Qiagen). Briefly, Buffer PB (5 vol) was added to the sample, mixed and applied to a spin column, centrifuged at 13,000 *g* for 1 min and the eluate discarded. The column was washed twice with Buffer PE (750 µL) by centrifugation at 13,000 *g* for 1 min and the supernatant discarded after each step. The column was dried by an additional 1 min centrifugation step at 13,000 *g*.

The Cy_3 and Cy_5 dyes in the CyDye Post-Labelling Reactive Dye Pack (RPN5661, Amersham Biosciences/GE Healthcare) were resuspended in 20 µL 0.1 M sodium bicarbonate buffer, pH 9.0 (Appendix II) and applied to the spin column containing the amino-allyl labelled probe. The columns were incubated in the dark at RT for 1 h and the labelled cDNA eluted with RNase-free water (80 µL) by centrifugation at 13,000 *g* for 1 min. The samples were then purified twice using the QIAquick PCR Purification Kit as described previously, but using new spin columns. The bound probes washed twice with a reduced volume of Buffer PE (500 µL). The labelled probes were eluted with Buffer EB (500 µL), a 1 min incubation and centrifugation at 13,000 *g* for 1 min. These steps were repeated for the second purification step to remove any unincorporated dye.

Hybridisation of Microarray Slides

Microarray slides were baked at 80°C to fix the cDNA to the slides, then pre-hybridised in a coplin jar (or in a humidified hybridisation chamber) at 42°C for 45 min. The slides were then immersed in MQ H_2O, rinsed in isopropanol and dried by centrifugation at 130 *g* (Eppendorf 5810R centrifuge, rotor A-4-62). The microarray probes (Cy_3 and Cy_5 labelled samples for comparison) were combined and the following components added: 3 µL tRNA 4 mg/mL (R8759, Sigma), 3 µL Cot-1 DNA 10 mg/mL (18440-016, Invitrogen), 0.75 µL Poly $dA_{(50)}$ 8 mg/mL (Geneworks), 0.75 µL of 50x Denhardt's solution with herring sperm DNA (Appendix II). The volume was reduced to 10.3 µL in a CHRIST RVC 2-25 speedy-vac at 60°C for 20 min and then 20x SSC (4.7 µL) and deionised formamide 100% (15 µL) added. The samples were denatured at 100°C for 3 min; 10% SDS (0.3 µL) added and the probe applied to a clean coverslip,

which was then applied to the microarray slide. Hybridisation was carried out overnight at 42°C in a hybridisation chamber. The following day, hybridised microarray slides were immersed in Wash Buffer 1 (Appendix II) in a coplin jar to remove the cover slips. The slides were washed in a light-proof box on a shaking platform for 1 min in Wash Buffer 1, 3 min in Wash Buffer 2 (Appendix II), 3 min in Wash Buffer 3 (Appendix II), dried by centrifugation at 130 *g* (Eppendorf 5810R centrifuge, rotor A-4-62) and stored in a light-proof box.

Analysis of Microarray Slides

Microarray slides were scanned with an Agilent scanner and the images analysed with Versarray software (Bio-Rad) by Dr. Matthew Digby. Microarray analysis of differentially expressed genes was initially performed using GeneSpring analysis software (Agilent Technologies) and also using R 3.0.1 (http://www.r-project.org/) and the Bioconductor package (http://www.bioconductor.org/).

For the large microarray project, all data were analysed by a statistician, Peter Thomson (University of Sydney), as described (Thomson, 2006). Briefly, Lowess normalization was used to normalize data both within and between microarray slides (Yang *et al.*, 2002). The complete data set of 36 comparisons, involving 72 slides, which included the Cy_3/Cy_5 dye-swap experiments (a total of 718,848 red and green intensity pairs) was analysed simultaneously using a large-scale, linear mixed-model, which included random effects to account for the microarray experiment design, plus gene effects and gene-contrast effects (Thomson, 2006).

The Best Liner Unbiased Predictors (BLUPs) of gene effects were fitted to a mixture model: $\pi_1 N(0, \sigma^2_g+\sigma^2_\varepsilon)$ DE genes + $(1-\pi_1)$ $N(0, \sigma^2_\varepsilon)$ non-DE genes (where π_1 is the probability of a gene being differentially expressed (DE), σ^2_g is the variance of DE gene expression and σ^2_ε is the residual variance for both differentially and non-differentially expressed (non-DE) genes (Thomson, 2006)). The model was fitted using the EM algorithm, with probabilities of genes being DE determined based upon a significance level of $p<0.001$ and their estimated effects (direct and indirect) also determined. Clustering of genes with similar expression patterns was performed with either Cluster 3.0, an enhanced version of Cluster - originally developed by Michael Eisen (Eisen *et al.*, 1998), or using the Mfuzz program within the Bioconductor package (http://www.bioconductor.org/) (Gentleman *et al.*, 2004; Reimers and Carey, 2006) of the open-source R statistics package (http://www.r-project.org/).

Protein

Extraction of Whey from Whole Milk

Milk was thawed from -80°C, mixed, diluted 1:2 with 1x PBS and centrifuged at 18,000 *g* for 15 min at RT. The top layer of fat was removed and the whey collected into a new tube, leaving the casein pellet and cell debris. The whey was centrifuged for a further 20 min at 18,000 *g* at 4°C, transferred to a new tube and stored at -20°C.

Determination of Protein Concentration

Protein concentrations were determined using the trichloroacetic acid (TCA)-ponceau S assay (Pesce and Strande, 1973), but using the modifications made by Dr Kylie Cane for use in a 96-well plate assay format (Cane, 2005). A bovine serum albumin (BSA) standard solution (2 mg/mL) was prepared and subsequently diluted with MQ H_2O to a working range of 0-2 mg/mL. All samples and standards were prepared in triplicate. Briefly, a 10 µL aliquot (sample or BSA standard) was mixed with 100 µL TCA-Ponceau-S working reagent (8 mg ponceau-S dye in 100 mL of 3% w/v trichloroacetic acid solution). The 96-well plate was centrifuged at 3000 *g* for 10 min (Eppendorf 5810R centrifuge, rotor A-4-62), the supernatant removed and the plate inverted for 10 min to remove any liquid. Sodium hydroxide (2 M, 200 µL) was added to each well and mixed to dissolve the precipitate. The absorbance of the solution was quantitated at A_{540nm} in a microtitre plate reader Multiskan Ex (Thermo Electron Corporation, Vantaa, Finland).

1D Gel Electrophoresis of Proteins (SDS-PAGE)

SDS-PAGE was performed using the discontinuous buffer system described by (Laemmli, 1970) and the Bio-Rad Mini-PROTEAN 3 Cell instruction Manual. The protein sample, with SDS-PAGE sample buffer (Appendix II) added was heated at 95°C for 5 min and placed on ice for 2 min. Denatured samples were electrophoresed in a polyacrylamide gel; ranges of 10-20% commonly used (Sambrook and Russell, 2001) together with a lane of protein standard 4-250 kDa, SeeBlue Plus2 Pre-Stained Standard (LC5925, Invitrogen Corporation). Samples were electrophoresed for 40-50 min at 200 V in 1x Tris-glycine SDS electrophoresis buffer (Appendix II) and then stained either using Coomassie brilliant blue R250, (161-040, Bio-Rad; Appendix II) or the more sensitive silver-staining method. Proteins were transferred to nitrocellulose membranes by Western blotting (Methods section, Chapter 7).

Staining of SDS-PAGE Gels

Coomassie Staining

SDS-PAGE gels were rinsed in MQ H_2O and stained in Coomassie brilliant blue (Appendix II) for 2 h on a shaking platform at 180 rpm. Gels were rinsed in destain: 5% methanol (v/v), 7% acetic acid (v/v) in MQ H_2O (Appendix II), then destained overnight (until clear) on a shaking platform. Gels were rinsed in MQ H_2O before being dried down between cellophane sheets (Z377600-1PAK, Sigma-Aldrich). Gels were stained using the EZBlue Gel Staining Reagent (G-1041, Sigma) for identification using mass spectrometry. Briefly, the gel was subjected to 3x 5 min rinses in MQ H_2O, stained with 20 mL EZBlue Gel Staining Reagent for 1 h and then rinsed overnight in MQ H_2O. All steps of the protocol were performed on a shaking platform.

Silver Staining

Silver staining of protein gels was performed using modified protocols (Bassam *et al.*, 1991; Merril *et al.*, 1981). All solutions used are listed in Appendix II. After electrophoresis, SDS-PAGE gels were rinsed quickly in MQ H_2O and then soaked for 30 min. in fixative (50% methanol/10% acetic acid (v/v)). Gels were immersed in 5% methanol/7% acetic acid (v/v) for 5 min and then soaked in a freshly-made 10% glutaraldehyde solution (v/v) for 30 min. They were then rinsed with water and placed in a large volume of MQ H_2O overnight. The next day, gels were soaked in fresh MQ H_2O for an additional 30 min, incubated in a fresh 0.1% (w/v) silver nitrate ($AgNO_3$) solution for 30 min and rinsed briefly in water. Gels were then rinsed twice with developer solution (3% (w/v) Na_2CO_3 and 0.05% formalin (v/v)) and stained with 20 mL of developer until protein bands were of the desired intensity. The reaction was stopped by the addition of 2.3 M citric acid (1 mL). Gels were washed in several changes of MQ H_2O, stored at RT in the dark in 0.3% Na_2CO_3 (w/v) and dried between cellophane sheets.

Cell culture

Recovery of Cells from Liquid N_2 Storage

Cryovials were placed immediately into a 37°C water bath to thaw, resuspended in the appropriate culture media (Appendix II) and centrifuged at 500 *g* for 5 min. The supernatant containing dimethylsulphoxide (DMSO, C6295, Sigma), which is toxic to cells, was removed. The cells were resuspended in fresh media, plated out in either a flask or tissue culture plate and incubated (37°C, 5% CO_2; Thermo Forma 310 Series Stericycle tissue culture incubator Thermo Electron Corporation).

Transfection of Primary Cells Isolated from the Tammar Wallaby Mammary Gland and the HC11, HEK293T and KIM-2 Immortalised Cell Lines

Primary Cells

Transfection efficiencies of primary cells isolated from 'normal' mammary gland tissue of pregnant tammars (primary mammary gland cells; PMGCs) from ~day 24 of pregnancy were tested with Lipofectamine 2000 Reagent (LF 2000, 11668-027, Invitrogen) and FuGENE 6 (1815091, Roche). Optimization of transfection conditions were undertaken using a grid design of varying amounts of plasmid DNA and transfection reagent. The pEGFP-C1 vector (6084-1, Clontech), which ubiquitously expresses enhanced green fluorescent protein (EGFP) under the control of the CMV promoter was used as a visual marker to assess transfection efficiencies. The effect of minimal (serum-free) media versus Opti-MEM I Reduced Serum Media (31985-070, Invitrogen) was also tested. Cells were plated in 24-well plates in media without antibiotics, grown to 90-95% and 50-80% confluency when transfected the next day with L2000 or FuGENE 6 respectively.

LF 2000 was diluted in Opti-MEM(50 μL) and incubated at RT for 5 min. DNA was diluted in Opti-MEM (50 μL), combined with LF 2000 and incubated at RT for 20 min. Cells to be transfected were rinsed with warm (37 ˚C) 1x PBS and either warm serum-free media or Opti-MEM added (500 μL). LF 2000/DNA reagent complexes (100 μL) were added drop-wise to each well and mixed gently by rocking the plates. Plates were incubated (37°C, 5% CO_2), transfection complexes removed after 4-6 h and replaced with culture media and transgene expression assayed 48 h post-transfection. For FuGENE 6 transfections, Opti-MEM, or serum-free media was added to a tube, then FuGENE 6, and lastly, DNA (to a volume of 100 μL). The contents were gently mixed, incubated at RT for 15 min, added to the wells and transgene expression assayed 48 h post-transfection.

Immortalised Cell Lines

Lipofectamine 2000 was used to transfect immortalised cell lines, with the amounts of DNA and LF 2000 used dependent upon the cell line. In general, 1 μg DNA and 2 μL LF 2000 were used for HC11, HEK293T and KIM-2 cells (see Chapters 6 and 7).

In silico Analyses

Bioinformatics

Many different *in silico* analysis tools were utilised as part of this research (refer also to Chapters 3-7 inclusive for chapter-specific analyses). Genomic DNA, mRNA and

protein sequence data were obtained from NCBI National Centre for Bioinformatics Institute (http://www.ncbi.nlm.nih.gov/), the Ensemble Genome Browser (http://www.ensembl.org/index.html), provided by EMBL-EBI European Molecular Biology Lab - European Bioinformatics Institute) and the Sanger Institute, ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (http://au.expasy.org/) and the MEROPS database of proteases and protease inhibitors (http://merops.sanger.ac.uk/) (Rawlings *et al.*, 2014). Contiguous sequences were assembled with CAP3 (http://pbil.univ-lyon1.fr/cap3.php, Huang and Madan, 1999), whilst genomic structure predictions, i.e. location of promoter/exons/introns/ were performed with the GENSCAN Web Server at MIT Mattachut Institute of Technology; (http://genes.mit.edu/GENSCAN.html) Burge and Karlin, 1997). Spidey (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/) Sayers *et al.*, 2009; Wheelan *et al.*, 2001), the web-based mRNA-to-genomic-alignment program was used to identify/confirm splice sites.

Sequence characterisations, manipulations and translations were performed with the Baylor College of Medicine Human Genome Sequencing Center Search Launcher (BCM HGSC; http://searchlauncher.bcm.tmc.edu/; Smith *et al.*, 1996). Signal peptide predictions (SignalP) and post-translational protein modifications (NetNGlyc, NetPhos) were predicted with programs provided by the Centre for Biological Sequence Analysis CBS), Technical University of Denmark DTU, (http://www.cbs.dtu.dk/services) Bendtsen *et al.*, 2004; Blom *et al.*, 1999; Blom *et al.*, 2004; Gupta *et al.*, 2004; Nielsen *et al.*, 1997). The location of protein domains and structural modifications were also predicted with InterPro (http://www.ebi.ac.uk/interpro/), plus programs provided within the Australian National Genomic Information Service website ANGIS; (http://www.angis.org.au/). Phylogenetic analyses were performed MEGA4 (Molecular Evolutionary Genetics Analysis; http://www.megasoftware.net/; Kumar *et al.*, 2008; Tamura *et al.*, 2007) and also with PHYLIP 3.69 (Felsenstein, 1989).

Statistics

Statistical analyses were performed with SPSS v.16 (SPSS Inc., Chicago, Illinois, USA), R 3.0.1, or using the statistical functions available in Microsoft Excel 12 (Microsoft, Redmond, Washington, USA). Refer also to Chapters 3-6, inclusive for chapter-specific analyses.

Chapter Three:

The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene

Published in BMC Evolutionary Biology : 8th June 2012

Elizabeth A Pharo, Alison A De Leo, Marilyn B Renfree, Peter C Thomson, Christophe M Lefèvre and Kevin R Nicholas (2012) The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene. *BMC Evolutionary Biology*. **12**: 80.

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene

Elizabeth A Pharo[1,2*], Alison A De Leo[1,2], Marilyn B Renfree[1,3], Peter C Thomson[2,4], Christophe M Lefèvre[1,2,5] and Kevin R Nicholas[1,2,5]

Abstract

Background: The marsupial *early lactation protein* (*ELP*) gene is expressed in the mammary gland and the protein is secreted into milk during early lactation (Phase 2A). Mature ELP shares approximately 55.4% similarity with the colostrum-specific bovine colostrum trypsin inhibitor (CTI) protein. Although ELP and CTI both have a single bovine pancreatic trypsin inhibitor (BPTI)-Kunitz domain and are secreted only during the early lactation phases, their evolutionary history is yet to be investigated.

Results: Tammar *ELP* was isolated from a genomic library and the fat-tailed dunnart and Southern koala *ELP* genes cloned from genomic DNA. The tammar *ELP* gene was expressed only in the mammary gland during late pregnancy (Phase 1) and early lactation (Phase 2A). The opossum and fat-tailed dunnart *ELP* and cow *CTI* transcripts were cloned from RNA isolated from the mammary gland and dog *CTI* from cells in colostrum. The putative mature ELP and CTI peptides shared 44.6%-62.2% similarity. *In silico* analyses identified the *ELP* and *CTI* genes in the other species examined and provided compelling evidence that they evolved from a common ancestral gene. In addition, whilst the eutherian *CTI* gene was conserved in the Laurasiatherian orders Carnivora and Cetartiodactyla, it had become a pseudogene in others. These data suggest that bovine *CTI* may be the ancestral gene of the Artiodactyla-specific, rapidly evolving chromosome 13 *pancreatic trypsin inhibitor* (*PTI*), *spleen trypsin inhibitor* (*STI*) and the five placenta-specific *trophoblast Kunitz domain protein* (*TKDP1-5*) genes.

Conclusions: Marsupial *ELP* and eutherian *CTI* evolved from an ancestral therian mammal gene before the divergence of marsupials and eutherians between 130 and 160 million years ago. The retention of the *ELP* gene in marsupials suggests that this early lactation-specific milk protein may have an important role in the immunologically naïve young of these species.

Background

Marsupials and eutherians diverged between 130 and 160 million years ago [1-3] and evolved very different reproductive strategies [4-6]. Marsupials have an ultra-short gestation ranging from 10.7 days for the stripe-faced dunnart (*Smithopsis macroura*) [7] to 38 days for the long-nosed potoroo (*Potorous tridactylus*) [8] and deliver an altricial young [5].

Organogenesis is completed after birth supported by a long and physiologically complex lactation, during which there is an increase in maternal mammary gland size and milk production, and there are dramatic changes in milk composition [5,9-13]. In contrast, eutherians have a long pregnancy during which maternal investment is high [14,15]. During eutherian lactation, milk composition remains relatively constant apart from the initial production of colostrum 24–36 hr postpartum (pp) [16].

The tammar wallaby (*Macropus eugenii*) has a 26.5-day pregnancy after embryonic diapause [17]. After giving birth, the tammar produces milk for ~300 days until the young is weaned. Phase 1 of lactation is comprised of mammary development during pregnancy and lactogenesis around parturition. At birth, the altricial young (~400 mg) attaches to one of the four teats [5,9,13,18]. Lactation proceeds only in the sucked gland, whilst the remaining three glands regress [5,9]. The young remains permanently attached to the teat from the day of birth

* Correspondence: epharo@unimelb.edu.au
[1]Department of Zoology, The University of Melbourne, Melbourne, Victoria 3010, Australia.
[2]Cooperative Research Centre for Innovative Dairy Products
Full list of author information is available at the end of the article

BioMed Central

© 2012 Pharo et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

until day 100 pp (Phase 2A) followed by detachment from the teat and a period of intermittent sucking while confined in the pouch between days 100–200 pp (Phase 2B) [5,13,18]. The final phase is from day 200 to at least day 300 when the young suckles variably and begins to graze as well as maintaining a milk intake (Phase 3) [18]. These phases are highly correlated with changes in milk composition and mammary gland gene expression [10,13,19]. Milk protein genes such as *α-lactalbumin*, *β-lactoglobulin* (*LGB*), *α-casein*, *β-casein* and *κ-casein* are induced at parturition and expressed throughout lactation, whilst others are expressed and secreted in a phase-specific manner [13]. *Early lactation protein* (*ELP*) is expressed during Phase 2A only [13,20,21], *whey acidic protein* (*WAP*) is Phase 2B-specific [22] and *late lactation protein A* and *B* are characteristic to late Phase 2B/Phase 3 and Phase 3 respectively [23,24].

The *ELP* gene was first identified in an Australian marsupial, the brushtail possum (*Trichosurus vulpecula*) [25]. *ELP* encodes a small precursor protein with a single bovine pancreatic trypsin inhibitor (BPTI)-Kunitz domain characteristic to serine protease inhibitors. ELP is secreted in milk in multiple isoforms, which include an ~8 kDa peptide and a heavily N-glycosylated protein (~16 kDa) [25]. *ELP* was later identified in the tammar [13,20,21,26], the stripe-faced and fat-tailed dunnarts (*Sminthopsis macroura* and *Sminthopsis crassicaudata* respectively) and the South American grey short-tailed opossum (*Monodelphis domestica*) [27] (Refer to Additional file 1: Table S1 for the species in which the putative functional *ELP*/*CTI* gene, transcript and protein have been identified). Marsupial *ELP* expression is limited to the early phase of lactation [13,20,21,27,28] at the time the mother produces milk for an immunologically naïve young [29,30]. During this period, the tammar young is permanently attached to the teat and protected by humoral (passive) immunity acquired from its mother's milk and its own innate immunity [18,30].

Whilst an *ELP* orthologue is yet to be identified in eutherians, tammar and possum ELP share ~37% similarity with bovine colostrum trypsin inhibitor (CTI) [20,25]. CTI was discovered by chance in bovine colostrum over 60 years ago [31]. Putative CTI proteins with trypsin inhibitor activity were subsequently isolated from colostrum of the pig [32], cat, sheep, goat, dog, reindeer, ferret and Blue fox [33], but were not found in equine colostrum [34]. These glycosylated proteins inhibited serine endopeptidases such as trypsin, pepsin and chymotrypsin [31,32,35]. However, of these putative CTI proteins, only bovine CTI has been sequenced (Additional file 1: Table S1) and found to contain a Kunitz domain which generally indicates serine protease inhibitor activity (see below) [36]. Laskowski and Laskowski [31] hypothesised that bovine CTI protected immunoglobulins against proteolysis during the crucial period of immunoglobulin transfer from cow to calf via colostrum. However, its function is yet to be determined. Although CTI and ELP are expressed in early milk, bovine CTI secretion is brief (~1-2 days) [31,37], but marsupial ELP expression is prolonged (up to 100 days pp) [20,21,25,28]. However, their secretion in milk is correlated with the period of immuno-incompetence in the young [29,31].

The Kunitz domain was thought to have evolved over 500 million years ago [38] and is now ubiquitous in mammals, reptiles, birds, plants, insects, nematodes, venoms from snakes, spiders, cone snails and sea anemones and in viruses and bacteria [39-42]. The archetypal protein of the Kunitz domain and the BPTI-Kunitz family I2, clan IB of serine endopeptidase inhibitors in the MEROPS database [43,44] is the much studied bovine pancreatic trypsin inhibitor, also known as aprotinin (reviewed in [45]). The Kunitz domain is characterised by six conserved cysteine residues which form three disulphide bonds, producing a compact, globular protein of $\alpha + \beta$ folds [43,46,47]. Serine endopeptidase inhibition occurs through the binding of the P_1 reactive site residue within the 'binding loop' of the Kunitz domain to a serine residue within the catalytic cleft of the protease [47,48]. This is a reversible, tight-binding, 1:1 interaction [44,48]. Furthermore, the Kunitz domain P_1 residue determines protease-specificity [39,47].

Since its evolution, the Kunitz domain has been incorporated into many different genes [43,44]. In general, each domain is encoded by a single exon [43,49]. Some genes encode proteins with a single Kunitz domain, e.g. *ELP*, *CTI*, *PTI*, *spleen trypsin inhibitor* (*STI*), the five *trophoblast Kunitz domain protein* genes (*TKDP1-5*) and *serine protease inhibitor Kunitz-type-3* (*SPINT3*) and *SPINT4*. These genes, apart from the TKDPs, have 3 exons. The first exon encodes the signal- and pro-peptide, the second, a single Kunitz domain and the third, a short C-terminus. However, the TKDPs have a variable number of unique N domains inserted between the signal peptide and the Kunitz domain-encoding exon [50,51]. Genes that encode multiple Kunitz domains include: *hepatocyte growth factor activator inhibitor 1* and *2*, also known as *SPINT1* and *SPINT2* respectively (two domains), *tissue factor pathway inhibitor 1* and *2* (three domains); with up to 12 domains in the Ac-KPI-1 I nematode (*Ancylostoma caninum*) protein [38,43,44]. In addition, the Kunitz domain has been integrated into multi-domain proteins, some of which include: the collagen α3(VI), α1(VII) and α1(XXVIII) chains, WFDC6 and WFDC8, amyloid beta A4 protein, α1-microglobulin/bikunin precursor (AMBP), SPINLW1 [serine peptidase inhibitor-like, with Kunitz and WAP domains 1 (eppin)] and the WAP, follistatin/kazal, immunoglobulin, Kunitz

and netrin domain containing (WFIKKN)1 and 2 proteins [39]. Furthermore, each domain within a multi-Kunitz domain protein, may exhibit different protease activity, such as for the three tandemly repeated domains within both tissue factor pathway inhibitor 1 and 2 [43,44,52].

The early lactation/colostrum-specific expression of *ELP*/*CTI* suggests these Kunitz domain-encoding genes may play an important role in the neonate. The sequencing of the tammar genome [53], in addition to the availability of numerous vertebrate genomes including one other marsupial, the opossum, a monotreme, the platypus, many eutherians, birds (chicken, Zebra finch), fish (Zebrafish, Japanese medaka, Three-spine stickleback, Tiger and Green spotted puffers), amphibian (African clawed frog) and reptile (Green anole lizard), provides an invaluable resource with which to investigate the evolution of these genes. We used a comparative genomics approach based upon bioinformatics and PCR-based cloning of cDNA and genomic DNA to characterise the marsupial *ELP* and eutherian *CTI* genes and investigate their evolutionary history.

Results

ELP/*CTI* evolved from a common ancestral gene

To determine whether the marsupial *ELP* gene was present in other species, we used multiple approaches. We cloned the *ELP* genes of the koala and fat-tailed dunnart and isolated tammar *ELP* from a genomic library. *ELP*/*CTI* transcripts were cloned from the mammary gland of the cow, opossum and fat-tailed dunnart and the dog *CTI* transcript was cloned from epithelial cells isolated from canine colostrum. We performed BLAST searches of genomic databases (Ensembl, Release 62, April 2011 [49], NCBI GenBank nr and WGS [54] and UCSC [55]), using a cut-off of E-value ≤ 1e-8 (nucleotides) and E-value ≤ 1e-17 (proteins). To further refine the identification of *ELP*/*CTI* orthologues based upon protein sequence, we also compared gene structures (where possible) to identify genes with a similar three-exon structure to *ELP*/*CTI*. Based upon these methods, no genes orthologous to marsupial *ELP*/eutherian *CTI* were present in fish (Zebrafish, Tiger and green spotted puffers, Three-spined stickleback), birds (chicken, zebra finch), amphibian (African clawed frog), reptile (Green anole lizard), monotreme (platypus), nor sea squirts, fruit fly, nematode (*Caenorhabditis elegans*) or yeast. However, many of the current genomes available provide only low sequence coverage (e.g. anole lizard, 2x; green spotted pufferfish, 2.5x; chicken, zebra finch and platypus, 6x; elephant, 7x). Many assemblies are also incomplete (contain gaps) and may contain incorrect assemblies. Hence it is possible that *ELP*/*CTI* orthologues may be identified within these genomes with future improvements in sequence coverage and assemblies.

The *CTI* gene was present in the Laurasiatherian orders Cetartiodactyla (cow, pig, common bottle-nosed dolphin) and Carnivora (dog, cat, Giant panda). However, based upon current genome assemblies, it is a pseudogene in Afrotheria, Xenarthra, Euarchontoglires and the Laurasiatherian orders Chiroptera and Perissodactyla.

The mammalian *ELP*/*CTI* gene was composed of 3 exons and 2 introns (Figure 1). The marsupial *ELP* gene ranged from ~1.4 kb for the koala to ~4.8 kb for the stripe faced dunnart, whilst eutherian *ELP* spanned from ~2.5 kb for the panda to ~3.8 kb for the pig. *ELP* exon 1 and 2 sizes respectively were highly conserved across all mammals (Figure 1). Exon 1 encoded the putative signal peptide and the first four amino acids at the N-terminus of the protein. The 216 bp exon 2 (with the exception of the koala, 210 bp) encoded the remainder of the N-terminal region, plus a single BPTI-Kunitz domain towards its 3'-end. *ELP*/*CTI* exon 3 differed most and encoded a maximum of seven amino acids. The *ELP/CTI* transcripts (putative translation start site to the polyadenylation signal, inclusive) were short. Marsupial *ELP* and eutherian *CTI* transcripts ranged from 425–447 bp and 416–428 bp respectively and shared 56.1%-63.6% similarity at the nucleotide level (Additional file 2: Figure S1; Additional file 3: Tables S2A, S2B). A highly conserved marsupial-specific region (87%-100%) was also identified within the *ELP* 3'-UTR (nt 420–475, Additional file 2: Figure S1; Additional file 3: Table S2C).

Based upon signal peptide analysis [56], the putative ELP/CTI peptides identified in this study were predicted to be secreted in milk, as for tammar and possum ELP and bovine CTI [20,25,26,31]. The mature ELP and CTI peptides shared 44.6%-62.2% similarity (Table 1; Additional file 4: Table S3A). In addition, the conservation of the two Kunitz domain motifs in all species suggested they may inhibit the S1 family of serine endopeptidases like many other members of the BPTI-Kunitz family [43,44]. The BPTI KUNITZ 2 motif [C1-C6, C2-C4 and C3-C5, Prosite: PS00280] indicates the 3 disulphide bonds which determine the structure of the domain (Figure 2). This motif spanned the entire 51 amino acid Kunitz domain (aa 23–73, C23-C73, C32-C56 and C48-C69, Figure 2). The second shorter motif BPTI KUNITZ 1 [F-x(2)-{I}-G-C-x(6)-[FY]-x(5)-C; where x represents any residue, those within square brackets are permitted, but those within curly brackets are not, Prosite: PS00280] was located within BPTI KUNITZ 2 (aa 51–69, Figure 2). A putative trypsin interaction site within the Kunitz domain (from KU NCBI cd00109) [57], is also depicted (aa 30–34, 36, Figure 2).

Conserved amino acid residues within a protein provide an indication of sites essential for its structure and

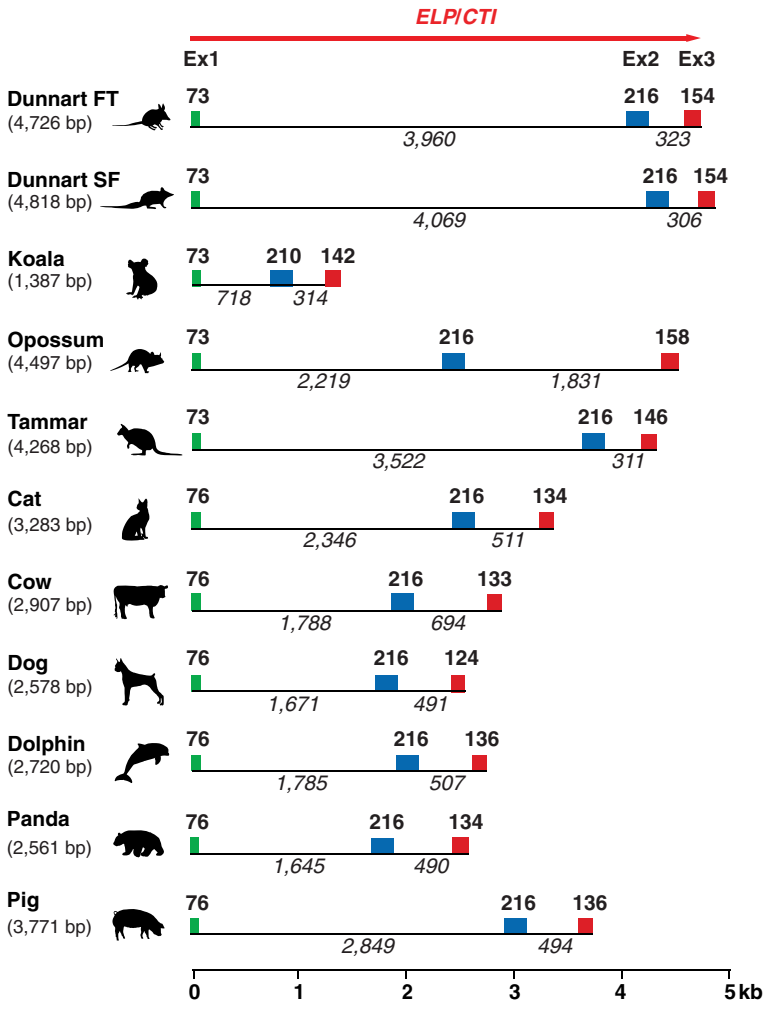


Figure 1 Structure of the marsupial *ELP* and eutherian *CTI* genes. The *ELP/CTI* genes of the stripe-faced (SF) dunnart (*Sminthopsis macroura*) [GenBank: AC186006], fat-tailed (FT) dunnart (*Sminthopsis crassicaudata*) [GenBank: JN191336], koala (*Phascolarctos cinereus*) [GenBank: JN191337], opossum (*Monodelphis domestica*) [GenBank: BK008085], tammar (*Macropus eugenii*) [GenBank: JN191335], cat (*Felis catus*, Abyssinian domestic cat) [GenBank: BK008083], cow (*Bos Taurus*, Hereford Breed) [Ensembl: ENSBTAG00000016127], dog (*Canis familiaris*, Boxer breed) [GenBank: BK008082], dolphin (*Tursiops truncatus*) [GenBank: BK008086], pig (*Sus scrofa domestica*) [Ensembl: F1SD34_PIG (ENSSSCG00000007398)] and Giant panda (*Ailuropoda melanoleuca*) [GenBank: BK008084] have 3 exons and 2 introns. Gene size is indicated within brackets and refers to the number of nucleotides from the putative translation start (ATG, exon 1) to the polyadenylation signal (AATAAA, inclusive, exon 3). Exons are colour-coded: exon 1 (green rectangle), the Kunitz domain-encoding exon 2 (blue) and exon 3 (red) and exon size is indicated in bold text. Intron sizes are italicised. The horizontal scale bar indicates the relative sizes of the *ELP/CTI* genes (kb), with the putative translation start site (ATG) of all sequences aligned with the origin (0 kb). Genes are drawn approximately to scale.

biological function. Comparison of the marsupial ELP and eutherian CTI precursor proteins showed that the signal peptide (57.1%-81.0% similarity), the 51 aa BPTI KUNITZ 2 motif (54.9%-68.6%), plus the shorter 19 aa BPTI KUNITZ 1 motif within it (63.2%-73.7%) were conserved. However, the 20–22 residue linear chain of the mature ELP/CTI N-terminus had marsupial-specific and eutherian-specific homology (59.1%-100%, Table 1; Additional file 4: Tables S3B, S3C, S3D, S3E). Conservation of the short (3–10 residue) C-terminus was variable (Additional file 4: Table S3F). This was in part due to the use of different stop codons in *ELP*/*CTI* transcripts across divergent species. The opossum and dunnart ELP proteins were truncated at the end of exon 2, with the stop codon encoded by one nucleotide in exon 2 and two in exon 3 (nt 323–325 inclusive; Additional file 2: Figure S1). For all other species, two different stop codons within exon 3 were used. For the panda, cat and dog, the TAA stop codon (nt 333–335) was used. However, for the pig, cow, dolphin and the remainder of the marsupials, the equivalent TGA stop codon (nt 344–346 inclusive) was used.

Surprisingly, there was little conservation of the amino acid residue type (physiochemical properties) at the P_1

Table 1 Homology between and within the marsupial ELP and eutherian CTI peptides[1]

Species comparisons	Signal peptide	Mature peptide	N-terminus	Kunitz motif2 (51 aa)	Kunitz motif1 (19 aa)	C-terminus
Marsupial ELP	85 - 95%	67.5 - 100%	59.1 - 100%	76.5 - 100%	84.2 - 100%	20 - 100%
Eutherian CTI	57.1 - 90.5%	70.7 - 88.6%	59.1 - 90.9%	76.5 - 94.1%	84.2- 100%	40 - 83.3%
Marsupial ELP vs Eutherian CTI	57.1 - 81.0%	44.6 - 62.2%	18.2 -59.1%	54.9 - 68.6%	63.2 - 73.7%	10 - 60%

Pairwise amino acids similarities were calculated using MatGAT 2.01 (BLOSUM62 matrix).
[1]Refer to Additional file 4: Tables S3 for individual species comparisons.

reactive site within the Kunitz domain (residue 33, Figure 2). Although the P_1 residue type (basic amino acid with a positively charged side chain) was conserved amongst eutherians: K (lysine) for the pig, cow and dolphin and R (arginine) for the cat, dog and panda, this was not so for marsupials. The opossum and possum ELP P_1 residue was acidic with a negatively charged side chain (D, aspartate). However, the P_1 residue for tammar (S, serine) and the koala and dunnarts (N, asparagine) was polar with uncharged side chains.

Although P_1 residues differed, all ELP/CTI peptides were predicted to be N-glycosylated at asparagine-42, consistent for bovine CTI [58] and therefore should be larger than their predicted masses (8.6 to 9.6 kDa, data not shown).

Selective pressure acting upon marsupial ELP and eutherian CTI

The evolutionary selection pressure acting upon different regions of the protein-coding marsupial *ELP* and eutherian *CTI* transcripts was determined by dN/dS analysis (Table 2). The dN/dS ratio measures the number of non-synonymous changes per non-synonymous site (those which produce amino acid substitutions) compared to the number of synonymous changes per synonymous site (no amino acid change) [59,60]. A ratio of $dN/dS = 1$ suggests a neutral condition, with nucleotide changes accumulating in the absence of selection pressure, i.e. both dN and dS occur at the same rates. $dN/dS < 1$ indicates purifying selection, with amino acid changes not tolerated.

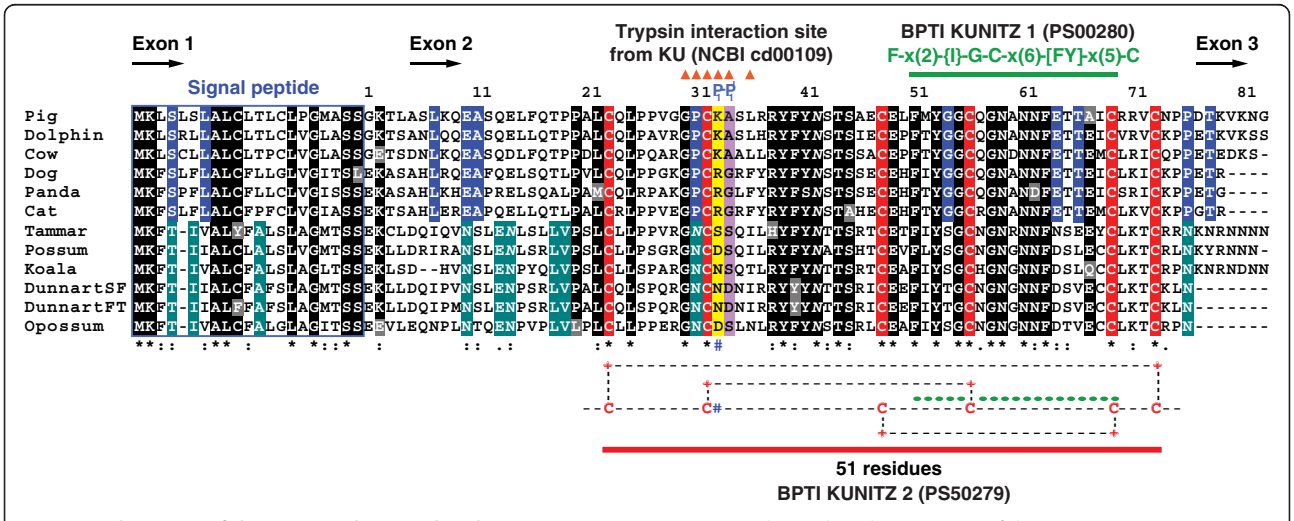


Figure 2 Alignment of the marsupial ELP and eutherian CTI precursor proteins. The nucleotide sequences of the *ELP/CTI* mRNA transcripts of the following species were conceptually translated and aligned with ClustalW2: tammar [GenBank: JN191338; UniProtKB/Swiss-Prot: O62845 (mature protein)], brushtail possum [GenBank: U34208], fat-tailed dunnart (FT) [GenBank: JN191339], opossum [GenBank: JN191340], cow (Holstein-Friesian breed) [GenBank: JN191341] and dog [GenBank: JN191342]. The stripe-faced dunnart (SF) [GenBank: AC186006], koala [GenBank: JN191337], cat [GenBank: BK008083], pig [Ensembl: F1SD34_PIG (ENSSSCT00000008098)], dolphin [GenBank: BK008086], and panda [GenBank: BK008084] *ELP/CTI* genes were conceptually spliced based upon conserved splice sites and translated. Amino acid residues are numbered based upon the start (N-terminus) of the mature ELP/CTI peptides. Black shading indicates nucleotide residues common to at least 10 of the species and grey, the remainder that differ. The six conserved cysteine residues (C1-C6, C2-C4 and C3-C5), which form the three disulphide bonds and produce a globular protein are shaded red. Teal shading indicates amino acids common to marsupials and blue, those common to eutherians. The location of exons is indicated by arrows. The predicted signal peptides are boxed (blue). The BPTI KUNITZ 1 and 2 motifs are indicated (green and red bars respectively) and the putative trypsin interaction site from the KU motif (NCBI cd00109) is depicted by orange triangles. The putative P_1 and P_1' reactive site residues are shaded yellow and purple respectively. Italicised asparagine (*N*) residues indicate predicted sites of post-translational N-glycosylation. Conservation between groups of amino acids with strongly similar properties, i.e., scoring > 0.5 in the Gonnet PAM 250 matrix is indicated (:). Conservation between groups of amino acids with weakly similar properties (scoring < 0.5 in the Gonnet PAM 250 matrix) is also noted (.). Gaps within the alignment are indicated (–).

In contrast, $dN/dS > 1$ is indicative of positive Darwinian selection for amino acid changes [59,61].

The protein-coding marsupial *ELP* and eutherian *CTI* transcripts and regions within them generally exhibited a trend towards purifying selection, with a dN/dS ratio <1 (Table 2). However, based upon codon-based Z-tests, only the eutherian *CTI* BPTI KUNITZ 1 motif (57 nt encoding 19 amino acids) was found to be undergoing purifying selection ($p < 0.05$). Although the regions encoding the marsupial BPTI KUNITZ 1 motif ($p = 0.103$) and the marsupial and eutherian BPTI KUNITZ 2 motifs ($p = 0.101$ and $p = 0.105$ respectively) exhibited a strong trend towards purifying selection, the test values ($dN < dS$) were not significant. This tendency was also consistent for the putative trypsin interaction site. In contrast, three regions of the *ELP*/*CTI* transcripts showed a trend towards positive selection ($dN/dS > 1$). These included the regions encoding the ELP/CTI N-terminus and the eutherian CTI signal peptide. However, based upon codon-based Z-tests ($dN > dS$), only the eutherian CTI signal peptide ($p < 0.05$) was undergoing positive selection.

Marsupial *ELP* and eutherian *CTI* share common flanking genes

In order to confirm that the marsupial *ELP* and eutherian *CTI* genes were orthologous, we characterised the location and arrangement of *ELP/CTI* and its flanking genes. We used fluorescence *in situ* hybridisation to map tammar *ELP* to chromosome 1q (Figure 3). The *ELP/CTI* gene was located on a syntenic segment in the marsupial (stripe-faced dunnart [27] and opossum) and eutherian genomes [49,55] and was generally flanked by one or both of the single-copy genes *phosphatidyl inositol glycan, class T* (*PIGT*) and *WAP four disulphide core domain 2* (*WFDC2*), confirming they were true orthologues (Figure 4).

The *PIGT-WFDC2* region of bovine chromosome 13 (~74.51-75.14 Mb) was unique. Bovine *CTI* was adjacent to *PIGT*, but there was an insertion of ~602 kb between the *CTI* and *WFDC2* genes [49,55] (data not shown). This region included 7 Artiodactyla-specific Kunitz domain-encoding genes including *PTI*, *STI*, plus the five placenta-specific *TKDP1-TKDP5* genes inclusive [50,63]. Furthermore, the *SPINLW1* gene which contains both a Kunitz and a WAP domain and the eutherian-specific *SPINT4* gene were located a further ~38 kb and ~90 kb respectively downstream from *WFDC2* [49,55] (data not shown). As mentioned previously, these genes, with the exception of *SPINLW1* and the *TKDP*s, share a similar 3-exon structure. However, the *TKDP*s differ due to the likely "exonisation" of an intron and its subsequent duplication to produce a variable number of tripartite N-domains between the exon encoding the signal peptide and the Kunitz domain [50,51].

CTI has been lost in some eutherians

Using the canine sequence as the basis for mVISTA comparative analysis [64], the region between the *PIGT*

Table 2 Average rates of synonymous (dS) and non-synonymous (dN) substitutions occurring in marsupial ELP and eutherian CTI

ELP/CTI protein-coding region		dN	SE	dS	SE	dN/dSRatio	(a) Neutral selection test ($dN \neq dS$)+*	(b) Purifying selection test ($dN < dS$)+*	(c) Positive selection test ($dN > dS$)+*
Precursor protein	***Marsupials***	0.145	0.022	0.190	0.033	0.763	0.256 (NS‡)	0.117 (NS)	1.000 (NS)
	Eutherians	0.194	0.026	0.225	0.033	0.862	0.232 (NS)	0.472 (NS)	1.000 (NS)
Mature protein	***Marsupials***	0.166	0.026	0.185	0.036	0.897	0.653 (NS)	0.334 (NS)	1.000 (NS)
	Eutherians	0.186	0.028	0.242	0.039	0.786	0.273 (NS)	0.130 (NS)	1.000 (NS)
Signal peptide	***Marsupials***	0.071	0.029	0.226	0.094	0.314	0.133 (NS)	0.064 (NS)	1.000 (NS)
	Eutherians	0.225	0.072	0.165	0.069	1.36	0.451 (NS)	1.000 (NS)	0.224 (NS)
N-terminus	***Marsupials***	0.240	0.064	0.116	0.048	2.07	0.064 (NS)	1.000 (NS)	0.041*
	Eutherians	0.242	0.050	0.224	0.065	1.08	0.842 (NS)	1.000 (NS)	0.424 (NS)
BTPI KUNITZ 2#	***Marsupials***	0.146	0.031	0.224	0.052	0.651	0.215 (NS)	0.101 (NS)	1.000 (NS)
	Eutherians	0.162	0.035	0.243	0.054	0.667	0.200 (NS)	0.105 (NS)	1.000 (NS)
BPTI KUNITZ 1~	***Marsupials***	0.095	0.030	0.223	0.098	0.426	0.212 (NS)	0.103 (NS)	1.000 (NS)
	Eutherians	0.066	0.026	0.264	0.110	0.250	0.122 (NS)	0.046*	1.000 (NS)
Trypsin interaction site^	***Marsupials***	0.230	0.136	0.323	0.181	0.712	0.740 (NS)	0.363 (NS)	1.000 (NS)
	Eutherians	0.175	0.093	0.228	0.131	0.768	0.689 (NS)	0.345 (NS)	1.000 (NS)

#PS50279 153 nt, 51 aa.
~PS00280 57 nt, 19 aa.
^18 nt, 6 aa site from KU (NCBI cd00109).
+Codon based Z-tests in MEGA5.
*$p < 0.05$.
‡ *NS* not significant.

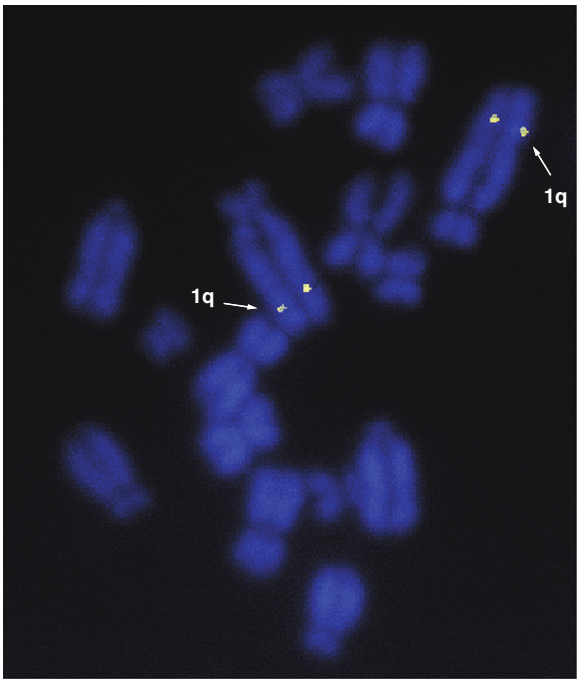


Figure 3 Localisation of the tammar *ELP* gene to *Macropus eugenii* chromosome 1q using FISH.

and *WFDC2* genes was examined using the available genome assemblies - which have variable sequence coverage, contain gaps and may contain misassembled sequences. Whilst the *ELP*/*CTI* gene was present in some mammals, it appeared to have become a disrupted pseudogene in others such as the African Savanna elephant and human (Figure 5). Exon 1 of the elephant and human *CTI* genes (signal- and pro-peptide) was present, but exon 2 (Kunitz domain) and exon 3 (C-terminus) were absent (red boxes, Figure 5), suggesting they had been excised or transposed, whilst the horse and mouse *CTI* genes initially appeared intact.

A closer examination of the nucleotide sequence between *PIGT* and *WFDC2* in these and other species using the Ensembl and UCSC genome databases revealed that different mutations had most likely disrupted the *CTI* gene. Exon 1 was disrupted in the elephant, Hoffmann's two-toed sloth (*Choloepus hoffmanni*), armadillo (*Dasypus novemcinctus*), human and other primates and horse, with exon 2 (Kunitz domain) also excised for these species, apart from the horse. Additional file 5: Figure S2A (i) depicts a nucleotide alignment of the functional/protein-coding dog *CTI* exon 1 compared with the putative disrupted *CTI* exon 1 of the elephant, sloth, human and horse. Additional file 5: Figure S2A (ii) shows the translated sequences to highlight mutations and/or deletions within the signal peptide region of CTI. The deletion of two nucleotides within human *CTI* exon 1 would produce a frame-shift (as depicted by the +1 and +2 reading frames). *CTI* exon 2 of the mouse, rat, large flying fox (*Pteropus vampyrus*) and horse also appeared to have been disrupted by deletions resulting in frame-shifts when compared to the functional/protein-coding dog *CTI* exon 2. The disruption of the protein-coding region of equine *CTI* exons 1 and 2 by at least one mutation and one deletion respectively would produce a frame-shift, suggested these were a recent occurrence (Additional file 5: Figure S2B (ii)).

Transposable elements within the *ELP/CTI* genes

Transposable elements integrate randomly into the genome, so the probability of the same element(s) integrating independently into orthologous positions in different species is extremely low. They therefore act as genetic markers and can be used to determine the phylogenetic relationship between genes and species [65]. Further evidence that marsupial *ELP* and eutherian *CTI* evolved from a common ancestral gene was provided by CENSOR retrotransposon analysis [66] (Additional file 6: Figure S3). Retroelements of conserved fragment size and orientation were located within the *PIGT-ELP/CTI* region. However, the elephant and human which appear to have lost *CTI* exons 2 and 3, had also lost retrotransposons in the corresponding region, but gained a MER5A element.

Bovine *CTI*, *PTI*, *STI* and the *TKDPs* share a common ancestral gene

The location of the 8 Kunitz-domain encoding genes (including *CTI*) on bovine chromosome 13 between the *PIGT* and *WFDC2* genes and the Artiodactyla-specific distribution of *PTI*, *STI* and *TKDP1-5* (cow and sheep [51,63]) suggested they may have evolved from *CTI*. This hypothesis was supported by phylogenetic analysis of the protein-coding regions of the mammalian *ELP*/*CTI*, bovine *PTI*, *STI* and *TKDP1-5* transcripts, with bovine *SLPI* used as an outgroup root (*SLPI* omitted, Figure 6). Several different methods in PHYLIP were used to determine the evolutionary relationships. These included the character-based maximum-likelihood (with/without a molecular clock) and maximum parsimony, as well as distance-based analysis (Fitch-Margoliash tree method using the Kimura distance model of nucleotide substitution). Trees were evaluated using the bootstrap method (100 replicates). Of the algorithms used, the maximum likelihood method using a molecular clock assumption, which assumes a constant evolutionary rate for all species, produced a tree with the highest bootstrap values. Huttley and colleagues [67] have shown that the eutherian nucleotide substitution rates are ~30% slower than for marsupials. However, all methods produced consensus trees which consistently separated the 19 sequences into the two groups depicted (Figure 6). The hypothesis that bovine *CTI* was the ancestral gene for

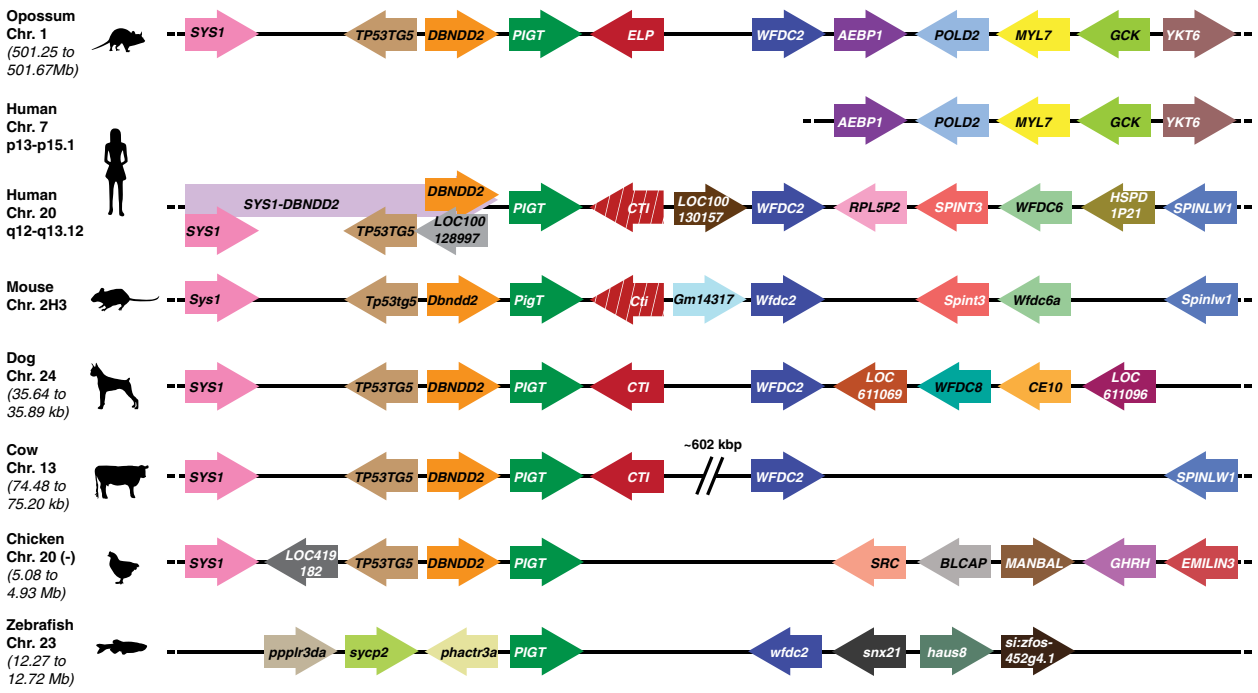


Figure 4 Chromosomal location of the *ELP/CTI* gene in different species. The *ELP/CTI* gene was located within a syntenic block on opossum Chr. 1 (~501.34 Mb), human Chr. 20q12-13.12, mouse Chr. 2 H3, dog chr. 24 (~35.7 Mb) and cow Chr. 13 (~74.5 Mb) [49,55]. However, *ELP/CTI* was reduced to a pseudogene in the human and mouse (red arrow, white diagonal stripes) and was absent in the chicken and zebrafish. The *ELP/CTI* gene was located on the reverse strand and was generally flanked by one, or both of the single-copy genes *PIGT* and *WFDC2*. The region upstream of *PIGT* was conserved in mammals and the chicken and included the *SYS1* [Golgi-localized integral membrane protein homolog (*S. cerevisiae*)], *TP53TG5* (*TP53-target gene 5 protein*), and *DBNDD2* [*dysbindin* (*dystrobrevin binding protein 1*) *domain containing 2*] genes. However, a chromosomal breakpoint was located downstream from the eutherian *WFDC2* gene. Opossum chromosome 1 contained the *AEBP1* (*Adipocyte enhancer binding protein 1*), *POLD2* [*polymerase (DNA directed), delta 2, regulatory subunit 50 kDa*], *MYL7* (*myosin, light chain 7*, regulatory) and *YKT6* [*YKT6* v-SNARE homolog (*S. cerevisiae*)] genes and was orthologous to human chromosome 7p13-p15.1. In contrast, the eutherian chromosomes contained a number of genes which encoded Kunitz and/or WAP domains. These included *SPINT3*, *SPINLW1*, *WFDC8* and *WFDC6*, which were likely to have arisen by gene and domain duplications [62]. Notably, there was an insert of ~602 kb between bovine *CTI* and *WFDC2*. Arrows indicate the arrangement and orientation of genes and are not drawn to scale.

\bovine *PTI*, *STI* and *TKDP1-5* was supported by both an alignment of precursor proteins and phylogenetic analysis of *CTI*, *PTI*, *STI*, *TKDP1-5* and the *SPINT4* protein-coding transcripts (Additional file 7: Figure S4; Additional file 8: Figure S5). Interestingly, the size of the Kunitz domain-encoding exon varied. Whilst the bovine *CTI* exon was 216 bp, those of the TKDPs were 196 bp, with 192 bp for *PTI* and *STI* and 175 bp for *SPINT4*. Furthermore, apart from CTI and SPINT4, none of the Kunitz domains were predicted to be N-glycosylated. Additional evidence of the evolutionary history of the *CTI*, *PTI*, *STI* and *TKDP1-5* genes was provided by mVISTA (Additional file 9: Figures S6A and S5B (i-viii) and CENSOR analysis (Additional file 10: Figure S7; Additional file 11: Table S4).

Tammar *ELP* expression is up-regulated at parturition and is mammary-specific

Northern analysis showed that tammar *ELP* was up-regulated at parturition, consistent with brushtail possum *ELP* [28] (Figure 7A). *ELP* transcripts were detected in the tammar mammary gland from ~ day 17 of pregnancy onwards, throughout early lactation (Phase 2A) until ~ day 87 of lactation. *ELP* was then down-regulated to minimal levels for the remainder of lactation. This was consistent with a previous study of late Phase 2A/Phase 2B mammary tissues, but the precise timing of *ELP* gene induction was not investigated [13,20,21]. Neither *ELP*, nor *LGB* was expressed in the virgin mammary gland and both genes were down-regulated postpartum in the non-sucked glands (Figure 7A), as in the brushtail possum [28].

LGB expression peaked in the mammary gland during Phase 3, consistent with [68].

Although cDNA microarray analysis of the tammar mammary gland (Figure 7B; Additional file 12: Table S5) was based upon comparative expression levels rather than actual transcript levels, the data was consistent with quantitative analysis of the Northern blot (data not shown) and microarray data reported by [69]. Lastly, Northern analysis of assorted tammar tissue samples indicated that expression of

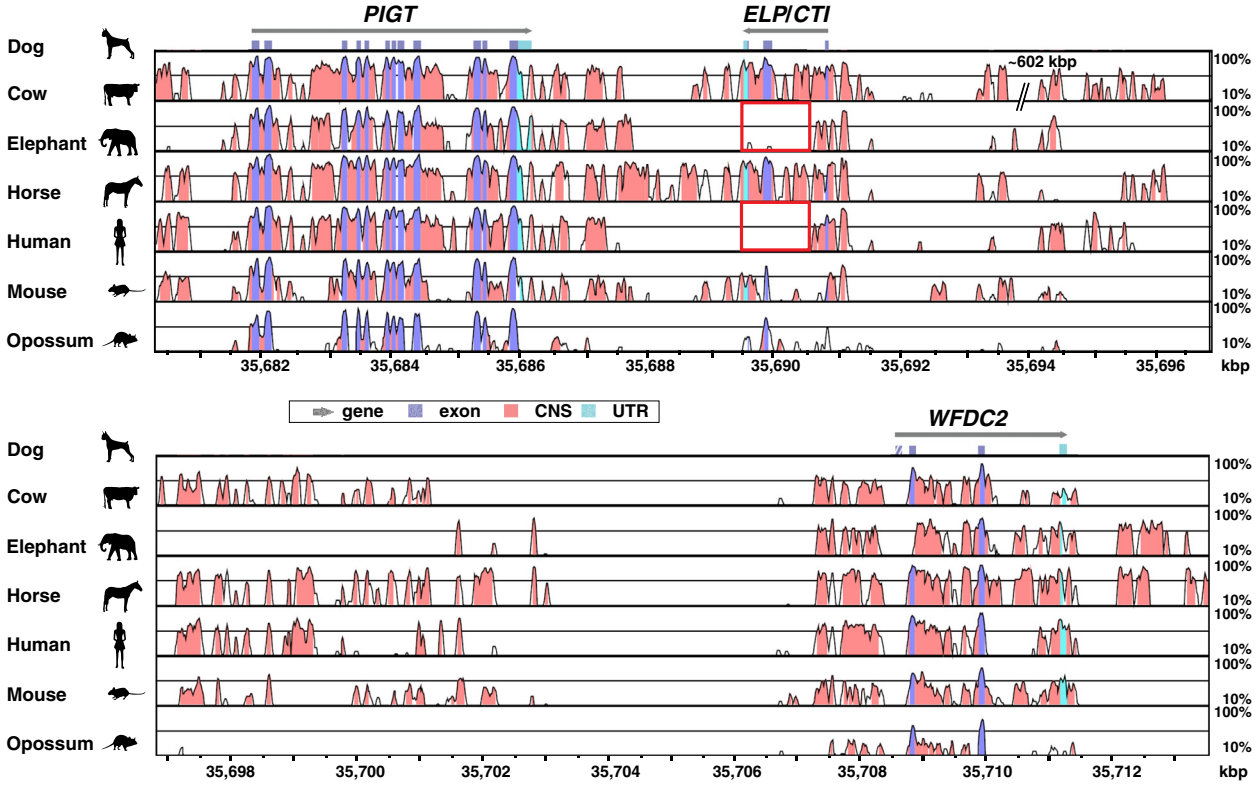


Figure 5 VISTA plot of pairwise alignments for selected mammals in the region containing the *PIGT*, *ELP/CTI* and *WFDC2* genes. Sequence homology within the *PIGT-ELP/CTI-WFDC2* region of the dog, cow, elephant, horse, human, mouse and opossum genomes was determined with mVISTA [64]. The dog sequence was used as the reference sequence (horizontal axis, dog chromosome 24 numbering). Grey horizontal arrows indicate gene location and direction of transcription. Blue rectangles indicate coding exons and untranslated regions (UTRs) of the gene are depicted by light green rectangles. Exon 1 of canine *WFDC2* was missing (gap in the current assembly) from the dog genome and is indicated by a blue rectangle with diagonal white stripes. The right axis indicates the percentage identity within a 100 bp window for each pairwise comparison, ranging from 10% to 100%. Regions sharing greater than 25% identity are shaded and the black horizontal line indicates 70% identity. The region containing the Kunitz domain-encoding *ELP/CTI* exon 2 was conserved in the cow, horse, mouse and opossum, but was absent in the elephant and human *CTI* genes (red boxes).

ELP and *LGB* was mammary gland-specific (Figure 8), unlike the ubiquitously expressed *cystatin C* (*CST3*) gene (data not shown).

Discussion

ELP was originally thought to be a marsupial-specific gene [19]. However, we have shown that the marsupial *ELP* and eutherian *CTI* genes evolved from a common therian ancestral gene (Figure 9). Mammalian *ELP/CTI* was generally flanked by one or both of the single copy *PIGT* and *WFDC2* genes in a region that was syntenic to that of other mammals. The conserved genomic structure of 3 exons and 2 introns and homologous transposable element fragments confirmed that *ELP* and *CTI* were true orthologues. *CTI* was also identified as the putative ancestral gene of the ruminant-specific *PTI*, *STI* and *TKDP1-5* genes. Based upon current genome sequencing and assemblies, *ELP/CTI* was not found in birds, fish, reptiles, nor amphibians, suggesting the gene was present in the therian ancestor before the divergence of marsupials and eutherians at least 130 million years ago [1,2,70].

Mammalian *ELP/CTI* and the evolution of bovine *PTI*, *STI* and the *TKDPs*

The Kunitz-type inhibitor domain has been duplicated many times throughout evolutionary history [38]. This was no more evident than for the region of bovine chromosome 13 on which *CTI* and the 7 *CTI*-like genes were located. The *PTI*, *STI* and *TKDP1-5* genes were specific to the order Cetartiodactyla, sub-order Ruminantia [50,51,63,72], strong evidence they evolved from *CTI* after the divergence of the Ruminantia ~25-35 MYA [1]. The *CTI*, *PTI* and *STI* genes had a similar 3-exon structure and conserved regions within both coding and non-coding segments. The *PTI* and *STI* genes and proteins were homologous and almost certainly arose by gene duplication [73]. However, the *TKDP1-5* genes had one or more additional exons inserted between

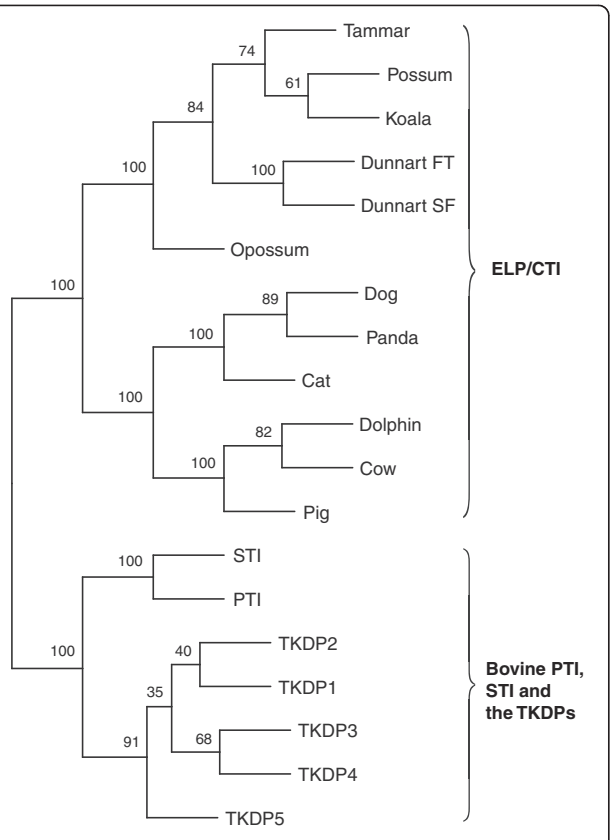


Figure 6 A phylogenetic tree of *ELP*/*CTI* and the *CTI*-like bovine *PTI*, *STI* and *TKDP1*, *2*, *3*, *4* and *5* family. The evolutionary relationship between the protein-coding regions of the marsupial *ELP*, eutherian *CTI* and bovine *TKDP1-5*, *PTI* and *STI* transcripts was determined by maximum likelihood analysis using a molecular clock assumption. The bovine *SLPI* transcript was used as an outgroup (data not shown). Two main groups were formed: 1. mammalian *ELP/CTI* and 2. bovine *CTI*, *PTI* and the *TKDP*s. Numbers at branch points indicate confidence levels as determined by bootstrap values (100 replicates). Phylogenetic trees were produced with Phylip software version 3.69. Transcripts were aligned with MUSCLE and boostrapped values generated with SEQBOOT. Maximum likelihood trees were generated with DNAMLK using a transition/transversion ratio of 1.34, a Gamma distribution shape of 1.39 with 5 Hidden Markov Model categories, global rearrangements and with a randomised input order jumbled once. The protein-coding regions of the following transcripts were used in the analysis: *ELP/CTI*, tammar [GenBank: JN191338], fat-tailed dunnart [GenBank: JN191339], stripe-faced dunnart [GenBank: AC186006], koala [GenBank: JN191337] opossum [GenBank: JN191340], brushtail possum, cow [GenBank: JN191341], dog [GenBank: JN191342], cat [GenBank: BK008083], pig [Ensembl: F1SD34_PIG (ENSSSCT00000008098)], Giant panda [GenBank: BK008084], and Common bottlenose dolphin [GenBank: BK008086], and the following bovine transcripts: *PTI* [GenBank: NM_001001554], *STI* [GenBank: NM_205786], *TKDP1* [GenBank: NM_205776], *TKDP2* [GenBank: NM_001012683], *TKDP3* [GenBank: XM_584746], *TKDP4* [GenBank: NM_205775], and *TKDP5* [GenBank: XM_614808] and *SLPI* [GenBank: NM_001098865].

the signal- and pro-peptide-encoding and Kunitz domain-encoding exons (equivalent to intron 1 of *CTI*, *PTI* and *STI*) resulting in an expansion to 4 (*TKDP5*), 6 (*TKDP2*, *3* and *4*) and 12 exons (*TKDP1*) [50,51,72]. These added exons encode tripartite N-domains which had no similarity to database sequences or motifs and evolved recently due to the "exonization" of an intron within an active MER retrotransposon and its subsequent duplication [50,63]. These elements have been associated with genetic rearrangements and deletions [74]. This may explain the excision of *CTI* exons 2 (Kunitz domain) and 3 (C terminus) for the elephant and primates, based upon current genome sequencing and assemblies.

Lack of conservation of the ELP/CTI putative P_1 reactive site residue

All putative ELP/CTI peptides were predicted to be secreted and shared a conserved single 51 amino acid Kunitz domain. The conserved location of the 6 cysteine residues which form three disulphide bonds suggested ELP/CTI would, like bovine CTI [75] and PTI [46] form a globular protein. However, neither the identity, physiochemical properties of the ELP/CTI P_1 reactive site residue, the trypsin interaction site, nor the N- and C-terminus of the proteins were conserved. The P_1 "warhead" residue plays an essential role in the interaction of a Kunitz inhibitor domain with a serine protease and a P_1 mutation may alter the protease specificity of the Kunitz domain to a particular substrate and the reaction kinetics [48,76]. Kunitz inhibitors with a basic residue, K (Cetartiodactyla) or R (Carnivora) at P_1 generally inhibit trypsin or trypsin-like serine endopeptidases such as chymotrypsin, pepsin, plasmin and kallikrein *in vitro* (e.g. bovine CTI and PTI) [31,38,77]. However, Kunitz domains with smaller, uncharged residues at P_1, such as serine, generally inhibit elastase-like proteases (eg. neutrophil elastase) [43,47,76]. In contrast, Kunitz domains with an acidic, negatively-charged P_1 residue (e.g. TKDP2) exhibit minimal antiprotease activity *in vitro* [72]. Comparison of BPTI Kunitz domains suggested that the marsupial ELP P_1 amino acids were quite rare [43,49,55]. Furthermore, the absence of purifying selection within the putative ELP/CTI trypsin interaction site and the lack of conservation of P_1 residues provides intriguing questions as to the role(s) of the marsupial ELP and eutherian CTI proteins *in vivo*.

Not all Kunitz domains act as protease inhibitors [43]. As mentioned previously, snake and spider venoms contain proteins with Kunitz domains [40]. Some domains inhibit trypsin or chymotrypsin via P_1, whilst others lack anti-protease activity but have neurotoxic effects by acting as potassium channel blockers [41]. Peigneur and colleagues [78] recently reported a sea anemone Kunitz domain protein, APEKTx1 (*Anthopleura elegantissima* potassium channel toxin 1) which had dual functions. It exhibited both trypsin-inhibitor activity and selectively blocked the Kv1.1 type of voltage-gated potassium channels. Furthermore, not all Kunitz protease inhibitors act via the P_1

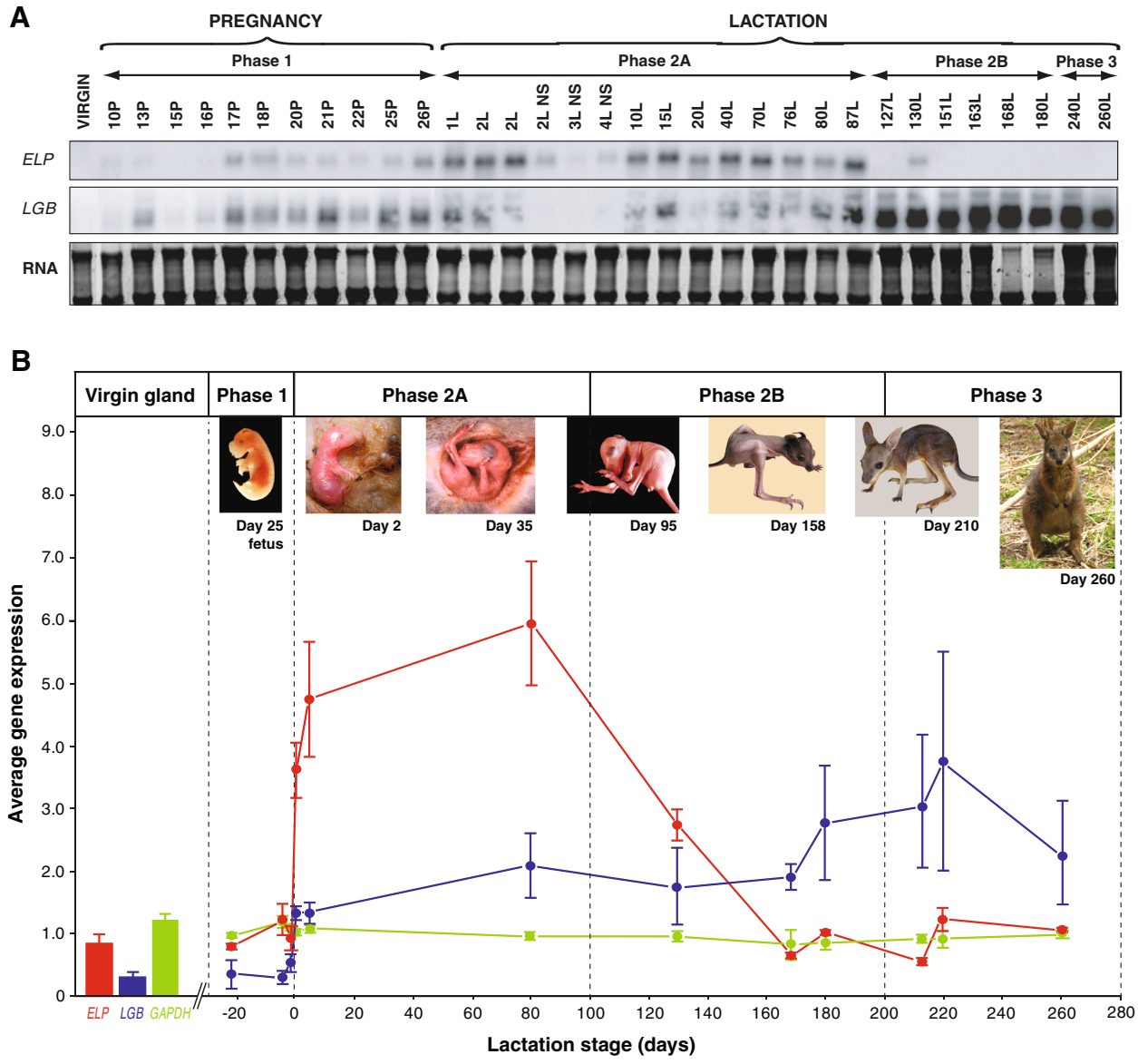


Figure 7 *ELP* expression in the tammar mammary gland throughout the reproductive cycle. A. Northern analysis of total RNA (10 μg) extracted from the mammary glands of tammar wallabies during pregnancy (P, Phase 1) and lactation (L, Phase 2A, Phase 2B and Phase 3), from non-sucked (NS) glands and from a virgin female (~220 days of age). Tammar *ELP* expression was undetected in the virgin gland, minimal during pregnancy (Phase 1) and then induced at parturition and expressed during early lactation (Phase 2A). *ELP* was down-regulated at mid-lactation (Phase 2B), consistent with [13,20,21]. *ELP* transcripts were not detected in Phase 3. *ELP* expression also declined postpartum in non-sucked glands. Tammar *LGB* was used as a positive control for lactation and exhibited a similar expression pattern to *ELP*, but with *LGB* expression increased (but not significantly so) during Phases 2B and 3, as reported previously [13,68,69]. Ribosomal RNA bands indicate RNA integrity and loading. **B.** Microarray analysis of the tammar mammary gland [ArrayExpress: E-MTAB-1057] supported the quantitative analysis of Northern blot (data not shown) and microarray data reported by [69]. Expression of the *ELP* and *LGB* milk protein genes and the housekeeping gene *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase) is depicted as average normalised raw intensity based upon the expression n = 3, 7 and 2 clones on each microarray respectively ± SEM (Additional file 12: Table S5). Whilst *ELP* (red) and *LGB* (blue) expression differed during the reproductive cycle, *GAPDH* (green) expression was constant.

residue. The tick anticoagulant peptide (TAP) inhibits Factor X, Factor Xa and thrombin but the reactive site is located towards the N-terminus of the protein, rather than at the P_1 residue of the Kunitz domain [79].

ELP/CTI – a conserved N-glycosylation site predicted within the Kunitz domain

All ELP/CTI proteins shared a putative conserved N-glyco sylation site within the Kunitz domain at asparagine-42 (as-

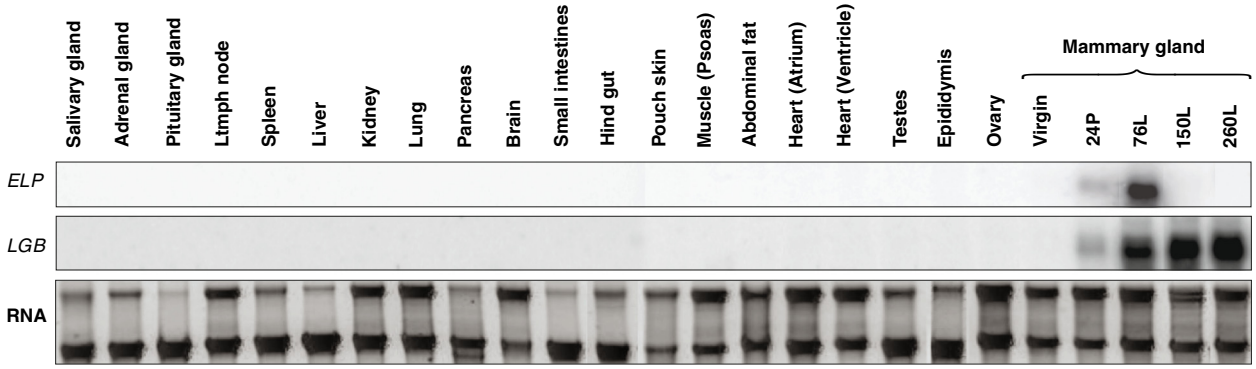


Figure 8 Tammar *ELP* expression was specific to the mammary gland. Northern analysis of total RNA (10 μg) extracted from assorted tammar tissues indicated that both *ELP* and *LGB* expression were specific to the mammary gland. Ribosomal RNA bands indicate RNA integrity and loading.

paragine-40 for koala ELP), consistent with the site identified for bovine CTI *in vitro* [58]. The proportion of sugars attached to glycosylated bovine CTI, possum ELP and tammar ELP varies, 25-40% [58,80], 60% [25] and ~47-55% [20,21,26], respectively. However, as the N-glycosylation site occurs at the base of the pear-shaped protein and at the opposite end to the P_1 site, it is unlikely to affect protease-binding activity [58]. Unlike bovine CTI, the Kunitz domains of neither bovine PTI, STI, nor for the placenta-specific TKDPs are predicted to be N-glycosylated. In fact, very few Kunitz domains are N-glycosylated, or predicted to be so [43,49,55]. The exceptions are SPINT4, SPINLW1, the first Kunitz domains of bikunin and hepatocyte growth factor activator inhibitor, the second domain of tissue factor pathway inhibitor 1, as well as selected sea anemone peptides. The precise effect of N-glycosylation is uncertain, but it may enhance protein hydrophilicity and solubility, reduce proteolysis, influence cell surface signalling and adhesion

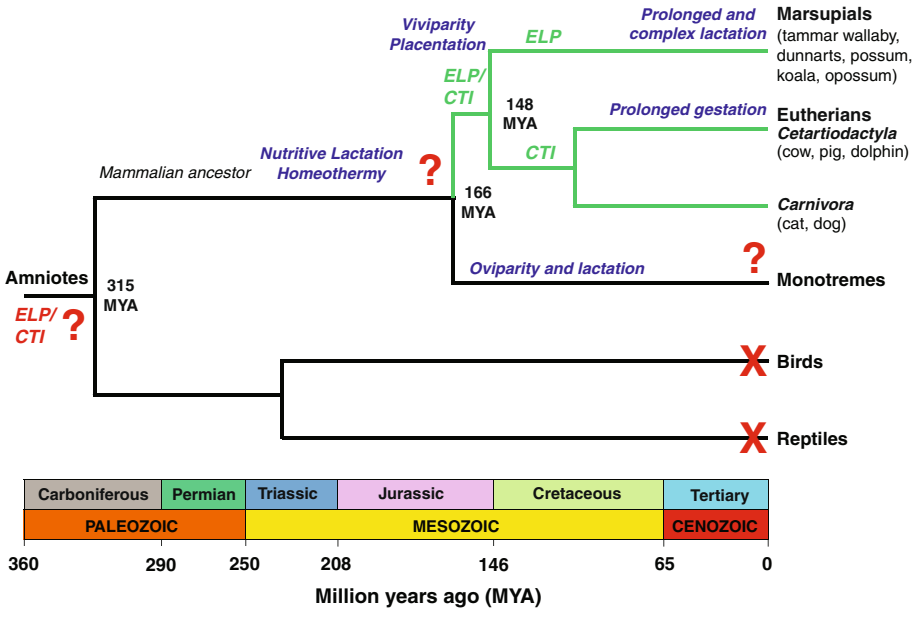


Figure 9 Evolution of the *ELP/CTI* gene in therians. Tree depicting the relationship between the amniotes: birds, reptiles, monotremes, marsupials and eutherians [1,3,70,71] and the distribution of the *ELP/CTI* gene. The divergence times used are based upon the analysis by Bininda-Emonds and colleagues [1]. Extant species which have a functional *ELP/CTI* gene are indicated by green tree branches. Extant species in which the *ELP/CTI* gene has not been detected are indicated by a red cross. Lineages on the tree for which the presence or absence of the *ELP/CTI* gene remains inconclusive are indicated by a red question mark. Based upon current analyses, the functional *ELP/CTI* gene evolved at least 130 million years ago (MYA) and has been retained by extant marsupials and the Laurasiatherian orders Cetartiodactyla and Carnivora. Whether the *ELP/CTI* gene is present in monotremes is unknown.

and affect protein folding, turnover and quality control [81-83]. Furthermore, oligosaccharides may act as soluble receptor analogues for bacterial and viral pathogens, preventing them from attaching to the wall of the intestines, thereby stopping their passage through the gastrointestinal and urinary tracts of the young [84,85].

The lack of conservation of the ELP/CTI N- and C-terminus was intriguing, particularly the positive Darwinian selection ($p < 0.05$) acting upon the coil-like marsupial ELP N-terminus. In contrast, the eutherian CTI N-terminus tended towards neutral selection. The N- and C-termini of proteins have been associated with sub-cellular targeting, protein-protein and protein-lipid interactions and macromolecular complex formation [86]. The marsupial- and eutherian-specific homology of the mature ELP/CTI N-terminus suggested these regions may have different activities. However, the lack of conservation of the ELP/CTI C-terminus suggested these areas may have species-specific effects. Interestingly, the conservation of the TGA codon used by the tammar, koala, pig, dolphin and cow for all species but the cat (CGA) suggested it was the ancestral *ELP*/*CTI* stop codon, with more recent mutations producing a shortened ELP/CTI C-terminus in some species. Furthermore, a conserved marsupial-specific region within the 3' UTR may regulate *ELP* gene transcription.

ELP/CTI is expressed and secreted in milk during the early lactation/colostrogenesis period only [this study, [20,21,25-28,31,36,37]]. Furthermore, all mammalian neonates have an innate immune system but an immature adaptive immune system and a gut which is yet to undergo maturation or 'closure' and is therefore permeable to macromolecules [16,29,87-89]. For the calf, gut maturation occurs 24–36 hr pp [16], whereas for the tammar, this process does not occur until ~200 days pp [87]. Therefore, maternal milk immunoglobulins such as IgG can be passively transferred via colostrum and Phase 2A/2B milk to the gut of the young calf and tammar, respectively, where they are absorbed by the intestines and enter the circulatory system [16,89]. Hence ELP/CTI may enhance the survival of the young by preventing the proteolytic degradation of maternal immunoglobulins [31], or by protecting the young against pathogens [25]. Although sequence comparisons predict the ELP/CTI peptides are likely to inhibit serine endopeptidases, their true function(s) will only be determined through *in vitro* and/or *in vivo* studies.

The importance of local control mechanisms in the regulation of the tammar mammary glands and *ELP* were highlighted in this study. Whilst *ELP* expression proceeds in the sucked gland, the gene is down-regulated and milk production ceases in the non-sucked glands, as for the possum [28]. However, this partitioning of mammary glands and lactation does not occur in eutherians [6]. Marsupial *ELP*/eutherian *CTI* expression was specific to the mammary gland and lactation (Figure 8), unlike the genes that most likely evolved from bovine *CTI*. PTI and STI are produced in mast cells, which have a protective role and are distributed throughout the body to tissues such as the duodenum, pancreas, lung, pituitary gland, spleen and chondrocytes [90]. In contrast, the five bovine TKDPs are differentially expressed in trophoblast cells of the ruminant placenta only during the peri-implantation period, suggesting they have an important role in the maintenance of the conceptus and pregnancy [51,63,72]. Hence, the bovine *PTI*, *STI* and *TKDP1-5* genes have undergone positive (adaptive) selection, changes in tissue-specific expression and function compared to the putative *CTI* ancestral gene, consistent with gene duplication and neofunctionalisation [91,92].

The location of the *CTI* gene in a rapidly evolving region of the eutherian chromosome [51,62] may explain the conversion of *CTI* into a putative pseudogene in Afrotheria (elephant), Xenarthra (sloth, armadillo), Euarchontoglires (humans, primates, rodents) and in selected Laurasiatherians such as the horse and flying fox.

This region included many additional genes with Kunitz and WAP 4-DSC domains [62], unlike for marsupials. It is possible that the role of CTI is fulfilled by one of these genes and hence the loss of the *CTI* gene is tolerated. Alternatively, CTI function may have become non-essential due to physiological changes in selected species. Notably, milk protein gene loss is not common amongst mammals, as genes involved in milk production are generally under negative selection [93]. However, the conservation of the *ELP*/*CTI* gene in marsupials and Laurasiatherian orders Carnivora (dog, cat, dolphin, panda) and Cetartiodactyla (cow, pig) suggests ELP/CTI has an important role in these species.

Conclusions

Marsupial *ELP* and eutherian *CTI* evolved from a common ancestral gene and encode a milk protein with a single BPTI-Kunitz serine protease inhibitor domain. Although *CTI* was identified as the putative ancestral gene of *PTI*, *STI* and the placenta-specific trophoblast *TKDP1-5* gene family, the origin of the *ELP*/*CTI* gene is inconclusive. *ELP*/*CTI* expression in the postpartum mammary gland is brief (~24-48 hrs) in eutherians but prolonged in the tammar and other marsupials (up to 100 days). However, this period correlates with the provision of milk to an immuno-incompetent young, suggesting ELP/CTI may play a vital role in immune protection of the young at this time.

Methods

Animals

Tammar wallabies (*Macropus eugenii*) were provided from two different marsupial colonies: VIAS (Victorian Institute of Animal Science), DPI (Department of Primary Industries), Attwood, Victoria and The University of

Melbourne, Victoria. Animals were kept in open grassy yards with *ad libitum* access to food, water and shelter, using standard animal husbandry conditions in accordance with the National Health and Medical Research Council guidelines [94]. All experiments were approved by the Animal Experimentation Ethics Committees of the Department of Primary Industries and The University of Melbourne.

Tissues

Tissues (salivary gland, adrenal gland, pituitary gland, lymph node, spleen, liver, kidney, lung, pancreas, brain, small intestines, hind gut, muscle, heart, ovaries) were collected from adult female tammars (n = 2). Mammary glands were also collected from adult females at different stages of pregnancy and lactation (n = 60). Mammary glands from virgin females were collected from tammar pouch young (~220 days of age, n = 3). Testes and epididymides were collected from adult tammar males (n = 2). Tissue samples derived from ear-tagging of a population of koalas (*Phascolarctos cinereus*) located on French Island, Victoria, were donated by Dr. Kath Handasyde and Dr. Emily Hynes from the Department of Zoology, The University of Melbourne. Total RNA extracted from a grey short-tailed opossum (*Monodelphis domestica*) mammary gland from day 15 of lactation (early-lactation) was provided by Dr Denijal Topcic (The University of Melbourne) from animals provided by Professor Norman Saunders (The University of Melbourne). Dr Peter Frappell (Latrobe University) provided fat-tailed dunnart mammary gland tissue from day 37 of lactation (Phase 2) and liver tissue. Dr Amelia Brennan (The University of Melbourne) provided total RNA isolated from the mammary gland of a late-pregnant (~8 months) Holstein-Friesian cow. A small quantity of dog colostrum (~20 μL) from a late-pregnant (~2 weeks prepartum) Labrador in its first pregnancy was also kindly donated by Cate Pooley (The University of Melbourne). All samples were snap frozen in liquid nitrogen and stored at −80°C until use, with the exception of the koala ear punches, which were stored at 4°C.

RNA extraction and northern analysis

Total RNA was extracted from tissues using the Qiagen RNeasy Midi Kit (Qiagen) and from cells isolated from colostrum using RNAWIZ (Ambion). RNA extracted from cells shed into milk during the lactation process provides a good representation of gene expression in the mammary gland [95] and therefore eliminates the need for destructive tissue sampling. RNA was electrophoresed through a 1% agarose, low-formaldehyde (1.1%) gel with 1X MOPS [3(N-Morpholino) Propane Sulfonic Acid] buffer at 4°C and then transferred to Zeta-Probe GT Blotting Membrane (BioRad) in 20X SSC (3.0 M sodium chloride, 0.3 M trisodium citrate, pH 7.0) overnight.

Membranes were rinsed in 2X SSC, UV crosslinked at 1200 J (Stratagene UV Stratalinker1800) and hybridized in 25 mL [30% deionised formamide, 5 X SSC, 50 mM sodium acetate, herring sperm DNA (100 μg/μL), 5 mL Denhart's 50X stock solution, 0.1% SDS] with an [α-^{32}P] dCTP-labelled probe [DECAprime II Random Priming DNA Labelling Kit (Ambion)] and incubated for ~16 hr at 42°C. The tammar *ELP*, *Rsa*I digested *LGB* (to detect both *LGB* transcripts [96]) and *CST3* probes were either amplified by RT-RCR from tammar mammary gland total RNA or sourced from clones in a tammar mammary gland EST library held by the Cooperative Research Centre for Innovative Dairy Products [19], with plasmid DNA isolated and the cDNA insert amplified by PCR. Membranes were washed (0.1X SSC, 0.1% SDS) twice for 15 min at 60°C, wrapped in cling film, sealed into plastic pockets and exposed to a General Purpose Storage Phosphor screen and scanned on a Typhoon 8600 Scanner (Molecular Dynamics/GE Healthcare). Membranes were stripped of probes by incubation with boiling (100°C) 1X SSC, 0.1% SDS on a shaking platform for two 15 min periods, then rinsed with RT 1X SSC, 0.1% SDS.

RT-PCR and cloning of *ELP/CTI*

cDNA was generated using Superscript III Reverse Transcriptase (Invitrogen), oligo(dT)20 primer (50 μM; Sigma-Proligo) and 5 μg of total RNA isolated from mammary tissue or cells separated from milk. PCR was performed using 2 μL (10%) of the first strand reaction, the proof-reading Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen), plus the appropriate forward and reverse primers and conditions to amplify *ELP*/*CTI* transcripts (Table 3). PCR products were cloned into the pGEM-T Easy Vector System I (Promega) and sequenced. Full protein-coding *ELP*/*CTI* transcripts were cloned from total RNA extracted from the fat-tailed dunnart, cow and opossum mammary gland tissues and from cells in canine colostrum.

Genomic DNA isolation and cloning

Genomic DNA was isolated from koala and fat-tailed dunnart tissues as described [97]. The *ELP*/*CTI* genes were amplified by PCR (Table 3) using Platinum *Taq* DNA Polymerase and ~200 ng of genomic DNA template, cloned into pGEM-T Easy and sequenced.

Isolation of the tammar *ELP* gene from a genomic library

A tammar genomic library (liver) in the *E. coli* phage vector lambda EMBL3 T7/SP6 was screened with tammar *ELP* cDNA and a positive clone isolated. The clone was *Sal*I digested and the ~14.7 kb genomic

Table 3 Primer sequences and conditions used to amplify *ELP/CTI* genes and transcripts

ELP/CTI gene/ transcript	Name	Primer Sequence[1] 5' 3'	PCR Product Size (bp)	Primer Conditions
FT dunnart transcript	FT_ELP_F	GTCAAGTGTTATCTACTGGCAGCACCATG	488	94°C for 2 min; 35 cycles of 94°C for 30 sec; 59°C for 30 sec; 68°C for 1 min; 68°C for 10 min
	FT_ELP_R	CCCAAAGTGCTGTTAATGCTTTATTGTAGC		
Opossum transcript	mELP_NheI_F	***GCTAGC***AAGGTTTTCTCTCAGTGCCATC	488	94°C for 2 min; 35 cycles of 94°C for 30 sec; 60°C for 30 sec; 68°C for 30 sec; 68°C for 10 min
	mELP_BamHI_R	***GGATCC***TGTTAATGCTTTATTGTACCAG		
Tammar transcript	tELP_NheI_F	***GCTAGC***AAGTGTAGTCTACCAGTGGCACC	479	94°C for 2 min; 35 cycles of 94°C for 30 sec; 58°C for 30 sec; 68°C for 30 sec; 68°C for 10 min
	tELP_BamHI_R	***GGATCC***TGTTAATGCTTTATTGTACCAG		
Dog transcript	Dog_ELP_Ex1_F	GCCTAGAACATTCAGCTATTGGCACC	449	94°C for 2 min; 35 cycles of 94°C for 30 sec; 55°C for 30 sec; 68°C for 1 min; 68°C for 10 min
	Dog_ELP_Ex3_R	TGAATGTTTTATTGACCTAGACCTGGAGG		
Cow transcript	bELP_NheI_F	***GCTAGC***AACTCACAGCTCCTCACACCATG	463	94°C for 2 min; 35 cycles of 94°C for 30 sec; 58°C for 30 sec; 68°C for 30 sec; 68°C for 10 min
	bELP_BamHI_R	***GGATCC***GAACACTTTATTGACCCAGTCCTG		
FT dunnart gene	FT_ELP_F	GTCAAGTGTTATCTACTGGCAGCACCATG	4771	94°C for 2 min; 35 cycles of 94°C for 30 sec; 55°C for 30 sec; 68°C for 6 min; 68°C for 10 min
	FT_ELP_R	CCCAAAGTGCTGTTAATGCTTTATTGTAGC		
Koala gene	tELP_Ex1_F	GGTAGCAAGTGTAGTCTACCAGTGGCACC	1428	94°C for 2 min; 35 cycles of 94°C for 30 sec; 52°C for 30 sec; 68°C for 4 min; 68°C for 10 min
	tELP_BamHI_R	***GGATCC***TGTTAATGCTTTATTGTACCAG		
Tammar gene (6.2 kbpromoter)	T7	TAATACGACTCACTATAGGG	6326	94°C for 2 min; 35 cycles of 94°C for 30 sec; 57°C for 30 sec; 68°C for 8 min; 68°C for 10 min
	tELP_Prom_R	GACTGATCAGACCAATATAAGCTT		
Tammar gene (7.9 kbpromoter)	T7	TAATACGACTCACTATAGGG	8044	94°C for 2 min; 35 cycles of 94°C for 30 sec; 57°C for 30 sec; 68°C for 8 min; 68°C for 10 min
	tELP_Ex1_R	GAGGGCCAACGATGGTAAATTTCAT		

[1]Restriction enzyme sites are indicated in bold, italicised text.

DNA fragment cloned into a modified pBeloBACII plasmid vector. Digestion of pBeloBACII-14.7kbtELP with *Sal*I and *Hind*III yielded three fragments, 6.2 kb *Sal*I/*Hind*III, 5.2 kb *Hind*III/*Hind*III and 3.3 kb *Sal*I/*Hind*III. These fragments were sub-cloned into pBluescript SK and the latter two clones sequenced by the Australian Research Genome Facility (Australia). The remaining 6.2 kb was sequenced (Department of Pathology, The University of Melbourne), providing the full sequence of the genomic clone (14.704 kb). BLAST [98] searches of the NCBI *Macropus eugenii* WGS (Whole Genome Shotgun) trace archives and assembly of hits with CAP3 [99,100] produced a contig of 54,363 bp which included *ELP* and the first 2 exons of *WFDC2*.

Fluorescence *in situ* hybridisation (FISH)

Metaphase spreads were prepared from the tammar and FISH performed as described [101]. The 14.7 kb tammar *ELP* genomic clone was used as a probe. Slides were examined using a Zeiss Axioplan microscope and images captured using the Spot Advance software package. Pictures were processed with Confocal Assistant, Image J, Adobe Illustrator and Adobe Photoshop. Chromosomal location of *ELP* was verified by at least ten metaphase spreads that had at least three or four signals out of a maximum of four.

cDNA microarray analysis of tammar *ELP* gene expression

ELP gene expression in the tammar mammary gland was investigated by analysing a microarray database [69,102-104] produced from custom-made cDNA microarray slides and total RNA collected from glands at each phase of the lactation cycle [69,102-104]. Glass microarray slides were printed by the Peter MacCallum Cancer Centre Microarray Core Facility, Melbourne, Australia and contained 10,368 tammar cDNA spots which were derived from a commercially prepared (Life Technologies, Rockville, MD, USA), normalised 15,001 tammar mammary gland EST (expressed sequence tag) library. The library was prepared using tammar mammary gland total RNA pooled from various time points in pregnancy (P), lactation (L) and involution (I). These included: day 26P, d55L, d87L, d130L, d180L, d220L, d260L and d5I (tissue from a d45L female 5 days after removal of the pouch young (RPY)) [19]. Gene expression changes in the tammar mammary gland during the reproductive cycle were investigated by a large-scale microarray experiment involving 36 comparisons (72 slides including dye swaps, 144 channels in total) [69,102-104].

Sixteen different time points were used in the experiment: virgin female ~ 300 days old (n = 3), pregnancy (Phase 1: d5P, d25P, d26P; n = 1 per time point), lactation (Phase 2A: d1L, d5L, d80L; Phase 2B: d130L, d168L, d180L; Phase 3: d213L, d220L, d260L; n = 1 per time

point) and involution (pouch young were removed at d264L and mammary tissue sampled 1, 5 and 10 days after RPY; n = 1 per time point). Microarray probes were prepared from total RNA (50 μg per sample) using a two-step procedure which involved incorporation of aminoallyl-modified dUTP and then coupling with either Cy3 or Cy5 fluorescent dye [102,104]. Slides were hybridised overnight (14–16 hr) in a humidified chamber [102,104], scanned (Agilent scanner) and the images analysed with Versarray software (Bio-Rad).

Quantile-quantile normalisation within and between microarray slides was implemented using the Limma Package of Bioconductor [105]. The complete data set was analysed simultaneously using a large-scale, linear mixed-model, which included random effects to account for the microarray experiment design, plus gene effects and gene-contrast effects [102,106]. For each time point during pregnancy and lactation, there were a total of 4 different microarray comparisons made; 8 including the Cy3/Cy5 dye swap experiments. For the virgin tissues, there were a total of 12 comparisons, with these values combined for each gene and the average determined. The relative gene expression levels were determined by exponentiation of the gene effects values. The expression levels of the *ELP* and *LGB* milk protein genes and the housekeeping gene *glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*) were based upon the average expression of n = 3, 7 and 2 non-identical clones on each microarray respectively ± SEM. Microarray experiment data (E-MTAB-1057) was submitted to the EBI Array Express Archive [107].

Sequence analysis

ELP/CTI genes and pseudogenes were identified by BLAST searches of the NCBI GenBank nr and WGS trace archives and BLAST searches of the Ensembl Release 62, April 2011 [49] and UCSC [55] genome databases. We used an Expect-value ≤ 1e-8 as a cut-off for orthologue identification for nucleotide comparisons and gene structure comparison and an E-value ≤ 1e-17 for protein comparisons. Contigs were assembled with CAP3. The following *ELP/CTI* genes and transcripts were submitted to GenBank: the *ELP* gene of the tammar (14.704 kb) [GenBank: JN191335], Southern koala [GenBank: JN191337] and fat-tailed dunnart [GenBank: JN191336], the *ELP* transcripts of the tammar [GenBank: JN191338], fat-tailed dunnart [GenBank: JN191339] and South American opossum [GenBank: JN191340] and *CTI* transcripts of the cow (Holstein-Friesian breed) [GenBank: JN191341] and dog (Labrador breed) [GenBank: JN191342]. Third party annotations of the *ELP*/*CTI* gene were also submitted to GenBank for the cat: [GenBank: BK008083], dog: [GenBank: BK008082], dolphin [GenBank: BK008086], opossum [GenBank: BK008085] and panda [GenBank: BK008084].

The genomic regions encompassing the *PIGT*, *ELP/CTI* and *WFDC2* genes in different species were sourced from either the Ensembl or UCSC genome databases for sequence comparisons using mVISTA [64]. These included: dog build *CanFam2* chr24: 35680293–35758485, elephant build *loxAfr3*:

SuperContig_scaffold_19:44809970–44903157, horse build *EquCab2* chr22: 34,465,586-34568786, human build *hg19* chr20: 436717–510935, mouse build *mm9/NCBI37* chr2: 164320020–164401749, opossum build *MonDom5* chr1: 501309327–501453154 and cow build *Btau_4.0* chr13: 74506302–74550554 (included the *PIGT* and *CTI* genes) and 75064658–75139756 (included the *WFDC2* gene). The tammar genome sequences used for comparisons included the incomplete *PIGT* gene in tammar build *Meug_1.0* GeneScaffold_3597: 2268–20682, and a 54,363 bp contig which included tammar *ELP* and the first 2 exons of *WFDC2*. The contig was compiled by BLAST searches of the NCBI *Macropus eugenii* WGS trace archives with the tammar *ELP* gene and assembly with CAP3. The following bovine chromosome 13 genes were also extracted for comparisons: *CTI* (74530701–74533686), *PTI* (75011365–75016221), *STI* (75065067–75069211), *TKDP1* (74843274–74860062), *TKDP2* (74913592–74923363), *TKDP3* (74567402–74577188), *TKDP4* (74874966–74883256), and *TKDP5* (74976879–74983345). The web-based CENSOR tool [108] was used to mask sequences and identify transposable elements by comparison to the Repbase database of repeat elements [66]. Putative exons, transcripts and proteins within genomic sequences were predicted using GENSCAN [109]. However, the third exon of *ELP*/*CTI* was incorrectly predicted by GENSCAN and was therefore determined by manual comparison to known *ELP*/*CTI* splice sites. Splice site location was confirmed by comparison of transcripts and putative proteins. Masked sequences were analysed with mVISTA [64]. Specifications used for each analysis are described in the relevant figure legends.

The ELP/CTI, PTI, STI, SPINT4 (bovine SPINT3 has not been detected) and TKDP family of proteins were subjected to a Prosite database scan [110] to identify putative conserved motifs and post-translational modifications. Putative leader sequences (indicative of secreted proteins) and N-glycosylation sites based upon the NX (S/T) motif were predicted by SignalP 3.0 and NetNGlyc 1.0 Server, respectively, using the Center for Biological Sequence analysis Prediction Servers [56]. Sequences were aligned with CLUSTALW2 [111] and homology within *ELP*/*CTI* transcripts and proteins assessed with MatGAT (Matrix Global Alignment Tool) 2.01 software

[112]. MatGAT produces pairwise alignments only and determines homology between each sequence pair based upon the BLOSUM50, BLOSUM62 (used for this study) or PAM250 matrix.

dN/dS analysis

Selection pressures acting upon different regions of the marsupial ELP and eutherian CTI precursor proteins were determined by dN/dS analysis with MEGA5 software [60]. The protein-coding regions of the marsupial and eutherian transcripts were analysed separately. For each region, the average transition/transversion ratio was calculated using the Maximum Composite Likelihood estimate of the pattern of nucleotide substitution based upon the Tamura-Nei model [113] and then used in the subsequent dN/dS analysis. All codon positions were used, but positions within the alignment containing gaps were eliminated from the analysis. In pairwise comparisons, dN (number of non-synonymous changes per non-synonymous site) and dS (number of synonymous changes per synonymous site) were estimated using the Nei-Gojobori method [114] with modified Jukes-Cantor correction [115] and their variances determined by boostrapping (1000 replications). Codon-based Z-tests for positive ($dN > dS$), purifying ($dN < dS$) and neutral ($dN = dS$) selection were carried out using the Modified Nei-Gojobori method with Jukes-Cantor correction in MEGA5.

Phylogenetic analysis

The phylogenetic relationship between the protein-coding regions of the marsupial *ELP*, eutherian *CTI*, bovine *TKDP1-5*, *PTI* and *STI* transcripts was investigated using PHYLIP software version 3.69 [116]. Bovine secretory leukocyte protease inhibitor (*SLPI*, GenBank: NM_001098865) was used as an outgroup for the analysis.

Transcripts were aligned with MUSCLE [117] and then 100 bootstrapped alignments generated with SEQBOOT (PHYLIP). The phylogenetic relationship between the sequences was determined using different methods including the character-based maximum likelihood and maximum parsimony methods, as well as distance-based methods. Maximum likelihood trees were generated with DNAMLK which uses a molecular clock assumption. A transition/transversion ratio of 1.34 and a coefficient of variation for the rate of substitution among sites of 0.848 (based upon a gamma distribution with a shape of 1.39) were also specified for the analysis. These values were derived from a Maximum Likelihood test of best fit for 24 different nucleotide substitution models with MEGA5. A Hidden Markov Model using 5 categories, global rearrangements and a randomized input order jumbled once were also used for the DNAMLK analysis. A consensus tree was generated with CONSENSE specifying *SLPI* as an outgroup root, redrawn with RETREE and plotted with DRAWGRAM. Bootstrapped trees were also generated without the molecular clock assumption (DNAML) and using maximum parsimony (DNAPARS). Distance-based analysis on bootstrapped alignments was carried out with DNADIST using the Kimura [118] model of nucleotide substitution. The values used for transition/transversion ratio and gamma distribution were the same as for the maximum likelihood analysis. Trees were generated with the FITCH joining method [119] using global rearrangements, a randomized input order jumbled 10 times and *SLPI* as an outgroup root. The bovine *CTI*, *TKDP1-5*, *PTI*, *STI*, *SPINT4* and *SLPI* protein-coding transcripts were also analysed with PHYLIP as described above. However, a transition/transversion ratio of 1.39 and a coefficient of variation for the rate of substitution among sites of 0.913 were used.

Additional files

Additional file 1: Table S1. Characterisation of the putative functional *ELP/CTI* gene, transcript and protein.

Additional file 2: Figure S1. Alignment of the marsupial *ELP* and eutherian *CTI* transcripts. Nucleotide sequences of the tammar [GenBank: JN191338], fat-tailed dunnart (FT) [GenBank: JN191339], opossum [GenBank: JN191340], cow (Holstein-Friesian breed) [GenBank: JN191341], dog (Labrador breed) [GenBank: JN191342] and brushtail possum, plus the transcripts predicted from the *ELP* genes of the stripe-face (SF) dunnart [GenBank: AC186006], koala [GenBank: JN191337], cat [GenBank: BK008083], pig [Ensembl: F1SD34_PIG (ENSSSCT00000008098)], Giant panda [GenBank: BK008084], and Common bottlenose dolphin [GenBank: BK008086] were aligned with ClustalW2. Black shading indicate nucleotide residues common to at least 10 of the species and grey, those that differ. Teal shading indicates nucleotides common to marsupials only, whilst those common to eutherians are shaded blue. The putative translation start site (ATG) is shaded green, the predicted stop codons red, and the polyadenylation signal (AATAAA) is indicated by red text. Nucleotides which encode the signal peptide are indicated by a blue arrow and the Kunitz domain motifs (BPTI KUNITZ 2, Prosite: PS50279 and BPTI KUNITZ 1, Prosite: PS00280) are indicated by red and green lines, respectively. The codons which encode the 6 cysteine residues that form the 3 disulphide bonds of the 51 amino acid Kunitz domain are boxed red. The putative P_1-P_1' reactive site residues are shaded yellow and purple respectively. Black arrows indicate the location of *ELP* exons and gaps within the alignment are indicated (–).

Additional file 3: Table S2. Percentage similarity between and within the marsupial *ELP* and eutherian *CTI* transcripts. Pairwise similarities were determined using MatGAT2.01 software [112] based upon alignment of sequence pairs using the BLOSUM62 matrix. **A.** *ELP/CTI* transcripts (translation start, ATG, to the polyadenylation signal, AATAAA inclusive), **B.** *ELP/CTI* transcripts (translation start, ATG, to the stop codon inclusive), **C.** Marsupial *ELP* 3'-UTR (untranslated region).

Additional file 4: Table S3. Percentage similarity between and within the marsupial ELP and eutherian CTI peptides. Pairwise similarities were determined using MatGAT2.01 software [112]. **A.** ELP/CTI signal peptide, **B.** ELP/CTI mature peptide, **C.** ELP/CTI N-terminus, **D.** ELP/CTI Kunitz domain motif 2 (51 amino acids), **E.** ELP/CTI Kunitz domain motif 1 (19 amino acids) and **F.** ELP/CTI C-terminus.

Additional file 5: Figure S2. Exon 1 and 2 mutations within selected putative eutherian *CTI* pseudogenes. **A (i).** ClustalW2 alignment of *CTI* exon 1 of the sloth, elephant, human and horse compared to dog exon 1 revealed different putative mutations and deletions. For sloth and horse *CTI* there was a point mutation within the putative translation start site (nt 1–3, methionine codon, ATG). However, human *CTI* exon 1 was disrupted by the deletion of 2 nucleotides (nt 26–27)

which would produce a frame-shift (A (ii)). The predicted GT splice site (nt 77–78, orange box) was also disrupted for both the elephant and horse *CTI* sequences. Interestingly, the mutation in the elephant GT splice site would produce a putative protein-coding open reading frame of (279 bp). If this region was transcribed and translated, a precursor protein of 92 amino acids would be secreted. Furthermore, SignalP analysis suggested a mature secreted protein of 70 residues would be produced (data not shown). Nucleotides common to at least four species are boxed black and the remainder, grey. **A (ii).** ClustalW2 alignment of the translated exon 1 region of the functional canine CTI protein revealed mutations in the methionine codon (translation start site) for horse and sloth CTI. In addition, the predicted deletion of 2 nucleotides in human *CTI* would produce a frame-shift. This is shown by the +1 and +2 reading frames of human CTI (2 yellow boxes). Amino acid residues identical to those of the dog are shaded black, with similar amino acid types in grey. **B (i).** ClustalW2 alignment of the functional canine CTI exon 2 with those of the horse, mouse and rat revealed multiple deletions within rodent *CTI*, but a single nucleotide deletion (nt 186) in equine *CTI*. Putative splice sites are indicated (orange boxes). Mutations were also present in the rodent intron 1 AG splice site (nt 20–21), whilst the GT splice site was intact (nt 243–244). The location of the BPTI KUNITZ 1 and 2 motifs within canine *CTI* exon 2 are indicated by green and reds bars respectively. Nucleotides that encode the cysteine residues of the Kunitz domain are shaded red. Mutations in the cysteine encoding nucleotides of C1 were detected for the rat and mouse and also in C2 of murine *CTI*. Nucleotides common to at least four species are shaded black and the fifth grey. **B (ii).** ClustalW2 alignment of the protein encoded by the functional dog *CTI* exon 2 with the putative horse, mouse, and rat proteins revealed multiple mutations and frame-shifts. Equine *CTI* was disrupted by a frame-shift, as shown by alignment of the +2 and +3 reading frames (2 yellow boxes). In contrast, rat CTI has been disrupted by multiple deletions, as evident from the comparison of the +1, +2 and +3 reading frames. This was also true for murine CTI (+1 and +2 reading frames only shown). Gaps have been added to the dog CTI sequence to assist with the alignment (.). Amino acid residues identical to those of the dog are shaded black, with similar residue types shaded grey.

Additional file 6: Figure S3. Transposable elements and simple repeats located within the *PIGT* and *ELP/CTI* genes and flanking regions. Conserved transposable elements in the region containing the *PIGT* and *ELP/CTI* genes of the opossum, tammar, dog, horse, human, elephant and cow were identified using CENSOR [66,108]. The horizontal axis indicates the relative sizes of the regions compared. Green and red arrows indicate the *PIGT* and *ELP/CTI* genes respectively, whilst red arrows with diagonal white stripes indicate the putative horse, human and elephant *CTI* pseudogenes. Exons are indicated by red rectangles. There was a gap in the tammar genome assembly between *PIGT* and *ELP* and the last exon of *PIGT* was missing (red dashed rectangle). Coloured rectangles indicate the different retroelement classes: Transposable elements: DNA transposon (maroon), LTR (long terminal repeat) retrotransposons (brown), Endogenous retrovirus (orange), Non-LTR retrotransposons (blue), interspersed repeat (black) and simple repeat (green). White space indicates the absence of retroelements. Solid lines indicate elements conserved between adjacent species as depicted. Dashed lines indicate elements not present in the adjacent species, but that are preserved in others. Conserved elements are shown in coloured text and those that differ are indicated by black text. Selected retroelements are identified.

Additional file 7: Figure S4. Alignment of the bovine CTI, PTI, STI, TKDP1-5 and SPINT4 precursor proteins. ClustalW2 alignment of the bovine CTI [GenBank: JN191341], PTI [GenBank: P00974], STI [GenBank: NP_991355], TKDP1 [GenBank: NP_991345], TKDP2 [GenBank: AF241777], TKDP3 [GenBank: DAA23071], TKDP4 [GenBank: AAF61250], TKDP5 [GenBank: XP_614808] and SPINT4 [GenBank: XP_614808] precursor proteins. Amino acid residues are numbered based upon the translation start of the precursor proteins and indicated on the right hand side of the alignment. The signal peptides were predicted by SignalP and boxed (blue). The region encoded by the Kunitz domain exon is also boxed (red). The six conserved cysteine residues (C1-C6, C2-C4 and C3-C5), which form the three disulphide bonds that produce a globular protein are shaded red. Notably, C2 and C4 are absent from the TKDP3 and TKDP4 proteins [63]. The BPTI KUNITZ 1 and 2 motifs are indicated (green and red bars respectively) and the putative trypsin interaction (TI) site from the KU motif (NCBI cd00109) is shown by orange triangles. The putative P_1 reactive site is indicated. Bold, italicised asparagine (*N*) residues indicate predicted sites of post-translational N-glycosylation. Only CTI and SPINT4 were predicted to be N-glycosylated within the Kunitz domain. Amino acid residues that overlap splice sites are shown in red text. Conservation between groups of amino acids with strongly similar properties, i.e., scoring > 0.5 in the Gonnet PAM 250 matrix is indicated (:). Conservation between groups of amino acids with weakly similar properties (scoring < 0.5 in the Gonnet PAM 250 matrix) is also noted (.). Gaps within the alignment are indicated (–).

Additional file 8: Figure S5. Relationship between bovine *CTI*, *PTI*, *STI*, *TKDP1-5* and *SPINT4*. The evolutionary history of the protein-coding regions of the bovine *CTI*, *PTI*, *STI*, *SPINT4* and *TKDP1*-5 transcripts was determined by maximum likelihood analysis based upon a molecular clock assumption using PHYLIP. Bovine *SLPI* was used as an outgroup (data not shown). Numbers at branch points indicate confidence levels as determined by bootstrap values (100 replicates). Transcripts were aligned with MUSCLE and bootstrapped values generated with SEQBOOT. Trees were generated with DNAMLK using a transition/transversion ratio of 1.39, a coefficient of variation for the rate of substitution among sites of 0.913, 5 Hidden Markov Model categories, global rearrangements and a randomised input order jumbled once. The protein-coding regions of the following bovine transcripts were used in the analysis: *CTI* [GenBank: JN191341], *PTI* [GenBank: NM_001001554], *STI* [GenBank: NM_205786], *TKDP1* [GenBank: NM_205776], *TKDP2* [GenBank: NM_001012683], *TKDP3* [GenBank: XM_584746], *TKDP4* [GenBank: NM_205775], *TKDP5* [GenBank: XM_614808], *SPINT4* [Ensembl: ENSBTAT00000039210] and *SLPI* [GenBank: NM_001098865].

Additional file 9: Figure S6. Genomic arrangement and mVISTA plot of pairwise alignments for the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes. **A.** Arrangement and orientation of the bovine chromosome 13 *CTI*, *PTI*, *STI*, *TKDP1 TKDP2*, *TKDP3*, *TKDP4* and *TKDP5* genes. **B. (i-viii)** Homology between the *CTI PTI*, *STI* and *TKDP1-5* genes as determined by mVISTA pairwise sequence alignment. Grey horizontal arrows indicate genes, coding exons are indicated by blue boxes and UTRs of the gene as light green rectangles. The right axis indicates the percentage identity for each pairwise comparison within a 100 bp window, ranging from 10% to 100%. Regions sharing greater than 25% identity are shaded and the black horizontal line indicates 70% identity. The horizontal axis indicates the size of the reference sequence used for each comparison: **(i)** Bovine *CTI*, **(ii)** *PTI*, **(iii)** *STI*, **(iv)** *TKDP1*, **(v)** *TKDP2*, **(vi)** *TKDP3*, **(vii)** *TKDP4*, and **(viii)** *TKDP5*. The *CTI* Kunitz domain was most similar to that of *PTI*, *STI*, and *TKDP3*, whilst *PTI* and *STI* homology was greatest within the *TKDP* gene family.

Additional file 10: Figure S7. Transposable elements located within the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes. Conserved transposable elements within the *CTI*, *PTI*, *STI*, *TKDP1*, *TKDP2*, *TKDP3*, *TKDP4* and *TKDP5* genes (translation start to polyadenylation site, inclusive) were identified using CENSOR [66,108]. The *TKDP* N-domain-encoding exons located between exon 1 (signal- and pro-peptide) and the Kunitz domain-encoding exon (light blue rectangle) were most likely to have arisen due to the "exonisation" of an intron [50]. Their evolutionary history including phylogenetic and dN/dS analysis is discussed in detail [50,51,63]. The second exon of *TKDP5* (A) which is adjacent to a MER21 element may be the ancestral exon of the unique 3-exon N-domains. The ancestral *TKDP5* gene was probably then duplicated to produce either the ancestral *TKDP4*, *TKDP3* or *TKDP2* gene. Within this copied gene, the retroelement (and the exonised intron) was most likely duplicated a further two times (B) and (C), producing a tripartite N-domain of 3 exons: C, B and A (yellow bar 1). This gene subsequently underwent 3 rounds of duplication, resulting in four genes with one N-domain, i.e. three exons which encode the N-domain. Three of these genes, *TKDP4*, *TKDP3* and *TKDP2*, retained the original N-domain. However, for the fourth (*TKDP1*) the tripartite N-domain was replicated twice (yellow bar 2 and yellow bar 3). The horizontal axis indicates the relative sizes of the regions compared, with all genes transcribed from left to right. Exons are indicated by red rectangles, with the exception of the Kunitz domain-encoding exon which is shown as a blue rectangle. Coloured rectangles indicate the different retroelement classes: Transposable elements: DNA

transposon (maroon), LTR (long terminal repeat) retrotransposons (brown), Endogenous retrovirus (orange), Non-LTR retrotransposons (blue), interspersed repeat (black) and simple repeat (green). White space indicates the absence of transposable elements. Coloured lines link elements conserved between genes. Selected retroelements are identified, with the conserved fragment size and orientation shown (d = direct, c = complementary). Conserved elements are indicated by coloured text and those that differ, by black text. Arrows indicate the relative orientation of each gene on bovine chromosome 13.

Additional file 11: Table S4. Location, identity and orientation of transposable elements within the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes. CENSOR [66,108] output tables showing the predicted identity, location and orientation of retroelement fragments within the bovine *CTI*, *PTI*, *STI*, *TKDP1*, *TKDP2*, *TKDP3*, *TKDP4* and *TKDP5* genes.

Additional file 12: Table S5. Tammar wallaby mammary gland cDNA microarray data presented in Figure 7. Global normalised Cy3/Cy5 gene expression data for (i) *ELP*, (ii) *LGB* and (iii) *GAPDH* - throughout the lactation cycle derived from custom-made tammar mammary gland cDNA microarrays. Stages of the reproductive cycle investigated included the virgin female mammary gland, Phase 1 (pregnancy), Phase 2A (early lactation), Phase 2B (mid-lactation) and Phase 3 (late lactation). The EBI ArrayExpress Accesion number (E-MTAB-1057) and GenBank Accession numbers for each clone (microarray spot) are provided, plus the average expression of *ELP*, *CTI* and *LGB* and the associated standard deviation and standard errors.

Abbreviations

4-DSC: Four disulphide core; aa: Amino acid; AMBP: α1-microglobulin/bikunin precursor; bp: Base pairs; Da: Daltons; EST: Expressed sequence tag; LTR: Long terminal repeat; MER: Medium Reiterated frequency repeat; MYA: Million years ago; nt: Nucleotide; pp: Postpartum; PY: Pouch young; RPY: Removal of pouch young; SPINLW1: Serine peptidase inhibitor-like with Kunitz and WAP domains 1; SPINT: Serine protease inhibitor Kunitz type; TFPI: Tissue factor pathway inhibitor; WAP: Whey acidic protein; WFDC: Wap four disulphide core; WFIKKN: WAP, follistatin/kazal, immunoglobulin, Kunitz and netrin domain containing protein; WGS: Whole genome shotgun.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Mammary glands were dissected by KRN and EAP from animals provided by MBR and the Department of Primary Industries, Attwood, Victoria. Experiments were conducted and analysed by EAP, with the exception of the fluorescence *in situ* hybridization of the tammar *ELP* gene, which was performed by AAD, the statistical analysis of microarray experiments which was performed by CML and PCT and the microarray experiments. All authors read, edited and approved the final manuscript.

Acknowledgements

We thank Dr Kaylene Simpson, Michael Wilson, Jenni Carfi and Dr Jane Whitley for their work in the cloning of the tammar *ELP* gene, Prof Geoff Shaw for his assistance in tammar tissue dissections and Scott Brownlees, Kerry Martin and Jenni Carfi for assistance with animal handling. We also thank Dr Matthew Digby and Sonia Mailer for their work on the microarray experiments. We thank Prof Geoff Shaw and Keng Yih Chew for the provision of tammar photos and Dr Andrew Pask for helpful comments on this manuscript. This research was funded by the Cooperative Research Centre for Innovative Dairy Products (CRC-IDP) and the Department of Zoology, The University of Melbourne. EAP was supported by an Australian Postgraduate Award, AAD by a Melbourne Research Scholarship and both EAP and AAD were recipients of a scholarship top-up from the CRC-IDP.

Author details

[1]Department of Zoology, The University of Melbourne, Melbourne, Victoria 3010, Australia. [2]Cooperative Research Centre for Innovative Dairy Products [3]ARC Centre of Excellence for Kangaroo Genomics. [4]Faculty of Veterinary Science, The University of Sydney Sydney, NSW 2006, Australia. [5]Institute for Technology Research and Innovation, Deakin UniversityGeelong, Victoria 3214, Australia.

Authors' note

After the submission of this manuscript, we identified the *ELP* gene in the Tasmanian devil (*Sarcophilus harrisii*) by *in silico* analysis of the DEVIL7.0 assembly.

Received: 29 July 2011 Accepted: 8 June 2012
Published: 8 June 2012

References

1. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: **The delayed rise of present-day mammals.** *Nature* 2007, **446:**507–512.
2. Luo ZX, Yuan CX, Meng QJ, Ji Q: **A Jurassic eutherian mammal and divergence of marsupials and placentals.** *Nature* 2011, **476:**442–445.
3. Luo Z-X, Ji Q, Wible JR, Yuan C-X: **An early Cretaceous Tribosphenic mammal and Metatherian evolution.** *Science* 2003, **302:**1934–1940.
4. Oftedal OT: **Lactation: land mammals, species comparisons**. In *Encyclopedia of Animal Science*. 2nd edition. Edited by Ullrey DE, Baer CK, Pond WG. New York: Taylor & Francis: Published online:19 Nov 2010; 2011:664–666.
5. Tyndale-Biscoe CH, Renfree MB: *Reproductive Physiology of Marsupials*. Cambridge, UK: Cambridge University Press; 1987.
6. Lefèvre CM, Sharp JA, Nicholas KR: **Evolution of lactation: ancient origin and extreme adaptations of the lactation system.** *Annu Rev Genomics Hum Genet* 2010, **11:**219–238.
7. Selwood L, Woolley PA: **A timetable of embryonic development, and ovarian and uterine changes during pregnancy, in the stripe-faced dunnart, *Sminthopsis macroura* (Marsupialia: Dasyuridae).** *J Reprod Fertil* 1991, **91:**213–227.
8. Hughes RL: **Reproduction in the macropod marsupial *Potorous tridactylus* (Kerr).** *Aust J Zool* 1962, **10:**193–224.
9. Stewart F: **Mammogenesis and changing prolactin receptor concentrations in the mammary glands of the tammar wallaby (*Macropus eugenii*).** *J Reprod Fertil* 1984, **71:**141–148.
10. Green B, Newgrain K, Merchant J: **Changes in milk composition during lactation in the tammar wallaby (*Macropus eugenii*).** *Aust J Biol Sci* 1980, **33:**35–42.
11. Joss JL, Molloy MP, Hinds L, Deane E: **A longitudinal study of the protein components of marsupial milk from birth to weaning in the tammar wallaby (*Macropus eugenii*).** *Dev Comp Immunol* 2009, **33:**152–161.
12. Tyndale-Biscoe CH: *Life of Marsupials*. Collingwood, VIC, Australia: CSIRO Publishing; 2005.
13. Nicholas K, Simpson K, Wilson M, Trott J, Shaw D: **The tammar wallaby: a model to study putative autocrine-induced changes in milk composition.** *J Mammary Gland Biol Neoplasia* 1997, **2:**299–310.
14. Hayssen V, Lacy RC, Parker PJ: **Metatherian reproduction - transitional or transcending.** *Am Nat* 1985, **126:**617–632.
15. Renfree MB: **Marsupial reproduction the choice between placentation and lactation.** *Oxf Rev Reprod Biol* 1983, **5:**1–29.
16. Kruse PE: **The importance of colostral immunoglobulins and their absorption from the intestine of the newborn animals.** *Ann Rech Vet* 1983, **14:**349–353.
17. Renfree MB, Fletcher TP, Blanden DR, Lewis DR, Shaw G, Gordon K, Short RV, Parer-Cook E, Parer D: **Physiological and behavioural events around the time of birth in macropodid marsupials**. In *Kangaroos, Wallabies and Rat-Kangaroos*. Edited by Grigg G, Jarman PJ, Hume ID. Sydney: Surrey Beatty and Sons Pty. Ltd; 1989:323–337.
18. Tyndale-Biscoe CH, Janssens PA: *The Developing Marsupial - Models for Biomedical Research*. Berlin, Germany: Springer; 1988.
19. Lefèvre CM, Digby MR, Whitley JC, Strahm Y, Nicholas KR: **Lactation transcriptomics in the Australian marsupial, Macropus eugenii: transcript sequencing and quantification.** *BMC Genomics* 2007, **8:**417.
20. Simpson K, Shaw D, Nicholas K: **Developmentally-regulated expression of a putative protease inhibitor gene in the lactating mammary gland of the tammar wallaby, *Macropus eugenii*.** *Comp Biochem Phys B* 1998, **120:**535–541.
21. Simpson K: **The tammar wallaby, *Macropus eugenii*, a model to study autocrine and endocrine regulation of lactation.** *PhD Thesis*. Latrobe University: Department of Agriculture; 1998.
22. Simpson KJ, Ranganathan S, Fisher JA, Janssens PA, Shaw DC, Nicholas KR: **The gene for a novel member of the whey acidic protein family encodes**

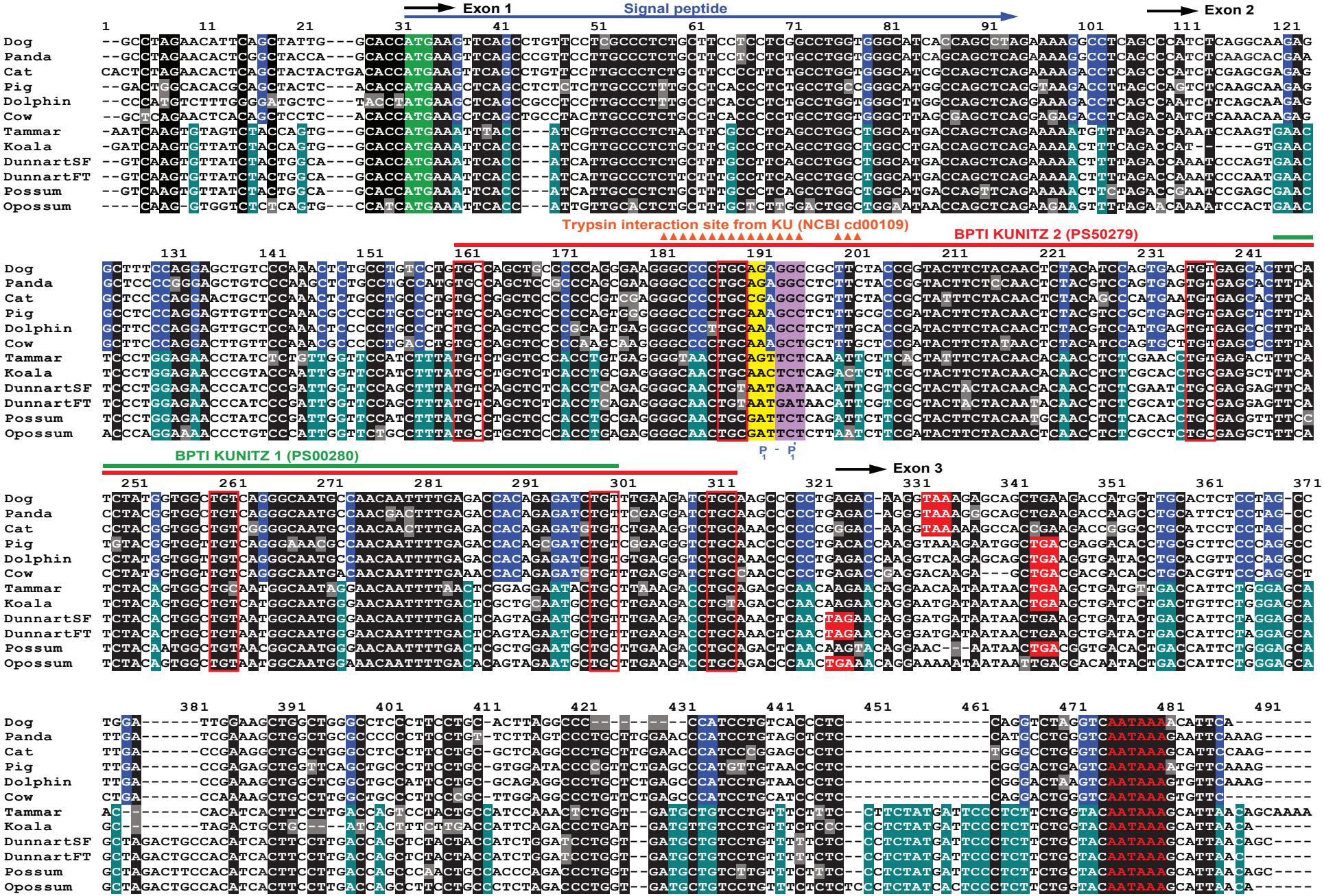
three four-disulfide core domains and is asynchronously expressed during lactation. *J Biol Chem* 2000, **275:**23074–23081.
23. Trott JF, Wilson MJ, Hovey RC, Shaw DC, Nicholas KR: **Expression of novel lipocalin-like milk protein gene is developmentally-regulated during lactation in the tammar wallaby, Macropus eugenii.** *Gene* 2002, **283:**287–297.
24. Nicholas KR, Messer M, Elliott C, Maher F, Shaw DC: **A novel whey protein synthesized only in late lactation by the mammary gland from the tammar (*Macropus eugenii*).** *Biochem J* 1987, **241:**899–904.
25. Piotte CP, Grigor MR: **A novel marsupial protein expressed by the mammary gland only during the early lactation and related to the Kunitz proteinase inhibitors.** *Arch Biochem Biophys* 1996, **330:**59–64.
26. Joss J, Molloy M, Hinds L, Deane E: **Proteomic analysis of early lactation milk of the tammar wallaby (*Macropus eugenii*).** *Comp Biochem Physiol Part D Genomics Proteomics* 2007, **2:**150–164.
27. De Leo AA, Lefevre C, Topcic D, Pharo E, Cheng JF, Frappell P, Westerman M, Graves JAM, Nicholas KR: **Characterization of two whey protein genes in the Australian dasyurid marsupial, the stripe-faced dunnart (*Sminthopsis macroura*).** *Cytogenet Genome Res* 2006, **115:**62–69.
28. Demmer J, Ross IK, Ginger MR, Piotte CK, Grigor MR: **Differential expression of milk protein genes during lactation in the common brushtail possum (*Trichosurus vulpecula*).** *J Mol Endocrinol* 1998, **20:**37–44.
29. Old JM, Deane EM: **Development of the immune system and immunological protection in marsupial pouch young.** *Dev Comp Immunol* 2000, **24:**445–454.
30. Basden K, Cooper DW, Deane EM: **Development of the lymphoid tissues of the tammar wallaby *Macropus eugenii*.** *Reprod Fertil Dev* 1997, **9:**243–254.
31. Laskowski M Jr, Laskowski M: **Crystalline trypsin inhibitor from colostrum.** *J Biol Chem* 1951, **190:**563–573.
32. Laskowski M, Kassell B, Hagerty G: **A crystalline trypsin inhibitor from swine colostrum.** *Biochim Biophys Acta* 1957, **24:**300–305.
33. Baintner K: **Occurrence of trypsin inhibitors in colostrum, meconium, and faeces of different species of ungulates and carnivores.** *Acta Vet Hung* 1984, **32:**91–95.
34. Baintner K, Csapo J: **Lack of acid-resistant trypsin inhibitor in mare's colostrum: short communication.** *Acta Vet Hung* 1996, **44:**95–97.
35. Kassell B, Laskowski M: **The comparative resistance to pepsin of six naturally occurring trypsin inhibitors.** *J Biol Chem* 1956, **219:**203–210.
36. Cechova D, Svestkova W, Keil B, Sorm F: **Similarities in primary structures of cow colostrum trypsin inhibitor and bovine basic pancreatic trypsin inhibitor.** *FEBS Lett* 1969, **4:**155–156.
37. Veselsky L, Cechova D, Jonakova V: **Secretion and immunochemical properties of the trypsin inhibitor from bovine colostrum.** *Hoppe Seylers Z Physiol Chem* 1978, **359:**873–878.
38. Ikeo K, Takahashi K, Gojobori T: **Evolutionary origin of a Kunitz-type trypsin inhibitor domain inserted in the amyloid beta precursor protein of Alzheimer's disease.** *J Mol Evol* 1992, **34:**536–543.
39. Rawlings ND, Tolle DP, Barrett AJ: **Evolutionary families of peptidase inhibitors.** *Biochem J* 2004, **378:**705–716.
40. Yuan CH, He QY, Peng K, Diao JB, Jiang LP, Tang X, Liang SP: **Discovery of a distinct superfamily of Kunitz-type toxin (KTT) from tarantulas.** *PLoS One* 2008, **3:**e3414.
41. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JD, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, de la Vega RC: **The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms.** *Annu Rev Genomics Hum Genet* 2009, **10:**483–511.
42. Schweitz H, Bruhn T, Guillemare E, Moinier D, Lancelin JM, Beress L, Lazdunski M: **Kalicludines and kaliseptine. Two different classes of sea anemone toxins for voltage sensitive K + channels.** *J Biol Chem* 1995, **270:**25121–25126.
43. **MEROPS: the peptidase database** http://merops.sanger.ac.uk/.
44. Rawlings ND: **Peptidase inhibitors in the MEROPS database.** *Biochimie* 2010, **92:**1463–1683.
45. Ascenzi P, Bocedi A, Bolognesi M, Spallarossa A, Coletta M, De Cristofaro R, Menegatti E: **The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor): a milestone protein.** *Curr Protein Pept Sci* 2003, **4:**231–251.
46. Huber R, Kukla D, Ruhlmann A, Steigemann W: **Pancreatic trypsin inhibitor (Kunitz). I. Structure and function.** *Cold Spring Harb Symp Quant Biol* 1972, **36:**141–148.
47. Laskowski M, Kato I: **Protein inhibitors of proteinases.** *Annu Rev Biochem* 1980, **49:**593–626.
48. Laskowski M Jr: **Protein inhibitors of serine proteinases - mechanism and classification.** *Adv Exp Med Biol* 1986, **199:**1–17.
49. **Ensembl Genome Browser** http://www.ensembl.org/.
50. Chakrabarty A, Green JA, Roberts RM: **Origin and evolution of the TKDP gene family.** *Gene* 2006, **373:**35–43.
51. Chakrabarty A, MacLean JA, Hughes AL, Roberts RM, Green JA: **Rapid evolution of the trophoblast Kunitz domain proteins (TKDPs) - A multigene family in ruminant ungulates.** *J Mol Evol* 2006, **63:**274–282.
52. Rawlings ND, Barrett AJ, Bateman A: **MEROPS: the peptidase database.** *Nucleic Acids Res* 2010, **38:**D227–233.
53. Renfree MB, Papenfuss AT, Deakin JE, Lindsay J, Heider T, Belov K, Rens W, Waters PD, Pharo EA, Shaw G, Wong ES, Lefèvre CM, Nicholas KR, Kuroki Y, Wakefield MJ, Zenger KR, Wang C, Ferguson-Smith M, Nicholas FW, Hickford D, Yu H, Short KR, Siddle HV, Frankenberg SR, Chew KY, Menzies BR, Stringer JM, Suzuki S, Hore TA, Delbridge ML, *et al*: **Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development.** *Genome Biol* 2011, **12:**R81.
54. **NCBI BLAST** http://www.ncbi.nlm.nih.gov/BLAST/.
55. **UCSC Genome Browser** http://genome.ucsc.edu/.
56. **Center for Biological Sequence analysis prediction servers (SignalP 3.0 and NetNGlyc 1.0)** http://www.cbs.dtu.dk/services/.
57. Perona JJ, Tsu CA, Craik CS, Fletterick RJ: **Crystal Structure of Rat Anionic Trypsin Complexed with the Protein Inhibitors APPI and BPTI.** *J Mol Biol* 1993, **230:**919–933.
58. Klauser R, Cechova D, Tschesche H: **The carbohydrate linkage site of cow colostrum trypsin inhibitor.** *Hoppe Seylers Z Physiol Chem* 1978, **359:**173–180.
59. Nei M, Kumar S: *Molecular Evolution and Phylogenetics*. New York: Oxford University Press; 2000.
60. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28:**2731–2739.
61. Hughes AL: **The evolution of functionally novel proteins after gene duplication.** *Proc Biol Sci* 1994, **256:**119–124.
62. Clauss A, Lilja H, Lundwall A: **A locus on human chromosome 20 contains several genes expressing protease inhibitor domains with homology to whey acidic protein.** *Biochem J* 2002, **368:**233–242.
63. MacLean JA 2nd, Chakrabarty A, Xie S, Bixby JA, Roberts RM, Green JA: **Family of Kunitz proteins from trophoblast: expression of the trophoblast Kunitz domain proteins (TKDP) in cattle and sheep.** *Mol Reprod Dev* 2003, **65:**30–40.
64. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I: **VISTA: computational tools for comparative genomics.** *Nucleic acids research* 2004, **32:**W273–W279.
65. Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J: **Retroposed elements as archives for the evolutionary history of placental mammals.** *PLoS Biol* 2006, **4:**e91.
66. Kohany O, Gentles AJ, Hankus L, Jurka J: **Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor.** *BMC Bioinformatics* 2006, **7:**474.
67. Huttley GA, Wakefield MJ, Easteal S: **Rates of genome evolution and branching order from whole genome analysis.** *Mol Biol Evol* 2007, **24:**1722–1730.
68. Bird PH, Hendry KA, Shaw DC, Wilde CJ, Nicholas KR: **Progressive changes in milk protein gene expression and prolactin binding during lactation in the tammar wallaby (*Macropus eugenii*).** *J Mol Endocrinol* 1994, **13:**117–125.
69. Sharp JA, Digby M, Lefevre C, Mailer S, Khalil E, Topcic D, Auguste A, Kwek JH, Brennan AJ, Familari M, Nicholas KR: **The comparative genomics of tammar wallaby and Cape fur seal lacation models to examine functions of milk proteins**. In *Milk Proteins: from Expression to Food*. Edited by Thompson A, Boland M, Singh H. San Diego: Academic Press/Elsevier; 2009:55–79. Food science and technology, international series.
70. Luo Z-X: **Transformation and diversification in early mammal evolution.** *Nature* 2007, **450:**1011–1019.
71. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ: **Molecular phylogenetics and the origins of placental mammals.** *Nature* 2001, **409:**614–618.

72. MacLean JA 2nd, Roberts RM, Green JA: **Atypical Kunitz-type serine proteinase inhibitors produced by the ruminant placenta.** *Biol Reprod* 2004, **71:**455–463.
73. Kingston IB, Anderson S: **Sequences encoding two trypsin inhibitors occur in strikingly similar genomic environments.** *Biochem J* 1986, **233:**443–450.
74. Jurka J, Kaplan DJ, Duncan CH, Walichiewicz J, Milosavljevic A, Murali G, Solus JF: **Identification and characterization of new human medium reiteration frequency repeats.** *Nucleic Acids Res* 1993, **21:**1273–1279.
75. Wagner G, Wutherich K, Tschesche H: **A 1 H nuclear-magnetic-resonance study of the conformation and the molecular dynamics of the glycoprotein cow-colostrum trypsin inhibitor.** *Eur J Biochem* 1978, **86:**67–76.
76. Laskowski M, Qasim MA: **What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes?** *Biochim Biophys Acta* 2000, **1477:**324–337.
77. Cechova D, Muszynska G: **Role of lysine 18 in active center of cow colostrum trypsin inhibitor.** *FEBS Lett* 1970, **8:**84–86.
78. Peigneur S, Billen B, Derua R, Waelkens E, Debaveye S, Beress L, Tytgat J: **A bifunctional sea anemone peptide with Kunitz type protease and potassium channel inhibiting properties.** *Biochem Pharmacol* 2011, **82:**81–90.
79. Jordan SP, Mao SS, Lewis SD, Shafer JA: **Reaction pathway for inhibition of blood coagulation factor Xa by tick anticoagulant peptide.** *Biochemistry* 1992, **31:**5374–5380.
80. Tschesche H, Klauser R, Cechova D, Jonakova V: **On the carbohydrate composition of bovine colostrum trypsin inhibitor.** *Hoppe Seylers Z Physiol Chem* 1975, **356:**1759–1764.
81. Walsh CT: *Posttranslational modification of proteins: expanding nature's inventory*. Greenwood Village, Colorado: Roberts and Company Publishers; 2006.
82. Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J: *(Eds): Essentials of Glycobiology*. 2nd edition. Cold Spring Harbor: Cold Spring Harbor Laboratories Press; 2009.
83. Roth J, Zuber C, Park S, Jang I, Lee Y, Kysela KG, Le Fourn V, Santimaria R, Guhl B, Cho JW: **Protein N-glycosylation, protein folding, and protein quality control.** *Mol Cells* 2010, **30:**497–506.
84. Shoaf-Sweeney KD, Hutkins RW: **Adherence, anti-adherence, and oligosaccharides preventing pathogens from sticking to the host.** *Adv Food Nutr Res* 2009, **55:**101–161.
85. Kunz C, Rudloff S: **Potential anti-inflammatory and anti-infectious effects of human milk oligosaccharides.** *Adv Exp Med Biol* 2008, **606:**455–465.
86. Chung JJ, Shikano S, Hanyu Y, Li M: **Functional diversity of protein C-termini: more than zipcoding?** *Trends Cell Biol* 2002, **12:**146–150.
87. Kwek J, De Iongh R, Nicholas K, Familari M: **Molecular insights into evolution of the vertebrate gut: focus on stomach and parietal cells in the marsupial, *Macropus eugenii*.** *J Exp Zool B Mol Dev Evol* 2009, **312:**613–624.
88. Yadav M: **The transmission of antibodies across the gut of pouch young marsupials.** *Immunology* 1971, **21:**839–851.
89. Brambell FWR: *The Transmission of Passive Immunity from Mother to Young*. Amsterdam: North-Holland Publishing Company; 1970.
90. Fioretti E, Angeletti M, Fiorucci L, Barra D, Bossa F, Ascoli F: **Aprotinin-like isoinhibitors in bovine organs.** *Biol Chem Hoppe Seyler* 1988, **369** (Suppl):37–42.
91. Rastogi S, Liberles D: **Subfunctionalization of duplicated genes as a transition state to neofunctionalization.** *BMC Evolutionary Biology* 2005, **5:**28.
92. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151:**1531–1545.
93. Lemay D, Lynn D, Martin W, Neville M, Casey T, Rincon G, Kriventseva E, Barris W, Hinrichs A, Molenaar A, Pollard K, Maqbool N, Singh K, Murney R, Zdobnov E, Tellam R, Medrano J, German JB, Rijnkels M: **The bovine lactation genome: insights into the evolution of mammalian milk.** *Genome Biology* 2009, **10:**R43.
94. NHMRC: *Australian Code of Practice for the Care and Use of Animals for Scientific Purposes*. 7th edition. Canberra: Australian Government; 2004.
95. Boutinaud M, Jammes H: **Potential uses of milk epithelial cells: a review.** *Reprod Nutr Dev* 2002, **42:**133–147.
96. Collet C, Joseph R, Nicholas K: **A marsupial β-lactoglobulin gene: characterization and prolactin-dependent expression.** *J Mol Endocrinol* 1991, **6:**9–16.
97. Sambrook J: *Russell DW: Molecular Cloning: A Laboratory Manual*. 3rd edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2001.
98. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403–410.
99. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome Res* 1999, **9:**868–877.
100. **CAP3** http://pbil.univ-lyon1.fr/cap3.php.
101. Toder R, Wilcox SA, Smithwick M, Graves JA: **The human/mouse imprinted genes *IGF2, H19, SNRPN* and *ZNF127* map to two conserved autosomal clusters in a marsupial.** *Chromosome Res* 1996, **4:**295–300.
102. Daly KA, Digby M, Lefèvre C, Mailer S, Thomson P, Nicholas K, Williamson P: **Analysis of the expression of immunoglobulins throughout lactation suggests two periods of immune transfer in the tammar wallaby (*Macropus eugenii*).** *Vet Immunol Immunopathol* 2007, **120:**187–200.
103. Khalil E: **Identification of novel milk proteins and their functions by exploiting the lactation strategy of the tamar wallaby (Macropus eugenii).** *PhD Thesis*, . University of Melbourne: Department of Zoology; 2007.
104. Khalil E, Digby MR, Thomson PC, Lefèvre C, Mailer SL, Pooley C, Nicholas KR: **Acute involution in the tammar wallaby: Identification of genes and putative novel milk proteins implicated in mammary gland function.** *Genomics* 2011, **97:**372–378.
105. **Bioconductor** http://www.bioconductor.org.
106. Thomson PC: **Analysis of microarray data: a mixed model finite-mixture approach [Abstract].** In *Proceedings of the XXIIIrd International Biometric Conference*. Montréal: McGill University; 2006. International Biometric Society.
107. **ArrayExpress Archive** http://www.ebi.ac.uk/arrayexpress/.
108. **CENSOR** http://www.girinst.org/censor/.
109. Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268:**78–94.
110. **Prosite** http://prosite.expasy.org/.
111. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23:**2947–2948.
112. Campanella JJ, Bitincka L, Smalley J: **MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences.** *BMC Bioinformatics* 2003, **4:**29.
113. Tamura K, Nei M, Kumar S: **Prospects for inferring very large phylogenies by using the neighbor-joining method.** *Proc Natl Acad Sci U S A* 2004, **101:**11030–11035.
114. Nei M, Gojobori T: **Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions.** *Mol Biol Evol* 1986, **3:**418–426.
115. Zhang J, Rosenberg HF, Nei M: **Positive Darwinian selection after gene duplication in primate ribonuclease genes.** *Proc Natl Acad Sci U S A* 1998, **95:**3708–3713.
116. Felsenstein J: **Phylip (Phylogeny Inference Package) version 3.69** http://evolution.genetics.washington.edu/phylip.html.
117. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32:**1792–1797.
118. Kimura M: **A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.** *J Mol Evol* 1980, **16:**111–120.
119. Fitch WM, M E: **Construction of phylogenetic trees.** *Science* 1967, **155:**279–284.

doi:10.1186/1471-2148-12-80
Cite this article as: Pharo *et al.*: **The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene.** *BMC Evolutionary Biology* 2012 **12**:80.

Additional file 1 - Table S1. Characterisation of the putative functional *ELP*/*CTI* gene, transcript and protein

Gene	Species	Gene (*in vitro*)	Gene (*in silico* only)	Transcript (*in vitro*)	Transcript (*in silico* only)	Protein (milk/ colostrum)	References
ELP	Fat-tailed dunnart	✓		✓	✓		[this study, 27]
	Stripe-faced dunnart	✓			✓		[this study, 27]
	koala	✓			✓		[this study]
	Opossum		✓	✓	✓		[this study, 27]
	Possum	✓		✓	✓	✓	[this study, 25, 28]
	Tammar wallaby	✓		✓		✓	[this study, 11, 13, 20, 21, 26]
CTI	Cat		✓		✓		[this study]
	Cow		✓	✓		✓	[this study, 31, 36, 37]
	Dolphin		✓		✓		[this study]
	Dog		✓	✓			[this study]
	Panda		✓		✓		[this study]
	Pig		✓		✓		[this study]



Additional file 2 - Figure S1. Alignment of the marsupial *ELP* and eutherian *CTI* transcripts

Additional file 3 - Table S2. Percentage similarity between and within the marsupial *ELP* and eutherian *CTI* transcripts

Pairwise similarities were calculated with MatGat version 2.01 using the BLOSUM62 matrix.
2A. ELP/CTI transcripts (ATG to AATAAA, inclusive)**, 2B.** ELP/CTI transcripts (ATG to stop codon, inclusive)**,**
2C. Marsupial ELP 3'UTR

2A. *ELP/CTI Transcripts*[1]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. Tammar											
2. DunnartSF	83.5										
3. DunnartFT	82.8	99.1									
4. Koala	83.4	83.5	83.1								
5. Opossum	79.1	82.9	82.7	80.7							
6. Possum	85.9	86.9	86.9	85.7	82.5						
7. Cat	58.6	60.3	59.4	61.7	59.8	62.0					
8. Cow	56.8	58.2	57.8	61.2	60.7	58.9	75.6				
9. Dog	58.9	58.9	57.3	62.6	62.0	58.0	82.4	76.7			
10. Dolphin	56.8	59.4	59.1	60.3	58.2	58.9	79.7	86.0	79.4		
11. Panda	58.2	59.8	59.4	62.2	63.6	59.8	84.3	76.1	86.6	79.4	
12. Pig	56.1	57.6	57.6	61.4	58.9	57.5	75.2	81.5	75.0	86.7	75.5

[1] ATG to AATAAA, inclusive

2B. *ELP/CTI Transcripts*[2]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. Tammar											
2. DunnartSF	78.5										
3. DunnartFT	77.6	98.6									
4. Koala	85.6	80.7	80.1								
5. Opossum	74.0	81.4	81.1	78.4							
6. Possum	86.2	82.2	82.2	89.3	78.0						
7. Cat	59.6	61.4	61.4	63.4	61.7	65.4					
8. Cow	59.0	56.1	56.1	62.8	61.2	60.9	75.6				
9. Dog	60.6	59.7	59.7	63.1	64.4	60.5	86.5	78.2			
10. Dolphin	59.7	57.5	57.5	61.9	57.5	60.6	79.0	87.3	79.4		
11. Panda	58.7	60.7	60.7	64.1	64.4	62.1	85.1	75.6	90.4	78.7	
12. Pig	58.1	57.1	57.1	63.8	59.0	59.7	76.5	82.2	76.8	88.3	76.2

[2] ATG to stop codon, inclusive

2C. Marsupial *ELP* 3'UTR	**1**	**2**	**3**	**4**	**5**
1. Tammar					
2. DunnartSF	88.5				
3. DunnartFT	88.5	100.0			
4. Koala	88.5	88.5	88.5		
5. Opossum	87.0	87.0	87.0	87.0	
6. Possum	92.3	92.3	92.3	88.5	87.0

Additional file 4 - Table S3. Percentage similarity between and within the marsupial ELP and eutherian CTI peptides

3A. ELP/CTI mature peptide, **3B.**ELP/CTI signal peptide, **3C.** ELP/CTI N-terminus, **3D.** ELP/CTI Kunitz domain motif 2 (51 amino acids), **3E.** ELP/CTI Kunitz domain motif 1 (19 amino acids) and **3F.** ELP/CTI C-terminus.

3A. ELP/CTI Mature Peptide	1	2	3	4	5	6	7	8	9	10	11
1. Tammar											
2. DunnartSF	71.1										
3. DunnartFT	71.1	100.0									
4. Koala	79.5	76.5	76.5								
5. Opossum	67.5	81.6	81.6	76.5							
6. Possum	80.7	75.6	75.6	81.7	72.0						
7. Cat	54.2	54.4	54.4	59.3	50.6	54.9					
8. Cow	56.6	53.7	53.7	62.2	51.2	61.0	73.2				
9. Dog	56.6	57.0	57.0	60.5	55.7	61.0	88.6	74.4			
10. Dolphin	56.6	53.0	53.0	57.8	48.2	60.2	81.9	86.7	78.3		
11. Panda	49.4	53.2	53.2	58.0	46.8	51.2	84.8	70.7	86.1	73.5	
12. Pig	53.0	53.0	53.0	51.8	44.6	55.4	73.5	80.7	73.5	85.5	68.7

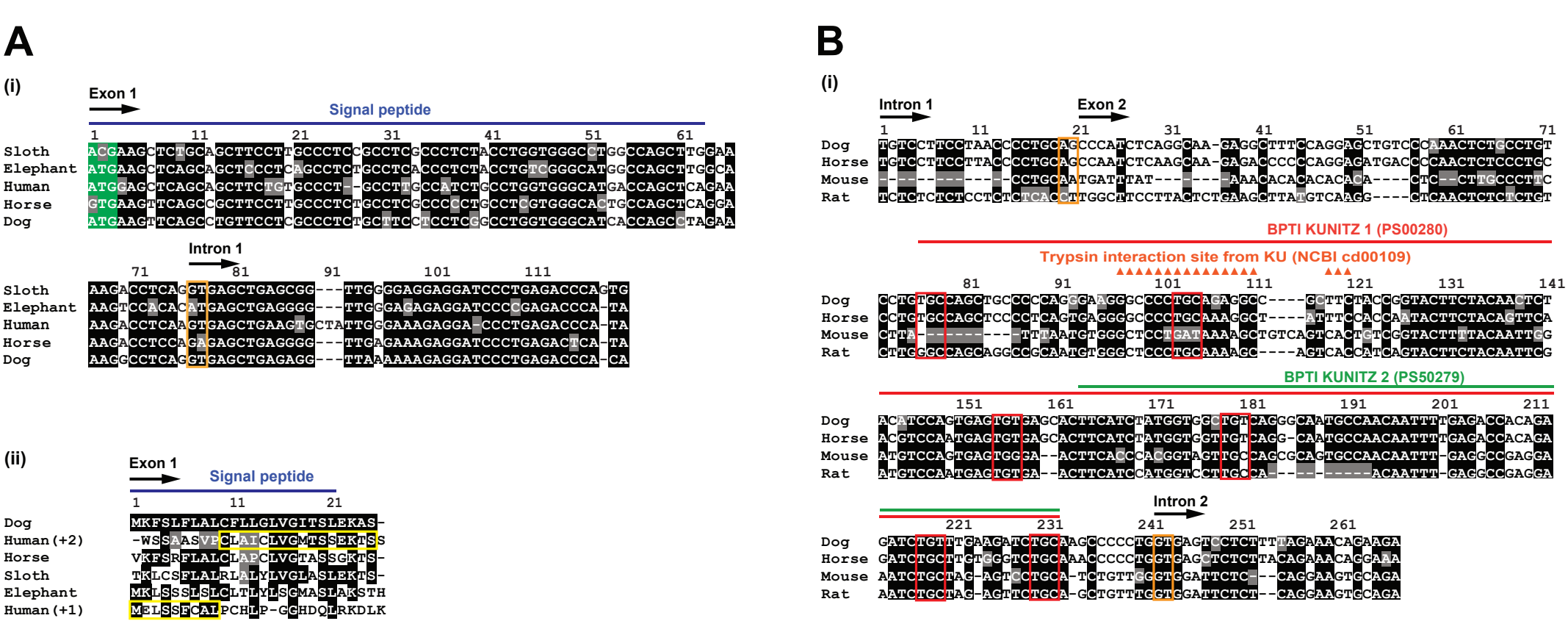
3B. ELP/CTI Signal Peptide	1	2	3	4	5	6	7	8	9	10	11
1. Tammar											
2. DunnartSF	90.0										
3. DunnartFT	95.0	95.0									
4. Koala	95.0	95.0	90.0								
5. Opossum	90.0	90.0	85.0	95.0							
6. Possum	85.0	85.0	80.0	90.0	85.0						
7. Cat	66.7	76.2	71.4	71.4	71.4	71.4					
8. Cow	57.1	61.9	57.1	61.9	61.9	71.4	71.4				
9. Dog	71.4	71.4	66.7	76.2	81.0	76.2	76.2	57.1			
10. Dolphin	61.9	61.9	57.1	66.7	66.7	76.2	71.4	90.5	61.9		
11. Panda	71.4	71.4	66.7	76.2	76.2	76.2	85.7	71.4	85.7	76.2	
12. Pig	61.9	61.9	57.1	66.7	66.7	71.4	71.4	81.0	61.9	85.7	71.4

3C. ELP/CTI N-terminus	1	2	3	4	5	6	7	8	9	10	11
1. Tammar											
2. DunnartSF	81.8										
3. DunnartFT	81.8	100.0									
4. Koala	63.6	77.3	77.3								
5. Opossum	59.1	72.7	72.7	59.1							
6. Possum	86.4	86.4	86.4	72.7	59.1						
7. Cat	50.0	45.5	45.5	59.1	27.3	50.0					
8. Cow	45.5	40.9	40.9	54.5	22.7	50.0	72.7				
9. Dog	50.0	45.5	45.5	50.0	36.4	54.5	77.3	68.2			
10. Dolphin	40.9	36.4	36.4	45.5	18.2	45.5	86.4	86.4	72.7		
11. Panda	50.0	45.5	45.5	59.1	18.2	54.5	81.8	59.1	81.8	68.2	
12. Pig	45.5	40.9	40.9	40.9	18.2	50.0	72.7	86.4	63.6	90.9	59.1

3D. ELP/CTI Kunitz Domain motif 2 (51 amino acids)	1	2	3	4	5	6	7	8	9	10	11
1. Tammar											
2. DunnartSF	76.5										
3. DunnartFT	76.5	100.0									
4. Koala	84.3	84.3	84.3								
5. Opossum	80.4	86.3	86.3	90.2							
6. Possum	80.4	78.4	78.4	86.3	86.3						
7. Cat	62.7	62.7	62.7	62.7	62.7	64.7					
8. Cow	64.7	64.7	64.7	66.7	66.7	68.6	78.4				
9. Dog	64.7	66.7	66.7	66.7	66.7	68.6	94.1	80.4			
10. Dolphin	64.7	66.7	66.7	62.7	64.7	68.6	86.3	88.2	84.3		
11. Panda	54.9	60.8	60.8	60.8	60.8	56.9	88.2	78.4	88.2	80.4	
12. Pig	62.7	66.7	66.7	60.8	62.7	64.7	80.4	80.4	82.4	84.3	76.5

3E. ELP/CTI Kunitz Domain motif 1 (19 amino acids)	1	2	3	4	5	6	7	8	9	10	11
1. Tammar											
2. DunnartSF	84.2										
3. DunnartFT	84.2	100.0									
4. Koala	84.2	100.0	100.0								
5. Opossum	84.2	100.0	100.0	100.0							
6. Possum	84.2	100.0	100.0	100.0	100.0						
7. Cat	63.2	68.4	68.4	68.4	68.4	68.4					
8. Cow	63.2	68.4	68.4	68.4	68.4	68.4	94.7				
9. Dog	68.4	73.7	73.7	73.7	73.7	73.7	94.7	89.5			
10. Dolphin	63.2	68.4	68.4	68.4	68.4	68.4	100.0	94.7	94.7		
11. Panda	63.2	68.4	68.4	68.4	68.4	68.4	100.0	94.7	94.7	100.0	
12. Pig	63.2	68.4	68.4	68.4	68.4	68.4	89.5	84.2	94.7	89.5	89.5

3F. ELP/CTI C-terminus	1	2	3	4	5	6	7	8	9	10	11
1. Tammar											
2. DunnartSF	20.0										
3. DunnartFT	20.0	100.0									
4. Koala	90.0	20.0	20.0								
5. Opossum	20.0	66.7	66.7	30.0							
6. Possum	70.0	33.3	33.3	70.0	22.2						
7. Cat	20.0	16.7	16.7	30.0	33.3	22.2					
8. Cow	40.0	22.2	22.2	50.0	33.3	44.4	44.4				
9. Dog	30.0	16.7	16.7	40.0	33.3	33.3	83.3	55.6			
10. Dolphin	50.0	20.0	20.0	60.0	30.0	50.0	50.0	80.0	60.0		
11. Panda	20.0	16.7	16.7	30.0	33.3	11.1	66.7	55.6	83.3	50.0	
12. Pig	20.0	10.0	10.0	30.0	10.0	20.0	40.0	70.0	50.0	80.0	50.0

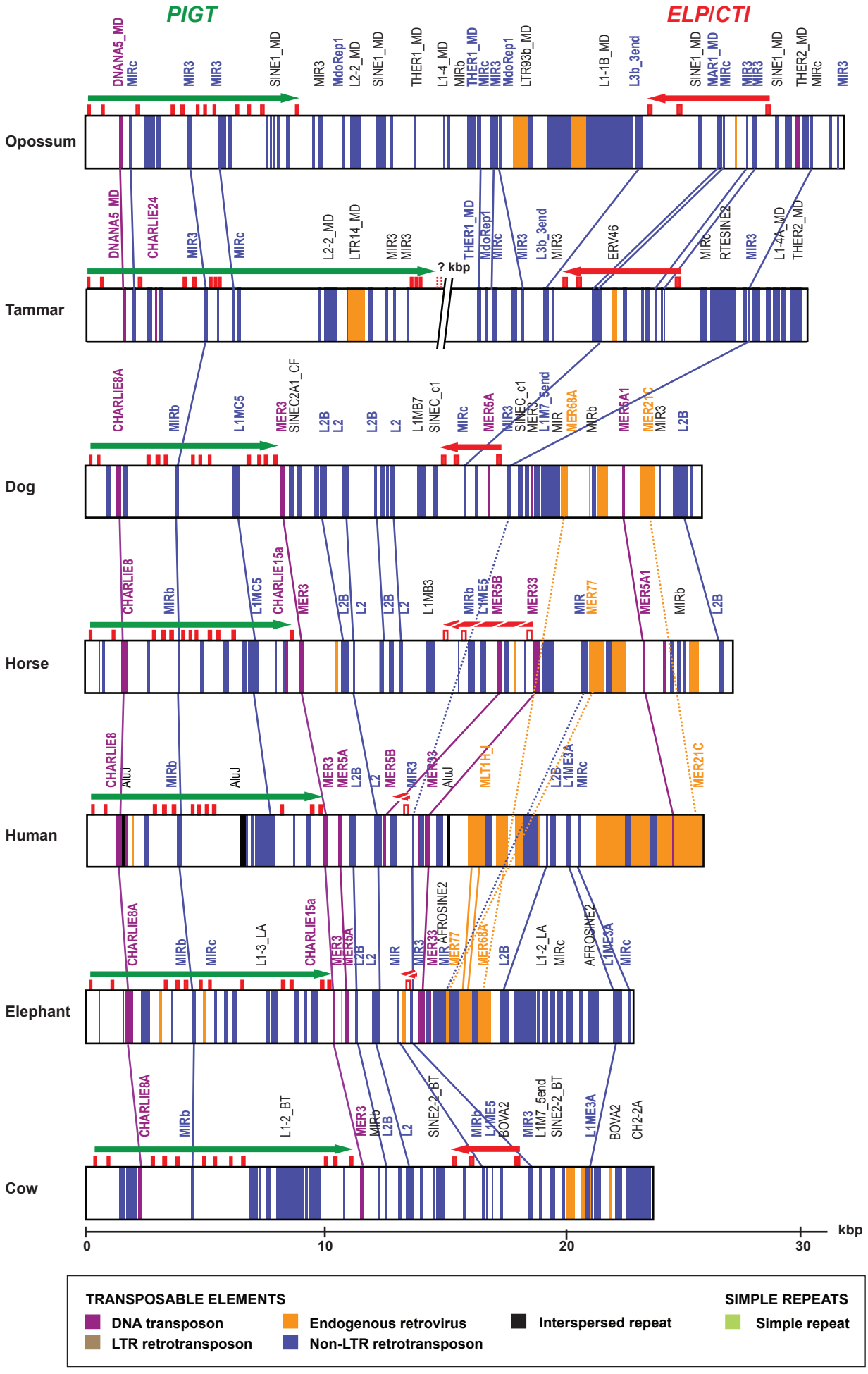


Additional file 5 - Figure S2. Exon 1 and 2 mutations within selected putative eutherian *CTI* pseudogenes.

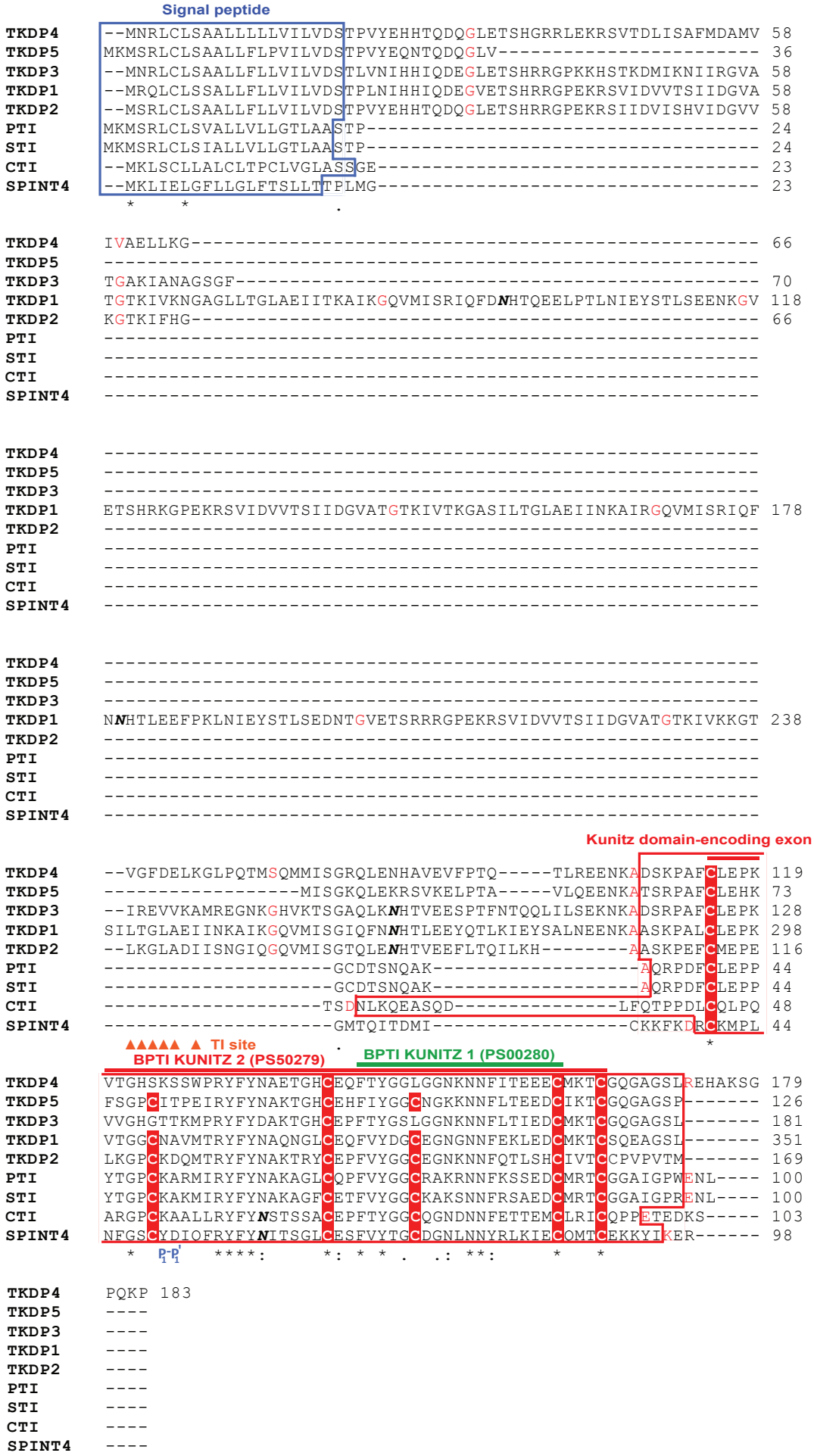
A (i). Alignment of the 'functional' canine *CTI* exon 1 with the putative disrupted exon 1 of the sloth, elephant, human and horse.**(ii)**. Alignment of the translated putative CTI exon 1 peptides of the dog, sloth, elephant and human. The +2 and +3 frames of human CTI are included in the alignment to show that deletions within exon 1 would produce a frameshift (yellow boxes).

B (i). Alignment of the functional canine *CTI* exon 2 with the putative disrupted exon 2 of the rat, mouse and horse. The BPTI KUNITZ 1 and 2 motifs are shown in green and red bars respectively. The putative trypsin interaction site from the KU motif (NCBI cd00109) is indicated by orange triangles.

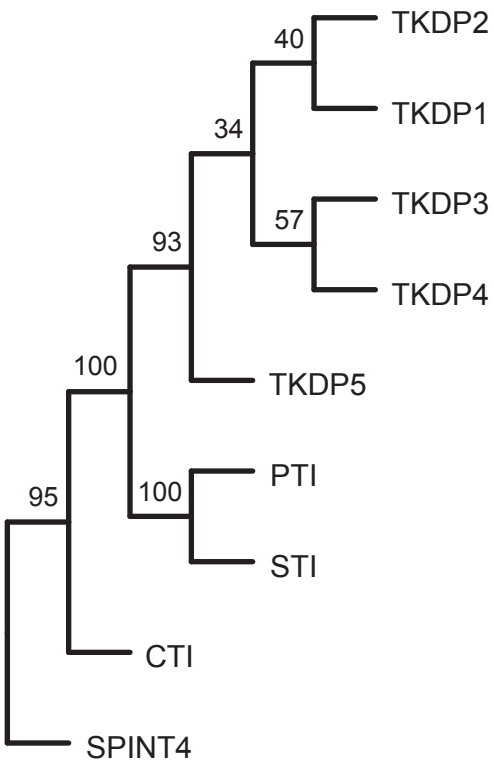
(ii). Alignment of the translated 'functional' canine *CTI* exon 2 region with the equivalent region in the rat, mouse and horse. The +2 and +3 reading frames are shown for equine CTI, the +1 and +2 frames for the mouse and all 3 reading frames for rat CTI. Yellow boxes indicate the putative frame-shift in equine ELP.



Additional file 6 - Figure S3. Transposable elements and simple repeats located within the *PIGT* and *ELP/CTI* genes and flanking regions

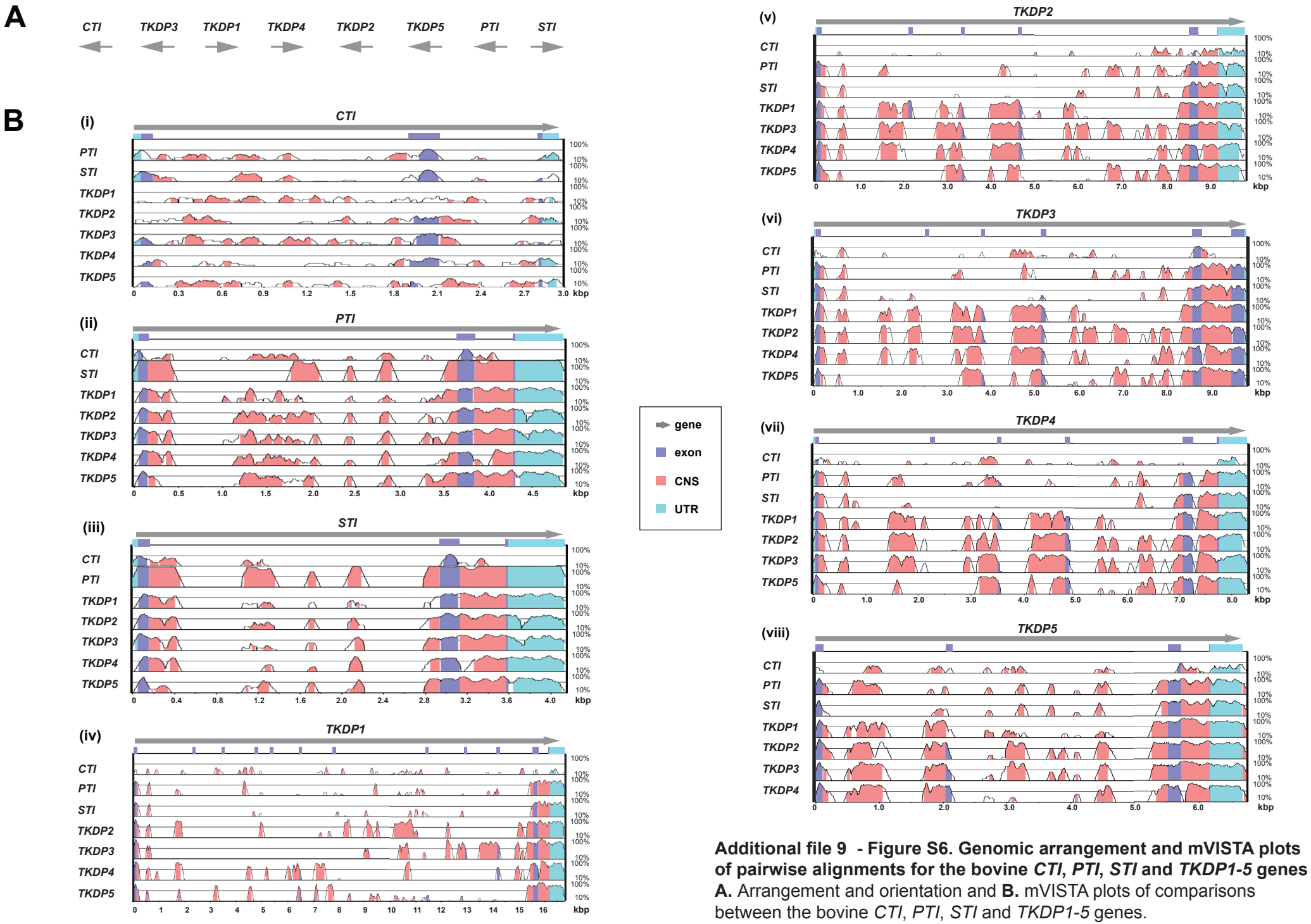


Additional file 7 - Figure S4. Alignment of the bovine CTI, PTI, STI, TKDP1-5 and SPINT4 precursor proteins

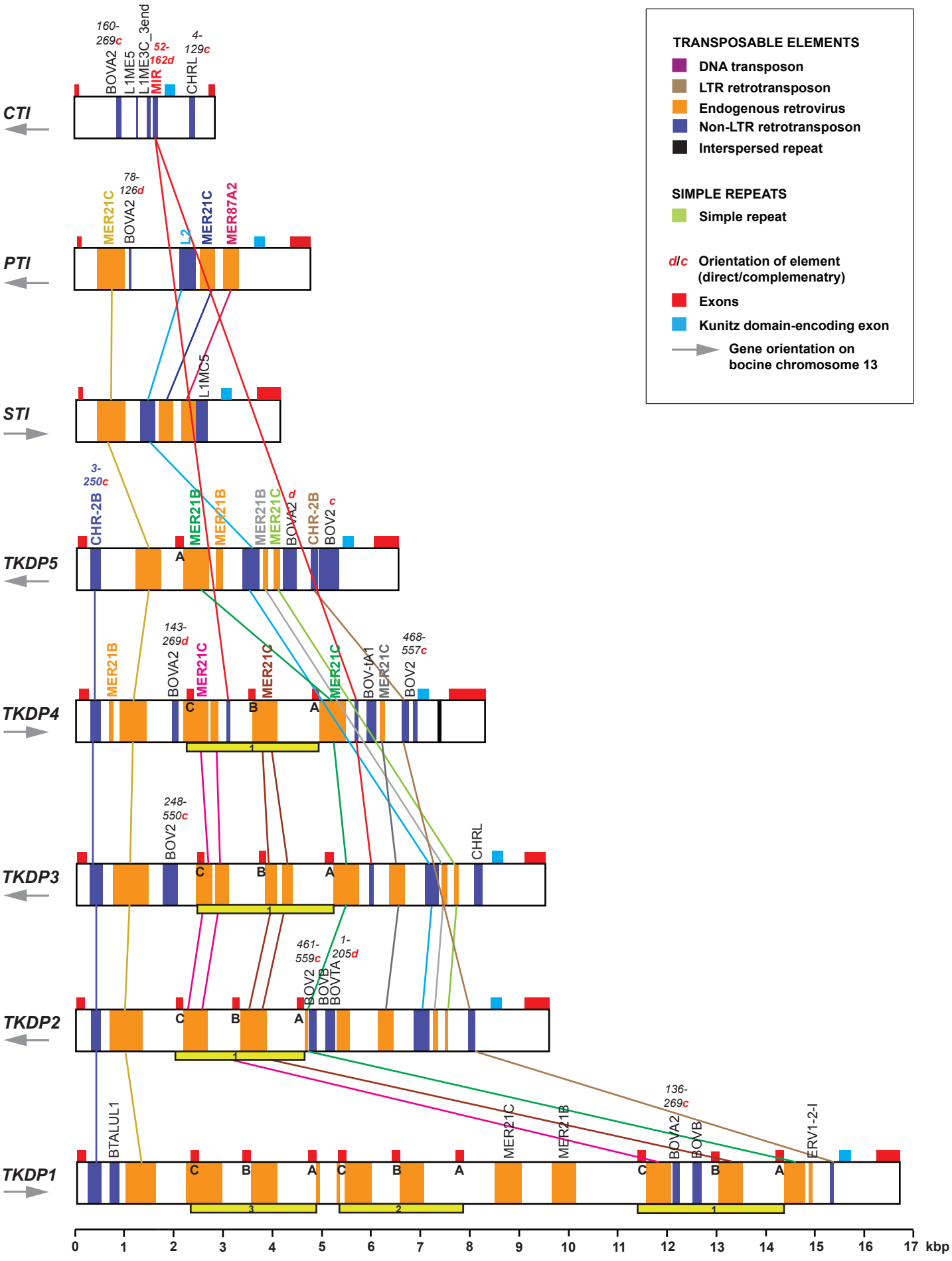


Additional file 8 - Figure S5. Relationship between bovine CTI, PTI, STI, the TKDPs and SPINT4

Phylogenetic relationship between the protein-coding regions of the *CTI*, *PTI*, *STI*, *TKDP1-5* and *SPINT4* transcripts.The tree was constructed using the maximum likelihood model of nucleotide substution based upon the assumption of a molecular clock. The *SLPI* transcript (omitted) was used to root the tree.



Additional file 9 - Figure S6. Genomic arrangement and mVISTA plots of pairwise alignments for the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes
A. Arrangement and orientation and **B.** mVISTA plots of comparisons between the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes.



Additional file 10 - Figure S7. Transposable elements located within the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes

A. Conserved repeat elements within the *CTI*, *PTI*, *STI* and *TKDP1-5* genes. **B.** Identity, location and orientation of transposable elements within the *CTI*, *PTI*, *STI* and *TKDP1-5* genes as determined using CENSOR.

Additional file 11 - Table S4. Location, identity and orientation of transposable elements within the bovine *CTI*, *PTI*, *STI* and *TKDP1-5* genes

CENSOR output tables showing the predicted identity, location and orientation of retroelement fragments within the bovine *CTI*, *PTI*, *STI*, *TKDP1*, *TKDP2*, *TKDP3*, *TKDP4* and *TKDP5* genes.

CTI

From	To	Name	From	To	Class	Dir
839	948	BOVA2	160	269	NonLTR/SINE	c
1266	1304	L1ME5	359	397	NonLTR/L1	c
1484	1566	L1ME3C_3end	426	505	NonLTR/L1	c
1612	1725	MIRb	52	162	NonLTR/SINE/SINE2	d
2388	2514	CHRL	4	129	NonLTR/SINE	d

PTI

From	To	Name	From	To	Class	Dir
400	993	MER21C	115	747	ERV/ERV3	d
1310	1624	L2	2698	3022	NonLTR/CR1	c
1705	2005	MER21C	117	403	ERV/ERV3	c
2171	2475	MER87A2_BT	500	811	ERV/ERV1	d
2484	2736	L1MC5	1859	2138	NonLTR/L1	c

STI

From	To	Name	From	To	Class	Dir
400	943	MER21C	115	690	ERV/ERV3	d
1026	1072	BOVA2	78	126	NonLTR/SINE	d
2028	2344	L2	2698	3022	NonLTR/CR1	c
2427	2721	MER21C	117	403	ERV/ERV3	c
2887	3191	MER87A2_BT	500	811	ERV/ERV1	d

TKDP1

From	To	Name	From	To	Class	Dir
221	345	CHR-2B	185	318	NonLTR/SINE	c
347	485	CHR-2B	5	144	NonLTR/SINE	c
656	847	BTALUL1	1	193	NonLTR/SINE/SINE2	c
971	1584	MER21C	238	909	ERV/ERV3	d
2200	2926	MER21C	105	911	ERV/ERV3	d
3518	4053	MER21C	316	903	ERV/ERV3	d
4833	4908	MER21C	320	402	ERV/ERV3	d
5241	5298	MER21C	115	174	ERV/ERV3	d
5410	5948	MER21C	317	909	ERV/ERV3	d
6522	7010	MER21C	280	815	ERV/ERV3	d
8437	8980	MER21C	317	909	ERV/ERV3	d
9587	10083	MER21B	171	715	ERV/ERV3	d
11498	12005	MER21C	317	848	ERV/ERV3	d
12033	12168	BOVA2	136	269	NonLTR/SINE	c
12438	12621	BovB	3664	3847	NonLTR/RTE	c
12965	13449	MER21B	171	715	ERV/ERV3	d
14282	14704	MER21C	316	786	ERV/ERV3	d
14783	14861	ERV1-2-I_BT	2917	2995	ERV/ERV1	c
15212	15289	CHR-2B	1	78	NonLTR/SINE	c

TKDP2

From	To	Name	From	To	Class	Dir
291	492	CHR-2B	3	245	NonLTR/SINE	c
674	1348	MER21C	115	848	ERV/ERV3	d
2191	2692	MER21C	316	864	ERV/ERV3	d
3371	3917	MER21C	306	877	ERV/ERV3	d
4698	4766	MER21C	318	387	ERV/ERV3	d
4789	4888	BOV2	461	559	NonLTR/SINE	c
4889	4943	BovB	3095	3149	NonLTR/RTE	c
5127	5328	BOVTA	1	205	NonLTR/SINE	d
5358	5629	MER21C	596	904	ERV/ERV3	d
6214	6536	MER21C	106	450	ERV/ERV3	d
6957	7276	L2	2713	3032	NonLTR/CR1	c
7353	7452	MER21B	207	308	ERV/ERV3	c
7594	7661	MER21C_BT	49	116	ERV/ERV3	c
8071	8208	CHR-2B	1	133	NonLTR/SINE	c

TKDP3

From	To	Name	From	To	Class	Dir
225	352	CHR-2B	174	315	NonLTR/SINE	c
359	482	CHR-2B	5	129	NonLTR/SINE	c
708	1433	MER21C	125	909	ERV/ERV3	d
1717	2016	BOV2	248	550	NonLTR/SINE	c
2390	2727	MER21C	103	473	ERV/ERV3	d
2788	3066	MER21B	466	763	ERV/ERV3	d
3805	4048	MER21C	285	542	ERV/ERV3	d
4168	4374	MER21C	596	825	ERV/ERV3	d
5213	5735	MER21C	316	863	ERV/ERV3	d
5947	6030	MIRc	62	162	NonLTR/SINE/SINE2	d
6352	6670	MER21C	106	445	ERV/ERV3	d
7083	7365	L2	2747	3027	NonLTR/CR1	c
7428	7539	MER21B	209	321	ERV/ERV3	c
7681	7772	MER21C_BT	23	116	ERV/ERV3	c
8081	8249	CHRL	1	168	NonLTR/SINE	c

TKDP4

From	To	Name	From	To	Class	Dir
291	497	CHR-2B	3	247	NonLTR/SINE	c
660	744	MER21C	1	81	ERV/ERV3	d
871	1411	MER21B	223	801	ERV/ERV3	d
1935	2061	BOVA2	143	269	NonLTR/SINE	d
2155	2654	MER21C	115	656	ERV/ERV3	d
2705	2858	MER21C	730	909	ERV/ERV3	d
3023	3099	MIRc	62	152	NonLTR/SINE/SINE2	d
3551	4046	MER21C	271	816	ERV/ERV3	d
4889	5420	MER21C	316	905	ERV/ERV3	d
5599	5670	MIRc	62	151	NonLTR/SINE/SINE2	d
5844	6041	Bov-tA1	4	211	NonLTR/SINE/SINE2	d
6105	6203	MER21C_BT	12	116	ERV/ERV3	c
6547	6682	CHR-2B	3	137	NonLTR/SINE	c
6772	6856	BOV2	468	557	NonLTR/SINE	d
7271	7339	LSU-rRNA_Hsa	2163	2234	Pseudogene/rRNA	c

TKDP5

From	To	Name	From	To	Class	Dir
268	476	CHR-2B	3	250	NonLTR/SINE	c
1169	1680	MER21C	316	864	ERV/ERV3	d
2142	2648	MER21B	213	755	ERV/ERV3	d
2785	2931	MER21B	213	355	ERV/ERV3	d
3318	3660	L2	2699	3046	NonLTR/CR1	c
3726	3832	MER21B	208	315	ERV/ERV3	c
3941	4069	MER21C	1	120	ERV/ERV3	c
4126	4394	BOVA2	1	269	NonLTR/SINE	d
4689	4823	CHR-2B	2	133	NonLTR/SINE	c
4843	5254	BOV2	150	560	NonLTR/SINE	c

Additional file 12 - Table S5. Tammar wallaby mammary gland cDNA microarray data presented in Figure 7

Global normalised Cy3/Cy5 gene expression data for (i) *EL* P (ii) *LGB* and (iii) *GAPDH* - throughout the lactation cycle from custom-made tammar mammary gland EST cDNA microarrays.
Microarray experiment data (E-MTAB-1057) was submitted to the EBI ArrayExpress Archive.
Phases include: the Virgin female mammary gland, Phase 1 (pregnancy), Phase 2A (early lactation), Phase 2B (mid-lactation) and Phase 3 (late lactation).
Clone identities (GenBank Accesssions) for *ELP* , *LGB* and *GAPDH* are provided below.

(i) *ELP* clones	GenBank	Virgin female mammary gland			Phase 1 - Pregnancy			Phase 2A - Early Lactation			Phase 2B - Mid-Lactation			Phase 3 - Late Lactation		
	Accession	a_V1	a_V2	a_V3	b_d005P	b_d022P	b_d025P	c_d001L	c_d005L	c_d080L	c_d130L	c_d168L	c_d180L	c_d213L	c_d220L	c_d260L
SGT20m3_B01	EX199339	0.810	0.590	0.710	0.770	0.830	0.640	3.440	3.570	4.370	2.490	0.580	1.000	0.580	1.480	1.040
SGT20p3_C08	EX200229	1.410	0.530	0.420	0.760	1.550	1.120	3.760	5.860	7.500	2.930	0.610	0.950	0.470	0.920	1.000
SGT20c1_G02	EX195759	1.350	0.550	0.480	0.900	0.980	0.760	2.390	2.780	4.880	2.080	0.600	1.240	0.590	0.970	0.980
Ave. expression		1.190	0.557	0.537	0.810	1.120	0.840	3.197	4.070	5.583	2.500	0.597	1.063	0.547	1.123	1.007
Std. Dev.		0.330	0.031	0.153	0.078	0.380	0.250	0.717	1.600	1.679	0.425	0.015	0.155	0.067	0.310	0.031
SEM		0.191	0.018	0.088	0.045	0.219	0.144	0.414	0.924	0.970	0.245	0.009	0.090	0.038	0.179	0.018

(ii) *LGB* clones	GenBank	Virgin female mammary gland			Phase 1 - Pregnancy			Phase 2A - Early Lactation			Phase 2B - Mid-Lactation			Phase 3 - Late Lactation		
	Accession	a_V1	a_V2	a_V3	b_d005P	b_d022P	b_d025P	c_d001L	c_d005L	c_d080L	c_d130L	c_d168L	c_d180L	c_d213L	c_d220L	c_d260L
SGT20j6_G09	EX198737	0.240	0.110	0.080	0.180	0.330	0.510	1.460	1.520	3.140	2.920	2.380	3.990	4.170	5.060	2.670
SGT20q3_B06	EX200470	0.200	0.140	0.130	0.200	0.240	0.360	1.640	1.750	2.370	1.980	1.980	4.480	4.830	6.290	3.450
SGT20f2_A11	EX196871	0.710	0.690	0.780	0.770	0.430	0.530	1.340	1.290	1.460	1.000	1.680	1.440	1.540	1.160	1.230
SGT20h1_B04	EX197428	0.760	0.520	0.480	0.660	0.260	0.510	1.550	1.480	1.480	0.980	2.000	1.860	2.030	1.280	1.330
SGT20i4_C12	EX198075	0.320	0.080	0.070	0.170	0.400	0.640	1.290	1.250	2.500	2.490	2.220	3.720	3.570	6.280	2.470
SGT20s2_G01	EX201427	0.830	0.700	0.690	0.790	0.380	0.610	1.570	1.520	1.290	0.870	1.680	1.520	1.430	1.190	1.120
SGT20n2_A09	EX199587	0.280	0.120	0.100	0.180	0.420	0.840	0.950	1.150	2.610	2.230	1.690	2.780	3.940	5.340	3.930
Ave. expression				0.382	0.421	0.351	0.571	1.400	1.423	2.121	1.781	1.947	2.827	3.073	3.800	2.314
Std. Dev.				0.285	0.301	0.077	0.148	0.235	0.205	0.710	0.829	0.281	1.255	1.380	2.464	1.127
SEM				0.062	0.114	0.029	0.056	0.089	0.077	0.268	0.313	0.106	0.474	0.522	0.931	0.426

(iii) *GAPDH* clones	GenBank	Virgin female mammary gland			Phase 1 - Pregnancy			Phase 2A - Early Lactation			Phase 2B - Mid-Lactation			Phase 3 - Late Lactation		
	Accession	a_V1	a_V2	a_V3	b_d005P	b_d022P	b_d025P	c_d001L	c_d005L	c_d080L	c_d130L	c_d168L	c_d180L	c_d213L	c_d220L	c_d260L
SGT20v2_A01	EX202667	1.010	1.120	1.000	0.920	1.060	1.080	0.940	0.990	0.970	0.990	1.030	0.930	0.930	1.020	0.960
SGT20j6_E09	EX198715	1.430	1.580	1.050	0.950	1.270	1.090	1.050	1.120	0.890	0.860	0.580	0.740	0.840	0.770	0.960
Ave. expression				1.198	0.935	1.165	1.085	0.995	1.055	0.930	0.925	0.805	0.835	0.885	0.895	0.960
Std. Dev.				0.246	0.021	0.148	0.007	0.078	0.092	0.057	0.092	0.318	0.134	0.064	0.177	0.000
SEM				0.100	0.015	0.105	0.005	0.055	0.065	0.040	0.065	0.225	0.095	0.045	0.125	0.000

Chapter Four:

A colostrum trypsin inhibitor gene expressed in the Cape fur seal mammary gland during lactation

Published in GENE: 27th Novemember 2015

Elizabeth A Pharo, Kylie N Cane, Julia McCoey, Ashley M Buckle, W H Oosthuizen, Christophe Guinet and John P Y Arnould (2016) A *colostrum trypsin inhibitor* gene expressed in the Cape fur seal mammary gland during lactation. *Gene*. **578**: 7-16.

Gene 578 (2016) 7–16

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

A *colostrum trypsin inhibitor* gene expressed in the Cape fur seal mammary gland during lactation

Elizabeth A. Pharo [a,b,*,1], Kylie N. Cane [a,b], Julia McCoey [c], Ashley M. Buckle [c], W.H. Oosthuizen [d], Christophe Guinet [e], John P.Y. Arnould [a,b,f]

[a] School of BioSciences, The University of Melbourne, Melbourne, VIC 3010, Australia
[b] Cooperative Research Centre for Innovative Dairy Products, Australia
[c] Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia
[d] Oceans and Coasts, Department of Environmental Affairs, Private Bag X2, Roggebaai 8012, South Africa
[e] Centre d'Etudes Biologiques de Chizé, CNRS, 79360 Villiers en Bois, France
[f] School of Life and Environmental Sciences, Deakin University, Burwood, VIC 3125, Australia

ARTICLE INFO

Article history:
Received 19 July 2015
Received in revised form 13 October 2015
Accepted 25 November 2015
Available online 27 November 2015

Keywords:
Pinniped
Kunitz
BPTI
Protease
Walrus
Milk

ABSTRACT

The *colostrum trypsin inhibitor* (*CTI*) gene and transcript were cloned from the Cape fur seal mammary gland and *CTI* identified by *in silico* analysis of the Pacific walrus and polar bear genomes (Order Carnivora), and in marine and terrestrial mammals of the Orders Cetartiodactyla (yak, whales, camel) and Perissodactyla (white rhinoceros). Unexpectedly, Weddell seal *CTI* was predicted to be a pseudogene. Cape fur seal *CTI* was expressed in the mammary gland of a pregnant multiparous seal, but not in a seal in its first pregnancy. While bovine *CTI* is expressed for 24–48 h postpartum (pp) and secreted in colostrum only, Cape fur seal *CTI* was detected for at least 2–3 months pp while the mother was suckling its young on-shore. Furthermore, *CTI* was expressed in the mammary gland of only one of the lactating seals that was foraging at-sea. The expression of *β-casein* (*CSN2*) and *β-lactoglobulin II* (*LGB2*), but not *CTI* in the second lactating seal foraging at-sea suggested that *CTI* may be intermittently expressed during lactation. Cape fur seal and walrus *CTI* encode putative small, secreted, N-glycosylated proteins with a single Kunitz/bovine pancreatic trypsin inhibitor (BPTI) domain indicative of serine protease inhibition. Mature Cape fur seal CTI shares 92% sequence identity with Pacific walrus CTI, but only 35% identity with BPTI. Structural homology modelling of Cape fur seal CTI and Pacific walrus trypsin based on the model of the second Kunitz domain of human tissue factor pathway inhibitor (TFPI) and porcine trypsin (Protein Data Bank: 1TFX) confirmed that CTI inhibits trypsin in a canonical fashion. Therefore, pinniped CTI may be critical for preventing the proteolytic degradation of immunoglobulins that are passively transferred from mother to young via colostrum and milk.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Colostrum trypsin inhibitor (*CTI*) is a mammary-specific gene which is expressed and the protein secreted in bovine colostrum for only 24–48 h postpartum (pp) (Laskowski and Laskowski, 1951; Pineiro et al., 1978; Veselsky et al., 1978). CTI is a small ~10–15 kDa, N-glycosylated protein (Klauser et al., 1978; Laskowski and Laskowski 1951; Tschesche et al., 1975) with a single Kunitz/bovine pancreatic trypsin inhibitor (BPTI) domain, characteristic to the family I2 (Kunitz-BPTI) inhibitors of the S1 (chymotrypsin) family of serine endopeptidases (Rawlings et al., 2014). Although CTI inhibits trypsin and plasmin and is a weak inhibitor of α-chymotrypsin *in vitro* (Feeney and Allison, 1969; Laskowski and Laskowski, 1951; Pineiro et al., 1978), neither its target enzyme, nor its function *in vivo* is known.

The expression of *CTI* and the orthologous marsupial *early lactation protein* (*ELP*) gene (Pharo et al., 2012), coincides with the passive transfer of antibodies from the mother to a neonate/young that lacks an

Abbreviations: AMBP, α1-microglobulin/bikunin precursor; APP, amyloid (βA4) precursor protein; CSN2, β-casein; BPTI, bovine pancreatic trypsin inhibitor; cDNA, DNA complementary to RNA; CTI, colostrum trypsin inhibitor; Da, Dalton; ELP, early lactation protein; h, hour; IgG, immunoglobulin G; KD, Kunitz domain; LGB2, β-lactoglobulin II; PIGT, phosphatidyl inositol glycan, class T; pp, postpartum; PRSS1, protease, serine, 1 (trypsin 1); SLPI, secretory leukocyte protease inhibitor; SPINLW1, serine peptidase inhibitor-like, with Kunitz and WAP domains 1; TFPI, tissue factor pathway inhibitor; TKDP, trophoblast Kunitz domain protein; UTR, untranslated region; WAP, whey acidic protein; WFIKKN, WAP, follistatin/kazal, immunoglobulin, Kunitz and netrin domain containing; WFDC2, WAP four disulphide core domain 2.

* Corresponding author at: School of BioSciences, The University of Melbourne, Melbourne, VIC 3010, Australia.
E-mail addresses: e.pharo@unimelb.edu.au (E.A. Pharo), kylie.cane@gmail.com (K.N. Cane), julia.mccoey@gmail.com (J. McCoey), ashley.buckle@monash.edu (A.M. Buckle), oosthuiz@environment.gov.za (W.H. Oosthuizen), guinet@cebc.cnrs.fr (C. Guinet), john.arnould@deakin.edu.au (J.P.Y. Arnould).
[1] Present address.

http://dx.doi.org/10.1016/j.gene.2015.11.042
0378-1119/© 2015 Elsevier B.V. All rights reserved.

adaptive (acquired) immune system and the ability to mount a specific immune response (Brambell 1970; Edwards et al., 2012). Furthermore, during this period, the gut of the young is permeable to intact immunoglobulins and macromolecules and thus these molecules can pass through the intestines and into the circulatory system prior to 'gut closure', i.e., when mucosal enterocytes lose the capacity to absorb macronutrients and immunoglobulins (Kruse, 1983; McFadden et al., 1997). *CTI* expression is brief in eutherians (1–2 days), but ELP expression is extended (for up to 100–125 days pp) in marsupials such as the possum and tammar wallaby (the common and scientific names for mammals described in this study are listed in Supplementary file 1) (Nicholas et al., 1997; Pharo et al., 2012; Piotte and Grigor 1996). Therefore, CTI may prevent the proteolytic degradation of immunoglobulins (Laskowski and Laskowski, 1951), while ELP may protect the marsupial young against pathogens (Piotte and Grigor, 1996).

Neither *CTI*, nor *ELP* is found in birds, fish, reptiles or amphibians and their status in monotremes is inconclusive. They therefore evolved from a common ancestral gene prior to the divergence of marsupials and eutherians (Pharo et al., 2012) ~160 million years ago (Luo et al., 2011). Intriguingly, all marsupials investigated have a functional (putative protein-coding) *ELP* gene, but this is not so for eutherian *CTI*. *CTI* is conserved in species from the orders Carnivora (dog, cat) and Cetartiodactyla (dolphin, cow, pig); but is a pseudogene in the horse (order Perissodactyla), humans and other primates, the elephant, sloth and rodents (Pharo et al., 2012). Gene loss, or loss of function has occurred many times throughout evolution and is often the result of gene duplication (Lynch and Conery, 2000) and/or transposition of genomic DNA fragments within the genome by retro-elements (Cañestro et al., 2013).

Since its evolution over 500 million years ago, the Kunitz domain (KD) has been duplicated many times (Gojobori and Ikeo, 1994; Ikeo et al., 1992) in bacteria, viruses, insects, invertebrates, vertebrates (e.g. fish, birds, amphibians, reptiles (e.g. venoms and dendrotoxins), mammals) and plants (Fry et al., 2009; Jamal et al., 2013; Rawlings et al., 2014). KDs have a diverse range of functions, e.g., serine protease inhibition, antimicrobial, anticoagulant and anti-inflammatory activity; potassium and calcium channel blockers (e.g. neurotoxic venoms), non-neurotoxic venoms, plant protection against herbivores, etc. (Fry et al., 2009; Ranasinghe and McManus, 2013; Shigetomi et al., 2010). Genes encoding as few as one KD, e.g. *BPTI* (also known as *PTI*) (Ascenzi et al., 2003); *serine protease inhibitor Kunitz-type* and *-4* (*SPINT3* and *-4*); *trophoblast Kunitz domain proteins 1, 2, 3, 4* and *5* (*TKDP1-5*); two domains: *SPINT1* and *-2*; three tandemly repeated domains: *tissue factor pathway inhibitor 1* and *-2* (*TFPI1 and -2*), and up to 12 domains, e.g. nematode *Ac*-KPI-1 (*Ancylostoma caninum*)*-Kunitz protease inhibitor-1* have been characterised (Hawdon et al., 2003). Multi-domain type-encoding genes have also been identified, e.g. *serine peptidase inhibitor-like, with Kunitz and WAP domains 1 (eppin) (SPINLW1)*; *WAP, follistatin/kazal, immunoglobulin, Kunitz and netrin domain containing-1* and *-2* (*WFIKKN1* and *-2*); *α1-microglobulin/bikunin precursor (AMBP)* and *amyloid (βA4) precursor protein* (*APP*).

The most extensively structurally characterised KD is BPTI (Ascenzi et al., 2003), which most-likely evolved from bovine *CTI* after the divergence of ruminants from other Cetartiodactyla (Pharo et al., 2012) ~25–35 Myr (Bininda-Emonds et al., 2007). The KD belongs to the α + β fold and is stabilised by three disulphide bonds (Ascenzi et al., 2003; Huber et al., 1972). The P_1 'warhead' residue within the convex exposed 'binding' loop of the inhibitor determines serine protease specificity (Laskowski and Kato, 1980; Laskowski and Qasim, 2000). Kunitz inhibitors with a basic Lys or Arg residue at P_1 inhibit trypsin and trypsin-like enzymes (Ikeo et al., 1992; Laskowski and Kato, 1980), those with Ala or Ser inhibit elastase-like enzymes and a P_1 Met or Leu confers activity against chymotrypsin and chymotrypsin-like enzymes (Laskowski and Kato, 1980). A P_1 Leu is also active against neutrophil elastase (Garcia-Fernandez et al., 2015). Protease inhibition involves the cleavage of the P_1–P_1' peptide bond (inhibitor) and the docking of the convex inhibitor 'binding' loop into the catalytic cleft of the serine protease in a classical 'lock and key' interaction (Marquart et al., 1983; Grzesiak et al., 2000). A new, 1:1, reversible, tight-binding serine protease-Kunitz inhibitor complex is formed, involving P_1 (inhibitor) and the catalytic triad of residues: Ser, His and Asp (protease) (Marquart et al., 1983).

The identification of *CTI* in the dolphin (Cetartiodactyla) (Pharo et al., 2012), suggested that the gene may also be present in other marine mammals, such as the Pinnipedia, 'fin-footed' semi-aquatic mammals (Carnivora). Although monophyletic, the pinnipeds comprise a diverse range of species within three extant families: the Otariidae (eared seals: fur seals, e.g. Cape and subantarctic fur seals; and sea lions, e.g. California seal lion), its sister family, the Odobenidae, with one extant member, the walrus; and the Phocidae (earless or 'true' seals: e.g. Weddell seal, Hooded seal), [Fig. 1; (Higdon et al., 2007; Nyakatura and Bininda-Emonds 2012)]. While all pinnipeds give birth to a pup on land or ice (Atkinson 1997; Oftedal et al., 1987) and produce lipid-rich, high-protein and low-sugar milk for their young, each family has developed unique lactation strategies to ensure the survival of their young in marine environs (Fig. 1). Otariids such as the Cape and subantarctic fur seals fast for the first 4–10 days pp while nursing their pup on land (Bonner, 1984; Oftedal et al., 1987). The mother then commences a pattern of alternating trips to sea to feed interspersed with ~2–3 days on-shore suckling her pup (Bonner, 1984; Gentry and Kooyman, 1986; Oftedal et al., 1987). Each trip to sea usually lasts for up to 25 days (Gamel et al., 2005) and ~23–28 days (Kirkman et al., 2002) for the Cape and subantarctic fur seals respectively, but extreme foraging periods of up to 1–2 months have been reported (Georges and Guinet, 2000; Verrier et al., 2011). Therefore, it is vital that the mother produces milk that will ensure the survival of her pup during these periods.

The aims of this study were to identify the *CTI* gene and characterise its expression in the mammary glands of semi-aquatic members of the order Carnivora (Cape and subantarctic fur seals), identify *CTI* in other species using an *in silico* bioinformatics approach, and to use protein structural homology modelling to investigate whether pinniped CTI, like other KDs has the potential to inhibit trypsin.

2. Materials and methods

2.1. Animals

Mammary gland tissue was collected from six Cape fur seals and one subantarctic fur seal during the reproductive cycle (Table 1), as described previously (Cane, 2005). Approval for this research was obtained from the South African Government (Cape fur seals) and the Ethics Committee of the French Polar Institute (IPEV) and the Polar Environment Committee of Terres Australes et Antarctiques Françaises (subantarctic fur seal).

2.2. Isolation of Cape fur seal genomic DNA and cloning of the CTI gene

Genomic DNA was isolated from the mammary gland of an on-shore lactating Cape fur seal as described (Sambrook and Russell, 2001) and *CTI* amplified by PCR from ~200 ng of genomic DNA with forward 5′-GCCTAGAACATTCAGCTATTGGCACC-3′ and reverse 5′-TGAATGTTTTAT TGACCTAGACCTGGAGG-3′ primers and Platinum *Taq* (Invitrogen) as per the manufacturer's instructions. PCR conditions used were 94 °C for 2 min; 35 cycles of [94 °C for 30 s; 54 °C for 30 s; 68 °C for 4 min] and a final extension at 68 °C for 10 min. The PCR product was cloned into pGEM-T Easy (Promega Corporation) and sequenced.

2.3. In silico identification of CTI

CTI genes were compiled from BLAST searches of the NCBI GenBank nr (http://www.ncbi.nlm.nih.gov/BLAST/) and Ensembl

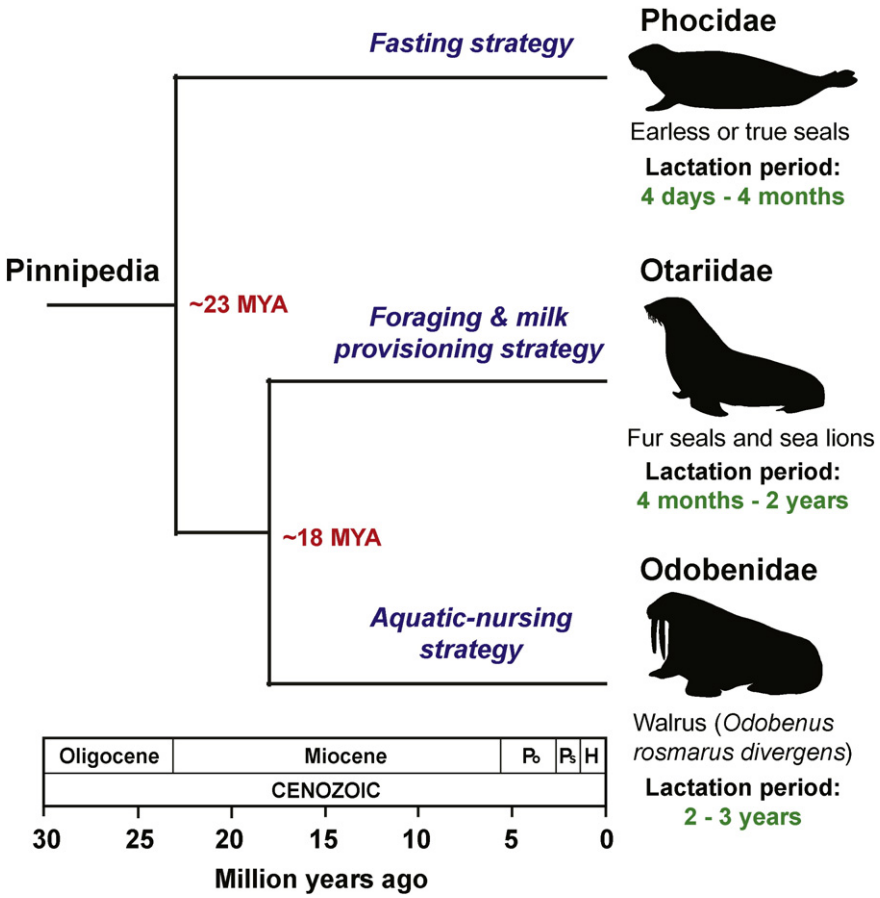


Fig. 1. Evolutionary history of the Pinnipedia. The Pinnipedia consists of three extant families with different reproductive strategies. The Phocidae diverged from the ancestor of the Otariidae and Odobenidae families ~33–23.3 Myr ago and these latter two families then diverged ~26–11 Myr ago (Higdon et al., 2007; Nyakatura and Bininda-Emonds, 2012). The divergence times shown are based on the study by Higdon and colleagues (Higdon et al., 2007). Po, Pliocene; Ps, Pleistocene; H, Holocene.

Release 80, May 2015 (http://www.ensembl.org/) databases or BLAT searches of the UCSC Genome Browser (http://genome.ucsc.edu/) using *CTI* mRNA and protein sequences. Expect-values ≤ 1e − 8 and E-value ≤ 1e − 17 were used as cut-offs for nucleotide and protein comparisons respectively. Putative exons, transcripts and proteins within genomic sequences were predicted using the web-based tools, GENSCAN (http://genes.mit.edu/GENSCAN.html) and Spidey (mRNA-to-genomic alignment) (http://www.ncbi.nlm.nih.gov/spidey/).

2.4. Preparation of total RNA

Total RNA for Northern analysis was isolated from pregnant and lactating Cape fur seal mammary gland tissue using the RNeasy Lipid Tissue Mini kit (Qiagen, Hilden, Germany) as per the manufacturer's instructions.

2.5. RT-PCR cloning of the fur seal CTI transcript

First strand cDNA was generated with Superscript II and total RNA (5 μg) isolated from mammary tissue of an on-shore, lactating Cape fur seal (O3). The Cape fur seal *CTI* transcript was amplified by PCR using the same primers as for the *CTI* gene (Section 2.2), high-fidelity proof-reading Platinum *Taq* (Invitrogen, Carlsbad, CA, USA) and 5% of the first strand reaction as a template. PCR conditions used were 94 °C for 2 min; 35 cycles of [94 °C for 30 s; 55 °C for 30 s; 68 °C for 1 min] and a final extension at 68 °C for 10 min and the PCR product cloned

Table 1
Reproductive stages of Cape and subantarctic fur seals used for this study.

Seal species	Reproductive stage	Label	Location
Cape fur seal	Late-pregnancy, primiparous (first pregnancy) female	P1	Small breeding colony at Robesteen, South Africa (33°37′60S, 18°23′60E)
Cape fur seal	Late-pregnancy, multiparous female	P2	Small breeding colony at Robesteen, South Africa (33°37′′60S, 18°23′′60E)
Cape fur seal	Early-lactation female (on-shore) ~2–3 months pp	O1	Large breeding colony in Kleinsee, South Africa (29°40′0S, 17°4′60E), near the Namibian border
Cape fur seal	Early-lactation female (on-shore) ~2–3 months pp	O2	Large breeding colony in Kleinsee, South Africa (29°40′′0S, 17°4′′60E), near the Namibian border
Subantarctic fur seal	Lactating female (on-shore) ~2 months pp	O*At*	La Mare aux Elephants (46°22′S, 51°40′E), Possession Island (Iles Crozet), southern Indian Ocean
Cape fur seal	Lactating female (at-sea)	S1	West of Cape Town, South Africa, past the continental shelf edge ~100 km from the nearest seal colony
Cape fur seal	Lactating female (at-sea)	S2	West of Cape Town, South Africa, past the continental shelf edge ~ 100 km from the nearest seal colony

into pGEM-T Easy (Promega Corporation) and sequenced using SP6 and/or T7 primers.

2.6. Northern blot analyses

The membrane used to evaluate *CTI* expression was prepared as previously described (Cane, 2005). Briefly, total RNA (10 μg) was electrophoresed at ≥5 V/cm through a 1% agarose, low-formaldehyde (1.1%) gel with 1× MOPS [3(N-Morpholino) Propane Sulfonic Acid] buffer at 4 °C. The RNA was transferred to Zeta-Probe GT Blotting Membrane (Bio-Rad) in 20× SSC (3.0 M sodium chloride, 0.3 M trisodium citrate, pH 7.0) overnight. The blot was rinsed briefly in 2× SSC, cross linked with UV light (120 mJ) and hybridised for 4 h at 42 °C in 25 mL [30% deionised formamide, 5× SSC, 50 mM sodium acetate, herring sperm DNA (100 μg/μL), 5 mL Denhart's 50× stock solution, 0.1% SDS]. The membrane was probed overnight at 42 °C with [α-^{32}P] dCTP-labelled fur seal *CTI* cDNA using the DECAprime II Random Priming DNA Labelling Kit (Ambion, Invitrogen, Carlsbad, CA, USA).

2.7. Sequence analysis

Physical characteristics of CTI were predicted using the web-based tools ProtParam (http://www.expasy.org/tools/protparam.html) and Pepstats (http://www.ebi.ac.uk/Tools/emboss/pepinfo/). Conserved protein motifs and post-translational modifications were predicted using the PROSITE database of protein domains, families and functional sites (http://prosite.expasy.org/) and the Center for Biological Sequence Analysis prediction servers: NetNGlyc 1.0 (N-glycosylation sites based on the Asn-Xaa-Ser/Thr sequon) (http://www.cbs.dtu.dk/services/NetNGlyc/) and NetPhos 2.0 (serine, threonine and tyrosine phosphorylation) (http://www.cbs.dtu.dk/services/NetPhos/). Signal peptides (indicative of secreted proteins) were predicted with SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP/). Sequences were aligned with Clustal W2 or Clustal Omega available at the EMBL-EBI website (http://www.ebi.ac.uk/services). BoxShade Server version 3.2 (http://www.ch.embnet.org/software/BOX_doc.html) was used to shade multiple sequence alignments. Similarity and identity between either nucleotide, or amino acid sequences were determined with MatGAT (Matrix Global Alignment Tool) 2.03 software (Campanella et al., 2003) using the BLOSUM50 matrix. MatGAT produces pairwise alignments and homology between each sequence pair. Genomic DNA and mRNA sequences and third party annotations were submitted to the GenBank nr database.

2.8. Structural homology modelling

Homology modelling was performed using Chimera (Pettersen et al., 2004) and Modeller (Šali and Blundell, 1993). BLAST searches of the NCBI GenBank nr (http://www.ncbi.nlm.nih.gov/BLAST/) and RSC PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank), (http://www.rcsb.org/pdb/home/home.do) were used to identify trypsinogen sequences and the latter to find the best KD 3D structure to model CTI, as well as the best model for the CTI-trypsin interaction. Models of Cape fur seal CTI [GenBank: AFZ99004; residues 33–94 inclusive], Pacific walrus trypsin [GenBank: XP_004416603; residues 24–246 inclusive] and a Cape fur seal CTI-Pacific walrus trypsin complex were constructed using the X-ray crystal structure of the second KD of tissue factor pathway inhibitor [GenBank: NP_006278; residues 121–178 inclusive] in complex with porcine trypsin [GenBank: NP_001156363; residues 24–246 inclusive], PDB ID: 1TFX. Pymol Version 1.3r1 (http://www.pymol.org/releases/v1.3r1) was used for model visualisation and preparation of Fig. 5.

3. Results

3.1. The Cape fur seal and Pacific walrus have a putative protein-coding CTI gene, but Weddell seal CTI encodes a pseudogene

To investigate whether pinnipeds have a functional *CTI* gene, we cloned the 2679 bp *CTI* gene from the Cape fur seal mammary gland [GenBank: KC152480] and identified the 2605 bp Pacific walrus *CTI* by BLAST searches of the walrus genome (Oros_1.0). Unexpectedly, Weddell seal *CTI* (LepWed1.0) was predicted to be a pseudogene, which if transcribed and translated would produce a ~6.7 kDa secreted protein lacking a KD (Supplementary file 2). Functional *CTI* genes were also predicted in the polar bear [GenBank: BK009384] (order Carnivora) and in marine mammals from the order Cetartiodactyla: killer whale, Scammon's minke whale and terrestrial species: sheep, goat, alpaca, wild yak, Tibetan antelope, Bactrian and Arabian camels and bison (Accession numbers are listed in Supplementary file 3). While *CTI* was thought to be a pseudogene in Perissodactyla, e.g. thoroughbred horse (Pharo et al., 2012) and Przewalski's horse (this study), another member of this order, the white rhinoceros, has a functional *CTI* gene (this study).

Consistent with other *CTI* genes (Pharo et al., 2012), pinniped *CTI* included three exons and two introns (Fig. 2; Supplementary file 4). However, pinniped *CTI* exon 1 was 12 nt shorter than that of other eutherians and encoded the signal peptide only. *CTI* gene identify was confirmed where possible, by the single-copy flanking genes, *phosphatidyl inositol glycan, class T* (*TPIG*) and *WAP four disulphide core domain 2* (*WFDC2*). A summary of *CTI* and *ELP* gene and *CTI* pseudogene distribution in mammals is provided (Supplementary file 5). An alignment of *CTI* and *ELP* transcripts is also provided (Supplementary file 6), as well as their evolutionary history (Supplementary file 7). Notably, *CTI* and *ELP* phylogeny was consistent with that of mammals.

3.2. Cape and subantarctic fur seal CTI are expressed in the mammary gland for at least 2–3 months postpartum

In order to investigate whether *CTI* is expressed in otariids, we amplified a 449 bp *CTI* transcript from the mammary gland of a pregnant Cape fur seal (P2) and used Northern analysis to characterise *CTI* expression in the mammary glands of pregnant, lactating at-sea, and lactating on-shore Cape fur seals, and from an on-shore lactating subantarctic fur seal (Fig. 3; Supplementary file 8). The expression of *LGB2* and *CSN2*, both of which encode major milk proteins of the whey and casein milk fractions respectively, were used as markers for prolactin-responsive whey protein gene expression and functional differentiation of the mammary gland (Rijnkels et al., 2013) respectively. *CTI* was detected in the mammary gland of the multiparous (P2), but not the primiparous (P1), pregnant seal, similar to the pattern of *CSN2* expression, while *LGB2* (Cane 2005; Sharp et al., 2005; Sharp et al., 2006a; Sharp et al., 2006b) was detected in both pregnant seals. *CTI*, *LGB2* and *CSN2* were expressed in the mammary glands of the lactating on-shore Cape (O3, O4) and subantarctic (O*At*) fur seals, which were nursing pups that were at least 2–3 months old. However, *CTI* transcripts were only detected in mammary tissue from one of the lactating females foraging at-sea. In contrast, *LGB2* and *CSN2* transcripts were detected in both lactating females at-sea.

3.3. Pinniped CTI, putative N-glycosylated Kunitz serine protease inhibitors

Next, we translated the pinniped *CTI* transcripts and determined that Cape fur seal and walrus *CTI* encode a putative 21 aa signal peptide and a 75 aa mature protein of ~8.47 and 8.59 kDa respectively, similar to other CTI and ELP peptides, 8.56–9.68 kDa (Supplementary file 9). As expected, mature Cape fur seal CTI was most similar to walrus CTI (93.3%), but less similar to Carnivora (78.5–83.5%), rhinoceros

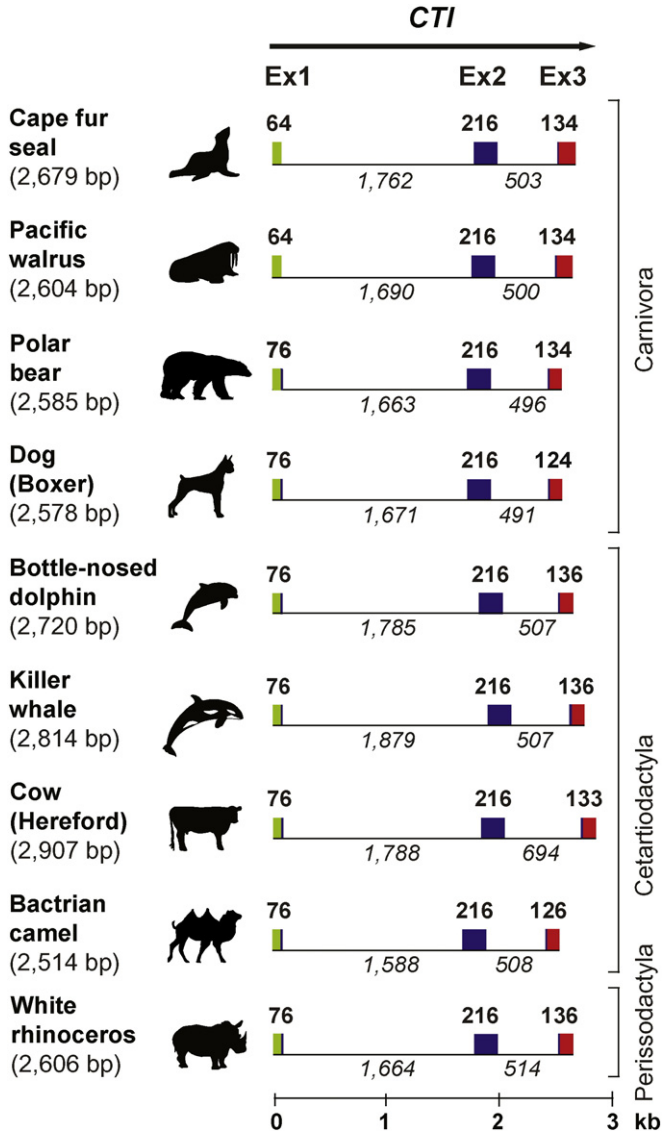


Fig. 2. Structure of pinniped *CTI* and other eutherian *CTI* genes. The conserved 3-exon structure of selected *CTI* genes is shown. Cape fur seal *CTI* [GenBank: KC152480], Pacific walrus [GenBank: LOC101362989], Polar bear [GenBank: BK009384; LOC103669871], dog [GenBank: BK008082], bottle-nosed dolphin [GenBank: BK008085; LOC101340157], killer whale [GenBank: LOC101273120], cow [GenBank: LOC104973896], Bactrian camel [GenBank: LOC105087824], white rhinoceros [GenBank: LOC101393639]. The *CTI* transcript encodes a signal peptide (green rectangle) and a mature secreted protein (blue), with a short non-coding 3′ UTR (red). Exon 2 which encodes the Kunitz domain is 216 bp. Exon size is shown in bold text and intron sizes are italicised. Gene size (within brackets) represents the number of nucleotides from the putative translation start (ATG, exon 1) to the polyadenylation signal (AATAAA, exon 3) inclusive.

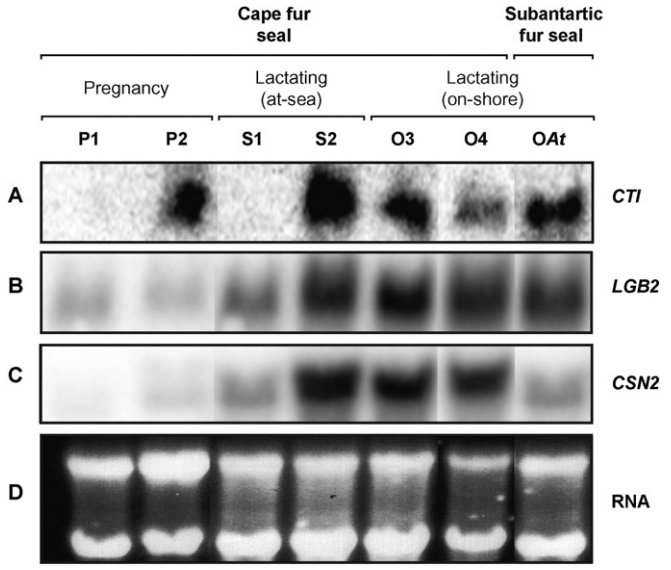


Fig. 3. Expression of *CTI* in the fur seal mammary gland. Northern analysis of total RNA (10 μg) extracted from the mammary glands of the Cape fur seal during late-pregnancy (~1 month pre-partum) and Cape and subantarctic fur seals during early-mid lactation (~2–3 months pp). A. *CTI* was not detected in the seal in its first pregnancy (P1), nor in one of the lactating seals foraging at-sea (S1). In contrast, *CTI* was expressed in the mammary glands of the multiparous seal during pregnancy (P2), in one of the lactating at-sea females (S2) and in the on-shore lactating Cape and subantarctic fur seals ~2–3 months pp (O3, O4, OA*t*). B. *LGB2* was detected at all reproductive stages, with expression highest in the mammary glands of the on-shore lactating seals. C. *CSN2*, indicative of mammary gland differentiation, was detected in all animals except the primarous seal, with strong expression in the on-shore lactating Cape fur seals and one of the lactating seals at-sea, D. Ethidium bromide-stained ribosomal RNA bands show total RNA integrity and loading.

(Perissodactyla; 73.5%) and Cetartiodactyla CTI (67.5–71.1%), and marsupial ELP (49.4–56.8%) (Supplementary file 10).

Alignment of the pinniped CTI precursor proteins with CTI of selected species, plus the ubiquitously expressed BPTI (shown as cow PTI, Fig 4) and human TFPI KD2, upon which Cape fur seal CTI was modelled (Section 3.4), highlighted regions of similarity and difference (Fig 4). An alignment of all CTI and ELP precursor proteins is also provided (Supplementary file 11). Characteristic to the Kunitz/BPTI domain [BPTI KUNITZ 2 motif, Prosite: PS50279 and BPTI KUNITZ 1 motif: Prosite: PS00280, F-x(2)-{I}-G-C-x(6)-[FY]-x(5)-C; residues within square brackets are permitted and those within curly brackets are not], pinniped CTI shares six conserved cysteine residues which form three disulphide bonds: C1–C6, C2–C4 and C3–C5, corresponding to Cys19–Cys69, Cys28–Cys52, Cys44–Cys65 (pinniped protein numbering, Fig. 4; Supplementary files 11 and 12). The 51 aa CTI KD is flanked by non-structured, low complexity regions of 18–22 aa and 3–10 aa at the N- and C-termini respectively.

The putative interaction of Cape fur seal CTI with a serine protease, e.g. trypsin, is most likely mediated by two protease recognition loops (blue box, 29–37; green box, 52–59; alignment numbering Fig. 4), which are stabilised by the C2-C4 disulphide bond (Xu et al., 1998). Like BPTI (Hanson et al., 2003), CTI is predicted to be stabilised by the internal hydrophobic core residues: Tyr39–Phe40–Tyr41, although the camel has Ser40 (Fig. 4). Notably, a basic P_1 reactive site residue (Arg/Lys) was conserved in all sequences (#, residue 33, Fig 4, Supplementary file 11). Although Carnivora CTI and human TFPI KD2 P_1–P_1' (Arg–Gly) and Cetartiodactyla CTI and bovine PTI P_1–P_1' (Lys–Ala) were conserved, the rhinoceros CTI P_1–P_1' (Arg-Gly) were a combination of the two.

In contrast to the non-glycosylated BPTI (Creighton and Charles, 1987), CTI has a conserved putative N-glycosylation site at Asn42 (alignment numbering Fig. 4, Supplementary files 11 and 12) [ASN glycosylation motif, Prosite: PS00001, N-{P}-[ST]-{P}, where N is the sugar attachment site]. The N-glycosylation site has been confirmed for bovine CTI and is located at the 'base' of the pear-shaped protein (Klauser et al., 1978). TFPI KD2 is also N-glycosylated, but at Asn117 (Nakahara et al., 1996) (position 43 alignment numbering, Fig. 4), adjacent to the CTI sugar-attachment site.

All CTI peptides were predicted to be phosphorylated at Ser43 (ϕ, alignment numbering Fig. 4, Supplementary files 11 and 12) and all Carnivora CTI peptides phosphorylated by casein kinase II at Thr44 (alignment numbering Fig. 4, Supplementary files 11 and 12), [CK2 PHOSPHO SITE, Prosite: PS00006, [ST]-x(2)-[DE], residues 44–47, phosphorylation at position 1 (Thr44)].

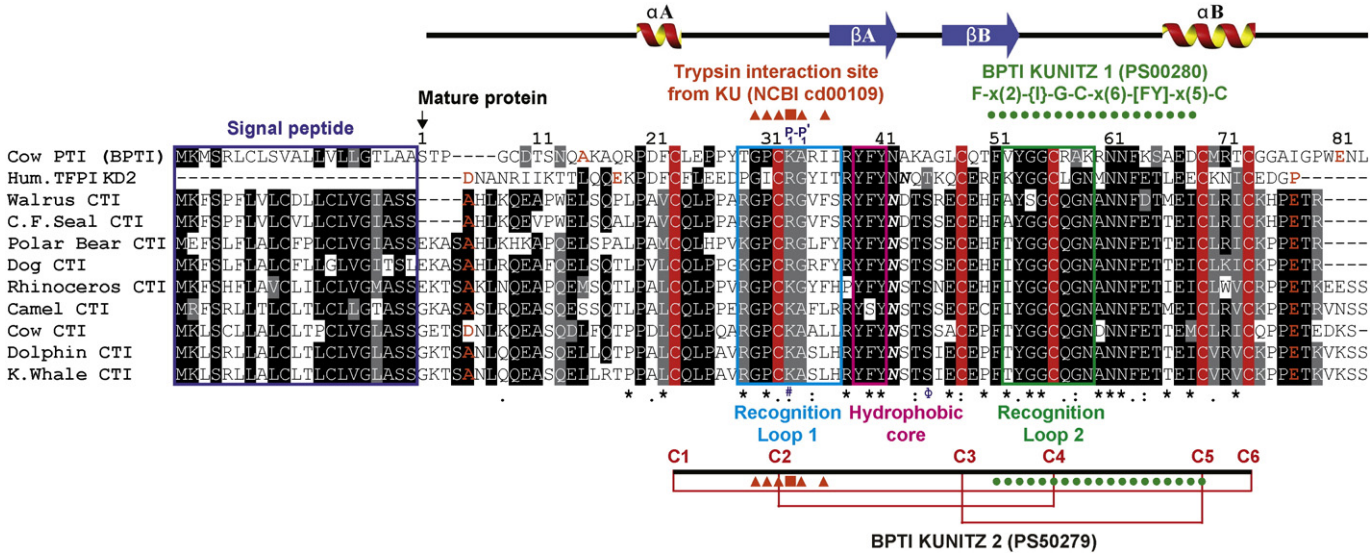


Fig. 4. Alignment of CTI, BPTI and the second Kunitz domain of human TFPI. Putative CTI and bovine PTI (BPTI) precursor proteins and KD2 of human TFPI were aligned with Clustal Omega. BPTI (cow PTI) [GenBank: AAI49369], human TFPI KD2 (includes part of the flanking regions, amino acids 79–150 inclusive shown, mature protein numbering) [GenBank: NP_006278] and CTI of the Pacific walrus [GenBank: XP_012422725], Cape fur seal CTI [GenBank: AFZ99004], Polar bear [GenBank: BK009384], dog [GenBank: AEN62470], white rhinoceros [GenBank: XP_004430613], Bactrian camel [GenBank: XP_006193597]; cow [GenBank: AEN62469], bottle-nosed dolphin [GenBank: BK008085; DAA35188] and killer whale [GenBank: XP_004272954]. Predicted signal peptides (dark blue border) and residues that span splice sites (orange text) are shown. The P_1 reactive site residue [#, the putative trypsin interaction site from the KU motif (NCBI cd00109)] and the BPTI_KUNITZ_1 and 2 motifs and are indicated. The two putative serine protease recognition loops (blue and green boxes respectively), the predicted hydrophobic core (pink box), N-glycosylation sites (bold, italicised text) and conserved serine phosphorylation site (ϕ) are shown. The predicted 3D-structure of CTI is indicated above the alignment: coiled regions, black bar; α-helices, red and yellow; and β-sheets, blue arrows. Conserved sites were shaded with BoxShade. Residues common to all species are indicated (*) and those shared by at least 19 of 27 species (>70%) are shaded black. Grey residues indicate conserved residue type. Conservation between groups of amino acids with strongly similar properties, i.e., score > 0.5 in the Gonnet PAM 250 matrix is indicated (:). Conservation between groups of amino acids with weakly similar properties (score < 0.5 in the Gonnet PAM 250 matrix) is also noted (.). Gaps within the alignment are indicated (−).

3.4. Structural homology modelling of Cape fur seal CTI suggests that it inhibits trypsin

In order to further investigate the protease inhibitory potential of Cape fur seal CTI, we constructed a homology model. High (54%) sequence identity with the second KD of human TFPI, for which an X-ray crystal structure exists [PDB ID: 1TFX, Chain D (Burgering et al., 1997)], predicted a high degree of structural conservation throughout the molecule and, in particular, in the protease-recognition loops (Fig. 5A). Apart from 3-residue and 2-residue insertions at the N- and C-termini respectively (both located distal to the inhibitory loop), there are no insertions or deletions.

We next constructed a model of the complex between Cape fur seal CTI and Pacific walrus trypsin, using the X-ray crystal structure of the second KD of human tissue factor pathway inhibitor in complex with porcine trypsin [PDB ID: 1TFX (Burgering et al., 1997)]. Neither Cape fur seal, nor any other otariid trypsin sequence was available and so we used the trypsin sequence of its closest relative, the walrus. As these species are from sister families, we expected minimal sequence difference. Notably, the active trypsin enzymes of the walrus (odobenid) and Weddell seal (phocid) share 89.7% identity, but less than with other Carnivora peptides (79.8–89.2%), Cetartiodactyla trypsin (69.5–77.1%), and human trypsin (68.3–71.4%) (Supplementary file 13).

Alignment of pinniped trypsinogen with the bovine, human, dog, polar bear and killer whale sequences shows the trypsin motif [TRYPSIN_DOM; Prosite: PS50240 (residues 9–229, alignment numbering; Fig. 6)], conservation of eight cysteine residues which form four disulphide bonds and stabilise the enzyme, as well as the conservation of the catalytic triad of residues (Ser185, His48 and Asp92; nucleophile–base–acid) which are necessary for protease-inhibitor binding (Vandermarliere et al., 2013). Like porcine trypsin (PDB ID: 1TFX, Chains A, B) and bovine trypsin (PDB ID: 4GUX, chain A), walrus trypsin is predicted to form two six-stranded beta-barrels (α-trypsin chain 1 and α-trypsin chain 2) separated by an interface comprising the catalytic residues (Vandermarliere et al., 2013).

Consistent with Kunitz inhibitor-serine protease complex formation (Burgering et al., 1997; Vandermarliere et al., 2013), homology modelling suggests that Cape fur seal CTI possesses the necessary structural and chemical features in order to inhibit trypsin (Fig. 5B). Strikingly, there were very few residue differences between the homology inhibitor-trypsin templates and the seal/walrus proteins in the vicinity of the protease-inhibitor interface, with no actual differences in the interface itself (Fig. 5B). The resulting features of the modelled walrus trypsin–Cape fur seal CTI complex, such as interface hydrogen bonding (Fig. 5C) and structural complementarity (Fig. 5D) were highly conserved with the X-ray crystallographically-determined mammalian complex (PDB ID: 1TFX). Interactions between the inhibitor P_1 residue (Arg in this case) and the protease S_1 specificity pocket typically make dominant energetic contributions to the complex stability and thus the inhibitory activity. Taken together, the modelling strongly suggests that seal CTI most likely possesses inhibitory activity against trypsin-like serine proteases in a highly similar manner to canonical inhibitors such as BPTI.

4. Discussion

A functional *colostrum trypsin inhibitor* gene was thought to be restricted to all species of the eutherian orders Carnivora and Cetartiodactyla (Pharo et al., 2012), but we have shown that although *CTI* is present in semi-aquatic marine mammals such as the Cape fur seal (Otariidae) and walrus (Odobenidae), it is a pseudogene in the Weddell seal (Phocidae), the only marine mammal with a disrupted *CTI* gene thus far. Furthermore, the identification of a putative protein-coding *CTI* gene in the white rhinoceros (family Rhinocerotidae, order Perissodactyla), suggests that *CTI* may also be present in other Perissodactyls, e.g. tapirs (family Tapiridae), with a disrupted *CTI* gene limited to members of the Equidae (this study; Pharo et al., 2012). The

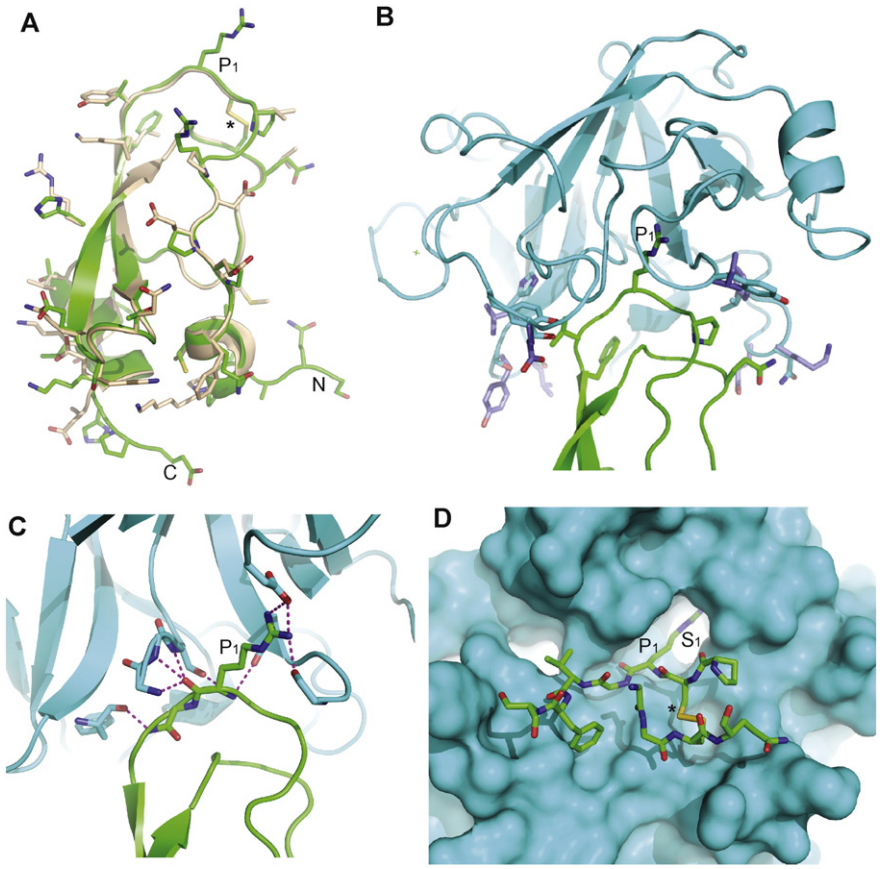


Fig. 5. Panel showing salient features of the modelled Cape fur seal CTI (green) and its complex with Pacific walrus trypsin (cyan). Orientations of the protease inhibitor are kept approximately the same throughout. A. Cartoon of Cape fur seal CTI superimposed on the second Kunitz domain of human tissue factor pathway inhibitor I (PDB ID: 1TFX chain D, wheat) on which it was modelled. The side chains of residues that differ in the sequence alignment are shown as sticks. N- and C-termini, showing 3- and 2-residue insertions, respectively, in Cape fur seal CTI are indicated; B. cartoon of model of Cape fur seal CTI (green) in complex with walrus trypsin (cyan). Interface residues that differ between the models and the homologues are shown as sticks (cyan = porcine trypsin, blue = walrus trypsin); C. Close-up cartoon of the walrus trypsin-Cape fur seal CTI complex showing hydrogen bonds (magenta) at the complex interface. A cross-section is shown where some residues have been removed for clarity; D. walrus trypsin-Cape fur seal CTI complex model showing trypsin molecular surface (cyan) and interacting residues of CTI (sticks). P_1 arginine of CTI fits snugly into the trypsin S_1 specificity pocket (both labelled). Panels B, C and D show the high degree of chemical and structural complementarity at the complex interface consistent with the proposed trypsin-inhibitory function of Cape fur seal CTI. The conserved disulphide bond in the protease inhibitor that anchors the inhibitory loop is shown as sticks (yellow S atoms) and marked with an asterisk. The P_1 residue is labelled in all panels.

prolonged period of *CTI* expression in the mammary gland of the Cape and subantarcic fur seals suggests that CTI is essential in these otariids. Additionally, the conservation of a basic P_1 residue (Lys/Arg) in all eutherian CTI peptides, unlike in present-day marsupial ELP, suggests that eutherian CTI acquired a basic P_1 residue after the divergence of the ancestral marsupials and eutherians ~160 million years ago (Luo et al., 2011). Structural homology modelling suggests that pinniped CTI binds to and inhibits trypsin in a canonical fashion, which may play an important role in neonate health.

4.1. Conservation of pinniped CTI gene structure

While the flanking genes of Cape fur seal *CTI* are unknown, the greater than 93% nucleotide identity shared been the Cape fur seal and walrus *CTI* genes, the latter of which is flanked by the single-copy *PIGT* and *WFDC2* genes, confirmed the identity of Cape fur seal *CTI*. The conservation of the 3-exon, 2-intron *CTI* structure and in particular, the 216 bp exon 2, which encodes a low-complexity 18 aa coiled region at its 5′ end (CTI N-terminus), a 51 aa Kunitz domain and a short 4 aa 3′ end suggests that conservation of exon 2 in its entirety may be important for CTI function. Furthermore, the eutherian Order-specific conservation of the 3–7 aa encoding CTI C-terminus suggests that this region may have an Order/species-specific function. The loss of *CTI* in the Weddell seal and other eutherians implies that either CTI function is provided by an alternate protein, or that CTI is not essential in these species and hence there is no evolutionary disadvantage to its loss.

4.2. The unique expression pattern of Cape fur seal CTI

The expression of *CTI* in the mammary glands of the on-shore lactating Cape and subantarctic fur seals 2–3 months pp, but the lack of *CTI* transcripts in one of the two lactating Cape fur seals foraging at-sea is consistent with reduced milk synthesis at-sea (>80%) in Antarctic (Arnould and Boyd, 1995) and Cape fur seals; and the resumption of milk production as the mother returns to shore to suckle her pup (Cane 2005; Sharp et al., 2005; Sharp et al., 2006a). Both the absence of lactose (the major osmole in most mammalian milks which directly affects milk volume) (Brew 2003; Neville et al., 1983) and a 'switch' in energy partitioning in the otariid mother, from the mobilisation of her body reserves for milk production on-shore, to replenishing her bodily stores while foraging at-sea (Costa 1991) may facilitate the 'suspension' of lactation at-sea without the gland involuting (Cane 2005; Sharp et al., 2005; Sharp et al., 2006a). In contrast, the cessation of sucking stimulus for ~48–72 h triggers mammary gland involution in most mammals (Watson 2006). Furthermore, the detection of *CSN2* and *LGB2* transcripts, but not *CTI* in one of the at-sea females suggests that *CTI* expression may be intermittent during lactation, i.e., up-regulated on-shore and down-regulated at-sea. Unlike other eutherians, otariid *CTI* may be expressed throughout lactation, rather than for only the first 24–

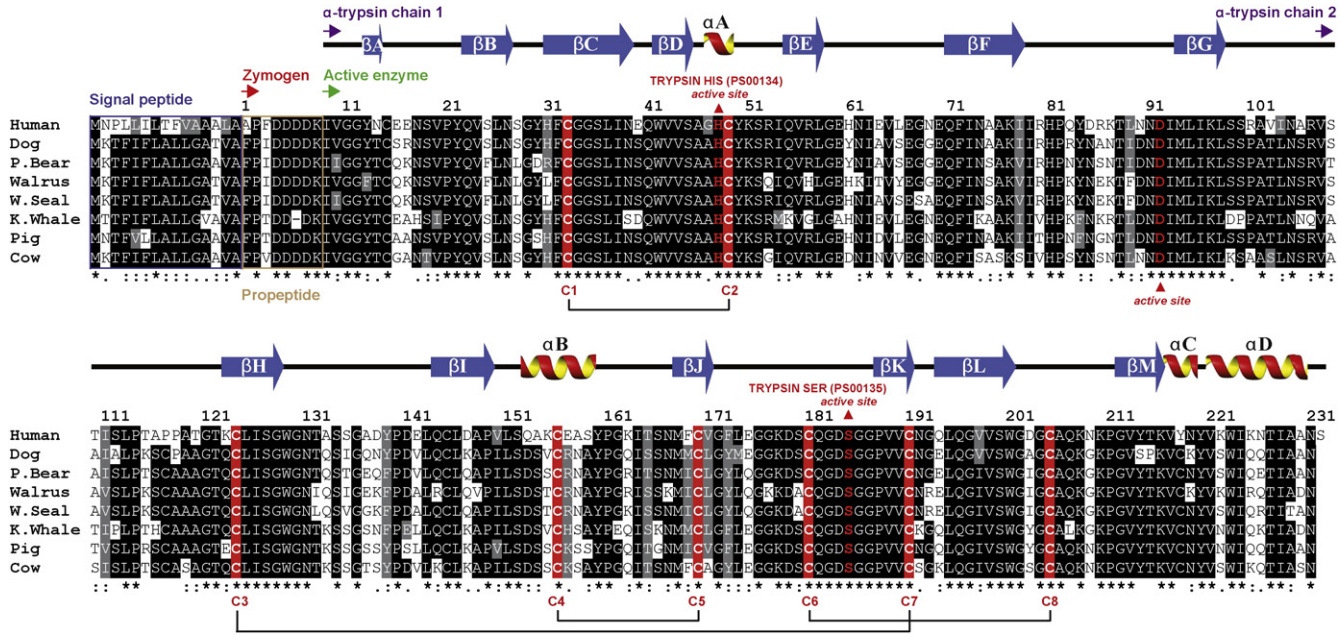


Fig. 6. Alignment of selected eutherian trypsinogen precursor proteins. Eutherian trypsinogen precursor proteins were aligned with Clustal Omega. Human [GenBank: NP_002760], dog [GenBank: P06871], Polar bear [GenBank: XP_008695541], Pacific walrus [GenBank: XP_004416603], Weddell seal [GenBank: XP_006743617], killer whale [GenBank: XP_004272168], pig [GenBank: NP_001156363], and cow [GenBank: NP_00110719]. The signal peptide and propeptide are boxed (blue and brown respectively) and the start site of the active trypsin enzyme shown (green arrow). The four conserved disulphide bonds are indicated, as well as the catalytic triad (bold red text). The 3D structure of trypsin [based on bovine trypsin (PDB ID: 1TFX chain A) and porcine trypsin (PDB ID: 1TFX chain A), is shown above the alignment: coiled regions, black bar; α-helices, red and yellow; and β-sheets, blue arrows. Conserved sites are shaded with BoxShade. Residues common to all species (*) are indicated and those shared by at least 6 of 8 species (>70%) are shaded black. Grey residues indicate conserved residue type. Conservation between groups of amino acids with strongly similar properties (score > 0.5, Gonnet PAM 250 matrix) is shown (:) and between those with weakly similar properties (score < 0.5, Gonnet PAM 250 matrix) also noted (.). Gaps within the alignment are indicated (—).

48 h pp. However, no conclusions could be drawn due to the limited number of animals available for this study. The lack of *CTI* expression in the pregnant, primiparous Cape fur seal mammary gland, in contrast to the multiparous seal at the same stage of gestation, may reflect the limited development of the primiparous gland (Cane, 2005), and its reduced secretory capacity (Lang et al., 2012).

4.3. Cape fur seal CTI — a putative trypsin inhibitor

Structural homology modelling based on human TFPI KD2 confirmed that Cape fur seal CTI forms a disulphide rich alpha + beta fold protein stabilised by 3 disulphide bonds. Although the Cape fur seal trypsin sequence is unknown, the high conservation between the active eutherian trypsin enzymes suggested that walrus trypsin was an excellent substitute for otariid trypsin for protease-inhibitor structural homology modelling using the human TFPI KD2-porcine trypsin template (PDB ID: 1TFX). Furthermore, the conservation of a basic P_1 Arg residue in Cape fur seal CTI and the catalytic triad of residues (Ser185, His48 and Asp92) in walrus trypsin strongly supported the inhibition of walrus trypsin by Cape fur seal CTI in a canonical mechanism of interaction characteristic to Kunitz inhibitors such as BPTI (Krowarsch et al., 2003; Laskowski and Kato 1980).

The precise *in vivo* target of eutherian CTI is unknown, but pinniped CTI may have similar anti-protease activity to bovine CTI. Hence, Cape fur seal and walrus CTI may inhibit trypsin, α-chymotrypsin and plasmin, but be inactive against β-chymotrypsin (Wu and Laskowski 1955), kallikrein, thrombin and renin (Pineiro et al., 1978). Similar to human TFPI KD2, CTI may also inhibit Factor Xa, a key component in the blood coagulation pathway (Burgering et al., 1997). In addition, the resistance of bovine CTI to pepsin digestion (Kassell and Laskowski, 1956) suggests that pinniped CTI may remain intact on its passage through the gastrointestinal tract and so may enter the circulatory system of the young pup. Therefore, CTI may have multiple sites of action against one or more serine peptidases in the maternal mammary gland and milk, as well as in the gastrointestinal tract and/or circulatory system/other organs of the young. Future studies of Cape fur seal CTI activity could involve the preparation and testing of recombinant CTI against serine endopeptidases.

4.4. Putative CTI functions in the mammary gland and neonate

Eutherian CTI may have multiple functions in the mother and young. In particular, as a putative serine protease inhibitor, it may prevent the digestion of milk proteins in the mammary gland and milk before their ingestion by the infant. The proteases in pinniped milk have yet to be characterised, but human milk has four major active proteases (Khaldi et al., 2014). These include a serum-derived trypsin-like enzyme and plasmin, plus elastase and cathepsin D, which are produced by the mammary gland. With the exception of cathepsin D, all are serine proteases, which may be inhibited by CTI. Notably, bovine milk contains neither trypsin, nor chymotrypsin, but does contain plasmin, one of the most active proteases in milk (Dallas et al., 2015). Hence, CTI may protect the maternal mammary gland and neonatal tissues from self-digestion by plasmin which is involved in tissue remodelling. CTI may also safeguard the neonate intestines and other tissues from maternal immune cells. These cells are absorbed across the intestinal epithelium and enter the newborn's circulatory system but release trypsin and elastase (Reber et al., 2008; Salmon, 2000). CTI may also influence cell-cell interactions by inhibiting trypsin and other enzymes that alter cell surface proteins.

The positive correlation of both CTI and IgG (immunoglobulin G) content in bovine colostrum (Laskowski and Laskowski, 1951; Pineiro et al., 1978; Quigley et al., 1995) and the passive transfer of immunoglobulins from mother to young via colostrum for the first 24–48 h pp (Veselsky et al., 1978) suggests that CTI protects IgG against proteolytic degradation in the neonate intestines (Brambell 1970; Quigley et al., 1995). Although the newborn calf has innate immunity, it has insufficient lymphocytes to mount an adaptive (specific) immune

response to protect its systemic and mucosal compartments (Brambell 1970; Salmon, 2000). Therefore, IgG transfer prior to gut 'closure' (cessation of passage of macromolecules) is essential for the survival of the young calf (Brambell, 1970; Quigley et al., 1995) and by analogy, the Cape fur seal pup.

4.5. Potential role of oligosaccharides in CTI function

The conservation of a sugar attachment site at Asn42 in all CTI and ELP peptides; with human TFPI KD2 glycosylated at an adjacent asparagine, in contrast to the non-glycosylated BPTI suggests that N-linked oligosaccharides are important for CTI function. The glycan attachment site at the protein's 'base' (Klauser et al., 1978) is unlikely to interfere with the P_1 protease reactive site at the 'opposite' end of the protein and so both 'ends' of CTI may have different biological roles. Oligosaccharides have anti-inflammatory and anti-adhesion properties acting as prebiotics, or as chemical messengers in cell-cell communication and may also bind directly to immune cells and influence their migration (Jost et al., 2015; Urashima et al., 2012; Zivkovic et al., 2011). Most importantly, oligosaccharides can act as soluble receptor analogues, or decoys for epithelial cell surface receptors which bind bacterial and viral pathogens (Jost et al., 2015; Kunz et al., 2014; Urashima et al., 2012). Hence microorganisms bind to these decoys, preventing infections in the young (Pacheco et al., 2015).

The cell- and species-specific nature of glycan production (Pacheco et al., 2015; Urashima et al., 2012) suggests that milk oligosaccharides may have unique functions. For example, milk oligosaccharides influence gut microbiota and gut maturation (Jost et al., 2015; Zivkovic et al., 2011). Many of these sugars are resistant to digestion (Pacheco et al., 2015) and so may pass intact through the intestines into the circulatory system and be transported to other tissues and organs (Kunz et al., 2014). Similarly, attached sugars may also increase the longevity of glycolipids and glycoproteins by preventing their non-specific degradation in the gut, as well as improving bioactivity, structure, stability and solubility (Roth et al., 2010). Hence N-glycosylated CTI may persist in the gut of the pup and enter the circulatory system. As the half-life of maternal IgG in newborn serum ranges from 12 to 20 days (Murphy et al., 2014), the extended period of *CTI* expression may provide prolonged protection for IgG and other milk proteins in the pup. This would be advantageous, particularly during the long periods of fasting that the pup experiences while its mother forages at-sea.

In summary, there appears to be no consistent set of characteristics that distinguish species that have retained the *CTI* gene from those that have a *CTI* pseudogene. However, *CTI* has not been reported in species that rely solely on the placental transfer of maternal immunoglobulins (humans, chimpanzees, rabbits). In contrast, *CTI* or the orthologous marsupial *ELP* gene are present in mammals that transfer immunoglobulins exclusively via colostrum (cow, pig, grey short-tailed opossum), or those that utilise both methods of transfer (dog, cat, pinnipeds, tammar wallaby) (Brambell 1970; Cavagnolo, 1979; Cavagnolo and Vedros, 1979; Deane et al., 1990). It would be of interest in the future to identify the target protease(s) of CTI and/or constituent(s) with which CTI interacts.

5. Conclusions

Colostrum trypsin inhibitor is expressed in the mammary gland of the lactating Cape and subantarctic fur seals (Otariidae) during the first 2–3 months pp, but unlike for other eutherians, *CTI* expression may be cyclical; up-regulated on-shore and down-regulated at-sea. The retention of a putative protein-coding *CTI* gene in the Pacific walrus (Odobenidae), but its disruption in the Weddell seal (Phocidae), was unexpected. Structural homology modelling of Cape fur seal CTI suggests that it, like other members of the BPTI/Kunitz family, inhibits trypsin-like serine proteases in a reversible, canonical mechanism. Therefore, pinniped CTI may prevent the proteolytic degradation of immunoglobulins and other milk proteins in the gastrointestinal tract and circulatory system of the young. Furthermore, glycosylated CTI may bind pathogenic bacteria in the pinniped pup gut, thereby preventing systemic infections and enhancing the survival of the young.

Acknowledgements

This research was funded by the Cooperative Research Centre for Innovative Dairy Products (CRC-IDP) and the Department of Zoology, The University of Melbourne. EAP was the recipient of an Australian Postgraduate Award (APA) and KNC a Dairy Australia Postgraduate Student Award and both EAP and KNC received a scholarship top-up from the CRC-IDP. AMB is a National Health and Medical Research Senior Research Fellow (1022688). The assistance of Mike Meyer, Chris Wilke, Deon Kotze, and Stefan Swanson in collecting Cape fur seal mammary tissue samples is gratefully acknowledged. We also thank the captain and crew of the *MV Sardinops* for logistical support without which this work could not have been conducted. Professor Marilyn Renfree provided insightful comments on this manuscript. This research was conducted under permits from the Department of Environmental Affairs and Tourism, Republic of South Africa.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2015.11.042.

References

Arnould, J.P.Y., Boyd, I.L., 1995. Temporal patterns of milk production in Antarctic fur seals (*Arctocephalus gazella*). J. Zool. (Lond.) 237, 1–12.
Ascenzi, P., Bocedi, A., Bolognesi, M., Spallarossa, A., Coletta, M., De Cristofaro, R., Menegatti, E., 2003. The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor): a milestone protein. Curr. Protein Pept. Sci. 4, 231–251.
Atkinson, S., 1997. Reproductive biology of seals. Rev. Reprod. 2, 175–194.
Bininda-Emonds, O.R., Cardillo, M., Jones, K.E., MacPhee, R.D., Beck, R.M., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., Purvis, A., 2007. The delayed rise of present-day mammals. Nature 446, 507–512.
Bonner, W.N., 1984. Lactation strategies in pinnipeds: problems for a marine mammalian group. Symp. Zool. Soc. Lond. 51, 253–272.
Brambell, F.W.R., 1970. The Transmission of Passive Immunity From Mother to Young, Vol 18. North-Holland Publishing Company, Amsterdam.
Brew, K., 2003. α-Lactalbumin. In: Fox, P.F., McSweeney, P.L.H. (Eds.), Advanced Dairy ChemistryProteins vol. 1. Kluwer Academic/Plenum Publishers, New York, pp. 387–419.
Burgering, M.J., Orbons, L.P., van der Doelen, A., Mulders, J., Theunissen, H.J., Grootenhuis, P.D., Bode, W., Huber, R., Stubbs, M.T., 1997. The second Kunitz domain of human tissue factor pathway inhibitor: cloning, structure determination and interaction with factor Xa. J. Mol. Biol. 269, 395–407.
Campanella, J.J., Bitincka, L., Smalley, J., 2003. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinf. 4, 29.
Cane, K.N., 2005. The Physiological and Molecular Regulation of Lactation in Fur Seals. PhD Thesis, Department of Zoology. University of Melbourne, Melbourne.
Cañestro, C., Albalat, R., Irimia, M., Garcia-Fernàndez, J., 2013. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. Semin. Cell Dev. Biol. 24, 83–94.
Cavagnolo, R.Z., 1979. The immunology of marine mammals. Dev. Comp. Immunol. 3, 245–257.
Cavagnolo, R.Z., Vedros, N.A., 1979. Serum and colostrium immunoglobulin levels in the northern fur seal *Callorhinus ursinus*. Dev. Comp. Immunol. 3, 139–146.
Costa, D.P., 1991. Reproductive and foraging energetics of pinnipeds: implications for life history patterns. In: Renouf, D. (Ed.), The Behaviour of Pinnipeds. Springer, Dordrecht, pp. 300–344.
Creighton, T.E., Charles, I.G., 1987. Sequences of the genes and polypeptide precursors for two bovine protease inhibitors. J. Mol. Biol. 194, 11–22.
Dallas, D.C., Murray, N.M., Gan, J., 2015. Proteolytic systems in milk: perspectives on the evolutionary function within the mammary gland and the infant. J Mammary Gland Biol. Neoplasia 20, 133–147.
Deane, E.M., Cooper, D.W., Renfree, M.B., 1990. Immunoglobulin G levels in fetal and newborn tammar wallabies (*Macropus eugenii*). Reprod. Fertil. Dev. 2, 369–375.
Edwards, M.J., Hinds, L.A., Deane, E.M., Deakin, J.E., 2012. A review of complementary mechanisms which protect the developing marsupial pouch young. Dev. Comp. Immunol. 37, 213–220.
Feeney, R.E., Allison, R.G., 1969. Evolutionary Biochemistry of Proteins : Homologous and Analogous Proteins from Avian egg Whites, Blood Sera, Milk, and Other Substances. Wiley, New York.

Fry, B.G., Roelants, K., Champagne, D.E., Scheib, H., Tyndall, J.D., King, G.F., Nevalainen, T.J., Norman, J.A., Lewis, R.J., Norton, R.S., et al., 2009. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annu. Rev. Genomics Hum. Genet. 10, 483–511.
Gamel, C.M., Davis, R.W., David, J.H., Meÿer, M.A., Brandon, E., 2005. Reproductive energetics and female attendance patterns of Cape fur seals (*Arctocephalus pusillus pusillus*) during early lactation. Am. Midl. Nat. 153, 152–170.
Garcia-Fernandez, R., Perbandt, M., Rehders, D., Ziegelmuller, P., Piganeau, N., Hahn, U., Betzel, C., Chavez Mde, L., Redecke, L., 2015. Three-dimensional structure of a Kunitz-type inhibitor in complex with an elastase-like enzyme. J. Biol. Chem. 290, 14154–14165.
Gentry, R.L., Kooyman, G.L., 1986. Fur Seals: Maternal Strategies on Land and at Sea. Princeton University Press, Princeton N. J.
Georges, J.-Y., Guinet, C., 2000. Maternal care in the subantarctic fur seals on Amsterdam Island. Ecology 81, 295–308.
Gojobori, T., Ikeo, K., 1994. Molecular evolution of serine protease and its inhibitor with special reference to domain evolution. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 344, 411–415.
Grzesiak, A., Helland, R., Smalas, A.O., Krowarsch, D., Dadlez, M., Otlewski, J., 2000. Substitutions at the P1′ position in BPTI strongly affect the association energy with serine proteinases. J. Mol. Biol. 301, 205–217.
Hanson, W.M., Beeser, S.A., Oas, T.G., Goldenberg, D.P., 2003. Identification of a residue critical for maintaining the functional conformation of BPTI. J. Mol. Biol. 333, 425–441.
Hawdon, J.M., Datu, B., Crowell, M., 2003. Molecular cloning of a novel multidomain Kunitz-type proteinase inhibitor from the hookworm *Ancylostoma caninum*. J. Parasitol. 89, 402–407.
Higdon, J.W., Bininda-Emonds, O.R., Beck, R.M., Ferguson, S.H., 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evol. Biol. 7, 216.
Huber, R., Kukla, D., Ruhlmann, A., Steigemann, W., 1972. Pancreatic trypsin inhibitor (Kunitz). I. Structure and function. Cold Spring Harb. Symp. Quant. Biol. 36, 141–148.
Ikeo, K., Takahashi, K., Gojobori, T., 1992. Evolutionary origin of a Kunitz-type trypsin inhibitor domain inserted in the amyloid beta precursor protein of Alzheimer's disease. J. Mol. Evol. 34, 536–543.
Jamal, F., Pandey, P., Singh, D., Khan, M.Y., 2013. Serine protease inhibitors in plants: nature's arsenal crafted for insect predators. Phytochem. Rev. 12, 1–34.
Jost, T., Lacroix, C., Braegger, C., Chassard, C., 2015. Impact of human milk bacteria and oligosaccharides on neonatal gut microbiota establishment and gut health. Nutr. Rev. 73, 426–437.
Kassell, B., Laskowski, M., 1956. The comparative resistance to pepsin of six naturally occurring trypsin inhibitors. J. Biol. Chem. 219, 203–210.
Khaldi, N., Vijayakumar, V., Dallas, D.C., Guerrero, A., Wickramasinghe, S., Smilowitz, J.T., Medrano, J.F., Lebrilla, C.B., Shields, D.C., German, J.B., 2014. Predicting the important enzymes in human breast milk digestion. J. Agric. Food Chem. 62, 7225–7232.
Kirkman, S.P., Bester, M.N., Hofmeyr, G.J.G., Pistorius, P.A., Makhado, A.B., 2002. Pup growth and maternal attendance patterns in subantarctic fur seals. Afr. Zool. 37, 13–19.
Klauser, R., Cechova, D., Tschesche, H., 1978. The carbohydrate linkage site of cow colostrum trypsin inhibitor. Hoppe Seylers Z. Physiol. Chem. 359, 173–180.
Krowarsch, D., Cierpicki, T., Jelen, F., Otlewski, J., 2003. Canonical protein inhibitors of serine proteases. Cell. Mol. Life Sci. 60, 2427–2444.
Kruse, P.E., 1983. The importance of colostral immunoglobulins and their absorption from the intestine of the newborn animals. Ann. Rech. Vet. 14, 349–353.
Kunz, C., Kuntz, S., Rudloff, S., 2014. Bioactivity of human milk oligosaccharides. In: Moreno, J.F., Sanz, M.L. (Eds.), Food Oligosaccharides: Production, Analysis and Bioactivity. John Wiley & Sons, Ltd., pp. 1–20.
Lang, S.L., Iverson, S.J., Bowen, W.D., 2012. Primiparous and multiparous females differ in mammary gland alveolar development: implications for milk production. J. Exp. Biol. 215, 2904–2911.
Laskowski, M., Kato, I., 1980. Protein inhibitors of proteinases. Annu. Rev. Biochem. 49, 593–626.
Laskowski Jr., M., Laskowski, M., 1951. Crystalline trypsin inhibitor from colostrum. J. Biol. Chem. 190, 563–573.
Laskowski, M., Qasim, M.A., 2000. What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate Complexes? Biochim. Biophys. Acta 1477, 324–337.
Luo, Z.X., Yuan, C.X., Meng, Q.J., Ji, Q., 2011. A Jurassic eutherian mammal and divergence of marsupials and placentals. Nature 476, 442–445.
Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155.
Marquart, M., Walter, J., Deisenhofer, J., Bode, W., Huber, R., 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Crystallogr. B 39, 480–490.
McFadden, T.B., Besser, T.E., Barrington, G.M., 1997. Regulation of immunoglobulin transfer into mammary secretions of ruminants. In: Welch, R.A.S., Burns, D.J.W., Davis, S.R., Popay, A.I., Prosser, C.G. (Eds.), Milk Composition, Production and Biotechnology. CAB International, New York, pp. 133–152.
Murphy, J.M., Hagey, J.V., Chigerwe, M., 2014. Comparison of serum immunoglobulin G half-life in dairy calves fed colostrum, colostrum replacer or administered with intravenous bovine plasma. Vet. Immunol. Immunopathol. 158, 233–237.
Nakahara, Y., Miyata, T., Hamuro, T., Funatsu, A., Miyagi, M., Tsunasawa, S., Kato, H., 1996. Amino acid sequence and carbohydrate structure of a recombinant human tissue factor pathway inhibitor expressed in Chinese hamster ovary cells: one N- and two O-linked carbohydrate chains are located between Kunitz domains 2 and 3 and one N-linked carbohydrate chain is in Kunitz domain 2. Biochemistry 35, 6450–6459.
Neville, M.C., Allen, J.C., Watters, C., 1983. The mechanisms of milk secretion. In: Neville, M.C., Neifert, M.R. (Eds.), Lactation: Physiology, Nutrition, and Breast-Feeding. New York, Plenum Publishing Co, pp. 49–92.
Nicholas, K.R., Simpson, K.J., Wilson, M., Trott, J., Shaw, D., 1997. The tammar wallaby: a model to study putative autocrine-induced changes in milk composition. J. Mammary Gland Biol. Neoplasia 2, 299–310.
Nyakatura, K., Bininda-Emonds, O.R., 2012. Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. BMC Biol. 10, 12.
Oftedal, O.T., Boness, D.J., Tedman, R.A., 1987. The behavior, physiology, and anatomy of lactation in the pinnipedia. Curr. Mamm. 1, 175–245.
Pacheco, A.R., Barile, D., Underwood, M.A., Mills, D.A., 2015. The impact of the milk glycobiome on the neonate gut microbiota. Annu. Rev. Anim. Biosci. 3, 419–445.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera — a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.
Pharo, E.A., De Leo, A.A., Renfree, M.B., Thomson, P.C., Lefevre, C.M., Nicholas, K.R., 2012. The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene. BMC Evol. Biol. 12, 8.
Pineiro, A., Brock, J.H., Esparza, I., 1978. Isolation and properties of bovine colostral trypsin inhibitor. Ann. Rech. Vet. 9, 281–286.
Piotte, C.P., Grigor, M.R., 1996. A novel marsupial protein expressed by the mammary gland only during the early lactation and related to the Kunitz proteinase inhibitors. Arch. Biochem. Biophys. 330, 59–64.
Quigley, J.D., Martin, K.R., Dowlen, H.H., 1995. Concentrations of trypsin inhibitor and immunoglobulins in colostrum of Jersey cows. J. Dairy Sci. 78, 1573–1577.
Ranasinghe, S., McManus, D.P., 2013. Structure and function of invertebrate Kunitz serine protease inhibitors. Dev. Comp. Immunol. 39, 219–227.
Rawlings, N.D., Waller, M., Barrett, A.J., Bateman, A., 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 42, D503–D509.
Reber, A.J., Donovan, D.C., Gabbard, J., Galland, K., Aceves-Avila, M., Holbert, K.A., Marshall, L., Hurley, D.J., 2008. Transfer of maternal colostral leukocytes promotes development of the neonatal immune system part II. Effects on neonatal lymphocytes. Vet. Immunol. Immunopathol. 123, 305–313.
Rijnkels, M., Freeman-Zadrowski, C., Hernandez, J., Potluri, V., Wang, L., Li, W., Lemay, D.G., 2013. Epigenetic modifications unlock the milk protein gene loci during mouse mammary gland development and differentiation. PLoS One 8, e53270.
Roth, J., Zuber, C., Park, S., Jang, I., Lee, Y., Kysela, K.G., Le Fourn, V., Santimaria, R., Guhl, B., Cho, J.W., 2010. Protein N-glycosylation, protein folding, and protein quality control. Mol. Cells 30, 497–506.
Šali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815.
Salmon, H., 2000. Mammary gland immunology and neonate protection in pigs. Homing of lymphocytes into the MG. Adv. Exp. Med. Biol. 480, 279–286.
Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual, Third edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Sharp, J.A., Cane, K.N., Lefevre, C., Arnould, J.P., Nicholas, K.R., 2006a. Fur seal adaptations to lactation: insights into mammary gland function. Curr. Top. Dev. Biol. 72, 275–308.
Sharp, J.A., Cane, K.N., Mailer, S.L., Oosthuizen, W.H., Arnould, J.P., Nicholas, K.R., 2006b. Species-specific cell–matrix interactions are essential for differentiation of alveoli like structures and milk gene expression in primary mammary cells of the Cape fur seal (*Arctocephalus pusillus pusillus*). Matrix Biol. 25, 430–442.
Sharp, J.A., Cane, K., Arnould, J.P., Nicholas, K.R., 2005. The lactation cycle of the fur seal. J. Dairy Res. 72, 81–89.
Shigetomi, H., Onogi, A., Kajiwara, H., Yoshida, S., Furukawa, N., Haruta, S., Tanase, Y., Kanayama, S., Noguchi, T., Yamada, Y., et al., 2010. Anti-inflammatory actions of serine protease inhibitors containing the Kunitz domain. Inflamm. Res. 59, 679–687.
Tschesche, H., Klauser, R., Cechova, D., Jonakova, V., 1975. On the carbohydrate composition of bovine colostrum trypsin inhibitor. Hoppe Seylers Z. Physiol. Chem. 356, 1759–1764.
Urashima, T., Fukuda, K., Messer, M., 2012. Evolution of milk oligosaccharides and lactose: a hypothesis. Animal 6, 369–374.
Vandermarliere, E., Mueller, M., Martens, L., 2013. Getting intimate with trypsin, the leading protease in proteomics. Mass Spectrom. Rev. 32, 453–465.
Verrier, D., Groscolas, R., Guinet, C., Arnould, J.P.Y., 2011. Development of fasting abilities in subantarctic fur seal pups: balancing the demands of growth under extreme nutritional restrictions. Funct. Ecol. 25, 704–717.
Veselsky, L., Cechova, D., Jonakova, V., 1978. Secretion and immunochemical properties of the trypsin inhibitor from bovine colostrum. Hoppe Seylers Z. Physiol. Chem. 359, 873–878.
Watson, C., 2006. Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. Breast Cancer Res. 8, 203.
Wu, F.C., Laskowski, M., 1955. Action of the naturally occurring trypsin inhibitors against chymotrypsins α and β. J. Biol. Chem. 213, 609–619.
Xu, Y., Carr, P.D., Guss, J.M., Ollis, D.L., 1998. The crystal structure of bikunin from the inter-α-inhibitor complex: a serine protease inhibitor with two Kunitz domains. J. Mol. Biol. 276, 955–966.
Zivkovic, A.M., German, J.B., Lebrilla, C.B., Mills, D.A., 2011. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. Proc. Natl. Acad. Sci. 108, 4653–4658.

Supplementary file 1 – Nomenclature of species listed in this study

INFRACLASS	SUPERORDER	ORDER	FAMILY	COMMON NAME	SPECIES NAME
EUTHERIA	**Laurasiatheria**	**Carnivora**	**Otariidae**	Cape (South African) fur seal	*Arctocephalus pusillus pusillus*
				Subantarctic fur seal	*Arctocephalus tropicalis*
				Antarctic fur seal	*Arctocephalus gazella*
				California sea lion	*Zalophus californianus*
			Odobenidae	Pacific walrus	*Odobenus rosmarus divergens*
			Phocidae	Weddell seal	*Leptonychotes weddellii*
				Hooded seal	*Cystophora cristata*
			Canidae	Dog (Labrador, Boxer breeds)	*Canis familiaris*
			Ursidae	Giant panda	*Ailuropoda melanoleuca*
				Polar bear	*Ursus maritimus*
			Felidae	Cat (Abyssinian breed)	*Felis catus*
			Mustelidae	Domestic ferret	*Mustela putorius furo*
		Cetartiodactyla*	**Bovidae**	Cow (Hereford breed)	*Bos taurus*
				American (Plains) bison	*Bison bison bison*
				Goat (Yunnan black breed)	*Capra hircus*
				Sheep (Texel breed)	*Ovis aries*
				Wild yak	*Bos mutus*
				Tibetan antelope (Chiru)	*Pantholops hodgsonii*
				Water buffalo	*Bubalus bubalis*
			Suidae	Pig (Duroc breed)	*Sus scrofa*
			Camelidae	Alpaca	*Vicugna pacos*
				Bactrian camel	*Camelus bactrianus*
				Arabian camel	*Camelus dromedarius*
			Delphinidae	Killer whale	*Orcinus orca*
				Bottle-nosed dolphin	*Tursiops truncatus*
			Balaenopteridae	Scammon’s minke whale	*Balaenoptera acutorostrata scammoni*
		Perissodactyla	**Rhinocerotidae**	Southern white rhinoceros	*Ceratotherium*
			Equidae	Thoroughbred horse	*Equus caballus*
				Przewalski’s horse	*Equus przewalskii*
	Afrotheria	**Proboscidae**	**Elephantidae**	African savanna elephant	*Loxodonta africana simum simum*
	Xenarthra	**Pilosa**	**Megalonychidae**	Hoffmann's two-toed sloth	*Choloepus hoffmanni*
		Cingulata	**Dasypodida**	Nine-banded armadillo	*Dasypus novemcinctus*
	Euarchontoglires	**Primates**	**Hominidae**	Human	*Homo sapiens*
METATHERIA	**Australidelphia**	**Diprotodontia**	**Macropodidae**	Tammar wallaby	*Macropus eugenii*
			Phascolarctidae	Koala	*Phascolarctos cinereus*
			Phalangeridae	Brushtail possum	*Trichosurus vulpecula*
		Dasyuromorphia	**Dasyuridae**	Tasmanian devil	*Sarcophilus harrisii*
				Stripe-faced dunnart	*Sminthopsis macroura*
				Fat-tailed dunnart	*Sminthopsis crassicaudata*
	Ameridelphia	**Didelphimorphia**	**Didelphidae**	South American (Gray short-tailed) opossum	*Monodelphis domestica*
PROTOTHERIA		**Monotremata**	**Ornithorhynchidae**	Platypus	*Ornithorhynchus anatinus*
			Tachyglossidae	Echidna	*Tachyglossus aculeatus*

*The order Cetartiodactyla is often divided into Artiodactyla (even-toed ungulates) and Cetacea (whales, dolphins and porpoises)

Supplementary file 2 - Weddell seal *CTI* pseudogene - putative transcript and protein

A. The putative Weddell seal *CTI* pseudogene if transcribed would most likely encode **B.** a 76 amino acid precursor protein, which is cleaved and secreted in milk, but lacks a Kunitz domain C. Weddell seal *CTI* predicted transcript translated in three frames (green box = start codon; red box = stop codon) .

A.

```
ATGAAGTTCAGCCCGTTCCTTGTCCTCTGCGTCCCCCTCTGCCTGGTTGGCATTGCCAGCTCAGCCCATCTCAAGCAAGAGGC
CCCCGGGAGCTGTCCCAAACTCTGCCTGCCACGTGTCAGCTCCCTCCAGCGAGGGGCCCCTGCAGAGGCCTCTTCTACTGGTA
CTTCTACGACAACATGTCCAGTGAGTGTGAGCACTTCGCCTACGGTGGCTGTCAGGGCGCTGACAACAATCTTGAGACCACGG
AGATCTGTTTGAGGATCTGCAAGCACGCTGAGACAAGGTAAAGAGCAGCTGAAGACCATGTCTGCATCCTCCTGGCCTTGACT
GAAAGCTGGCTGGGCCTCCTTTCCTGCACCTAGTCCCTGCTTGGAACCCATCCTGTAGCCCTCCAGGCCTGGGTCAATAAA
```

B.

Putative signal peptide

Putative mature protein

```
MKFSPFLVLCVPLCLVGIASSAHLKQEAPGSCPKLCLPRVSSLQRGAPAEASSTGTSTTTCPVSVSTSPTVAVRA
LTTILRPRRSV
```

C.

5'3' Frame 1

```
atgaagttcagcccgttccttgtcctctgcgtcccccctctgcctggttggcattgccagc
 M  K  F  S  P  F  L  V  L  C  V  P  L  C  L  V  G  I  A  S
tcagcccatctcaagcaagaggcccccgggagctgtcccaaactctgcctgccacgtgtc
 S  A  H  L  K  Q  E  A  P  G  S  C  P  K  L  C  L  P  R  V
agctccctccagcgaggggcccctgcagaggcctcttctactggtacttctacgacaaca
 S  S  L  Q  R  G  A  P  A  E  A  S  S  T  G  T  S  T  T  T
tgtccagtgagtgtgagcacttcgcctacggtggctgtcagggcgctgacaacaatcttg
 C  P  V  S  V  S  T  S  P  T  V  A  V  R  A  L  T  T  I  L
agaccacggagatctgtttgaggatctgcaagcacgctgagacaaggtaaagagcagctg
 R  P  R  R  S  V  -  G  S  A  S  T  L  R  Q  G  K  E  Q  L
aagaccatgtctgcatcctcctggccttgactgaaagctggctgggcctcctttcctgca
 K  T  M  S  A  S  S  W  P  -  L  K  A  G  W  A  S  F  P  A
cctagtccctgcttggaacccatcctgtagccctccaggcctgggtcaataaa
 P  S  P  C  L  E  P  I  L  -  P  S  R  P  G  S  I
```

5'3' Frame 2

```
atgaagttcagcccgttccttgtcctctgcgtcccccctctgcctggttggcattgccagct
  -  S  S  A  R  S  L  S  S  A  S  P  S  A  W  L  A  L  P  A
cagcccatctcaagcaagaggcccccgggagctgtcccaaactctgcctgccacgtgtca
 Q  P  I  S  S  K  R  P  P  G  A  V  P  N  S  A  C  H  V  S
gctccctccagcgaggggcccctgcagaggcctcttctactggtacttctacgacaacat
 A  P  S  S  E  G  P  L  Q  R  P  L  L  L  V  L  L  R  Q  H
gtccagtgagtgtgagcacttcgcctacggtggctgtcagggcgctgacaacaatcttga
 V  Q  -  V  -  A  L  R  L  R  W  L  S  G  R  -  Q  Q  S  -
gaccacggagatctgtttgaggatctgcaagcacgctgagacaaggtaaagagcagctga
 D  H  G  D  L  F  E  D  L  Q  A  R  -  D  K  V  K  S  S  -
agaccatgtctgcatcctcctggccttgactgaaagctggctgggcctcctttcctgcac
 R  P  C  L  H  P  P  G  L  D  -  K  L  A  G  P  P  F  L  H
ctagtccctgcttggaacccatcctgtagccctccaggcctgggtcaataaa
 L  V  P  A  W  N  P  S  C  S  P  P  G  L  G  Q  -
```

5'3' Frame 3

```
atgaagttcagcccgttccttgtcctctgcgtcccccctctgcctggttggcattgccagctc
   E  V  Q  P  V  P  C  P  L  R  P  P  L  P  G  W  H  C  Q  L
agcccatctcaagcaagaggcccccgggagctgtcccaaactctgcctgccacgtgtcag
 S  P  S  Q  A  R  G  P  R  E  L  S  Q  T  L  P  A  T  C  Q
ctccctccagcgaggggcccctgcagaggcctcttctactggtacttctacgacaacatg
 L  P  P  A  R  G  P  C  R  G  L  F  Y  W  Y  F  Y  D  N  M
tccagtgagtgtgagcacttcgcctacggtggctgtcagggcgctgacaacaatcttgag
 S  S  E  C  E  H  F  A  Y  G  G  C  Q  G  A  D  N  N  L  E
accacggagatctgtttgaggatctgcaagcacgctgagacaaggtaaagagcagctgaa
 T  T  E  I  C  L  R  I  C  K  H  A  E  T  R  -  R  A  A  E
gaccatgtctgcatcctcctggccttgactgaaagctggctgggcctcctttcctgcacc
 D  H  V  C  I  L  L  A  L  T  E  S  W  L  G  L  L  S  C  T
tagtccctgcttggaacccatcctgtagccctccaggcctgggtcaataaa
 -  S  L  L  G  T  H  P  V  A  L  Q  A  W  V  N  K
```

Supplementary file 3 – Accession numbers of sequences used in this study

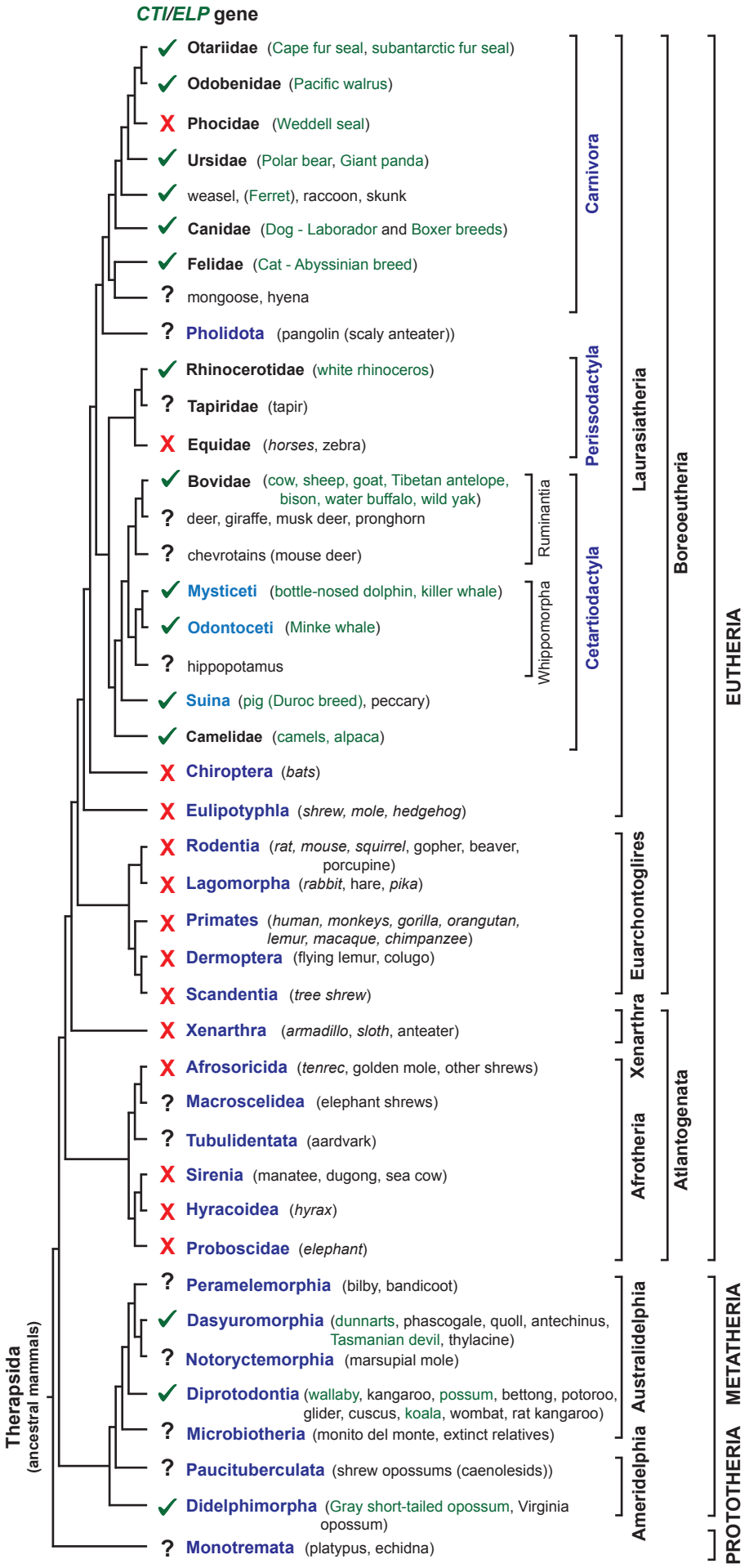
GENE*	COMMON NAME	ACCESSION NUMBER		
		Gene ID/locus	transcript	protein
CTI	Cape fur seal	KC152480	KC152479	AFZ99004
	Pacific walrus	LOC101362989	XM_012567271.1	XP_012422725
	Weddell seal	LOC102745256	-	-
	Dog (Labrador, Boxer breeds)	BK008082	JN191342	AEN62470
	Giant panda	BK008084	BK008084	DAA35187
	Polar bear	LOC103669871, BK009384	BK009384	BK009384
	Cat (Abyssinian breed)	BK008083	BK008083	DAA35186
	Cow (Hereford breed)	ENSBTAG00000016127, LOC104973896		
	Cow (Holstein-Friesian breed)		JN191341	AEN62469
	American (Plains) Bison	BK009385	BK009385	BK009385
	Goat (Yunnan black breed)	LOC102174856		XP_00568864
	Sheep (Texel breed)	LOC105613397		XP_004014646
	Wild yak	LOC102286180		ELR57883
	Tibetan antelope (Chiru)	LOC102333010		XP_005954742
	Water buffalo	LOC102405068		
	Pig (Duroc breed)	LOC100513767		ENSSSCT00000008098
	Alpaca	LOC102539473	XM_006202709	XP_006202771
	Wild Bactrian camel	LOC105087824	XM_006193535	XP_006193597
	Arabian camel	LOC105087824	XM_010978588	XP_010976890
	Killer whale	LOC101273120	XM_004272906	XP_004272954
	Bottle-nosed dolphin	BK008085, LOC101340157	BK008085	DAA35188
	Scammon's minke whale	LOC103003268	XM_007193070	XP_007193132
	Southern white rhinoceros	LOC101393639	XM_004430556	XP_004430613
ELP	Tammar wallaby	JN191335	JN191338	AEN62466
	Southern koala	JN191337		AEN62465
	Brushtail possum	KP260654		U34208
	Tasmanian devil	BK008719		BK008719
	Stripe-faced dunnart	AC186006		
	Fat-tailed dunnart	JN191339	JN191339	AEN62467
	South American (Gray short-tailed) opossum	BK008086	JN191340	AEN62468
BPTI	Cow (Hereford breed)	404172	BC149368	AAI49369
SLPI	Cow (Hereford breed)	504598	NM_001098865	NP_001092335
TFPI	Human	7035	NM_006287	NP_006278
PRSS1 (Trypsin)	Pacific walrus	LOC101378334	XM_004416546	XP_004416603
	Weddell seal	LOC102739415	XM_006743554	XP_006743617
	Polar bear	LOC103669003	XM_008697319	XP_008695541
	Dog	LOC475521	XM_532744	P06871
	Pig	LOC100302368	NM_001162891	NP_001156363
	Cow (Hereford breed)	LOC780933	NM_001113727	NP_00110719
	Killer whale	LOC101274113	XM_004272120	XP_004272168
	Human		NM_002769	NP_002760

**CTI, colostrum trypsin inhibitor*; *ELP, early lactation protein*; *BPTI, bovine pancreatic trypsin inhibitor*; *SLPI, secretory leukocyte protease inhibitor*; *TFPI, tissue factor pathway inhibitor*; *PRSS1 , protease, serine, 1 (trypsin 1).*

Supplementary file 4 – Structure of the eutherian *CTI*, marsupial *ELP* and bovine *PTI* genes

Gene	Species	Gene Size# (bp)	Exon 1 (bp)	Intron 1 (bp)	Exon 2 (bp)	Intron 2 (bp)	Exon 3 (bp)	3' utr (bp)+	Transcript Size# (bp)	Transcript Size (CDS*) (bp)
CTI	Cape fur seal	**2,679**	**64**	1,762	**216**	503	**134**	126	414	291
	Pacific walrus	**2,604**	**64**	1,690	**216**	500	**134**	126	415	291
	Polar bear	**2,561**	**76**	1,645	**216**	490	**134**	126	426	303
	Giant panda	**2,561**	**76**	1,645	**216**	490	**134**	126	426	303
	Dog (Labrador, Boxer breeds)	**2,578**	**76**	1,671	**216**	491	**124**	116	416	303
	Cat (Abyssinian breed)	**3,283**	**76**	2,346	**216**	511	**134**	126	426	303
	Cow (Hereford breed)	**2,907**	**76**	1,788	**216**	694	**133**	116	425	312
	Wild yak	**2,907**	**76**	1,788	**216**	694	**133**	116	425	312
	American (Plains) Bison	**2,907**	**76**	1,788	**216**	694	**133**	116	425	312
	Water buffalo	**2,901**	**76**	1,787	**216**	690	**132**	115	424	312
	Sheep (Texel breed)	**2,773**	**76**	1,666	**216**	682	**133**	116	425	312
	Goat (Yunnan black breed)	**2,778**	**76**	1,673	**216**	680	**133**	116	425	312
	Tibetan antelope (Chiru)	**2,767**	**76**	1,654	**216**	688	**133**	116	425	312
	Pig (Duroc breed)	**3,771**	**76**	2,849	**216**	494	**136**	116	428	315
	Alpaca	**2,513**	**76**	1,586	**216**	509	**126**	106	418	315
	Bactrian camel	**2,514**	**76**	1,588	**216**	508	**126**	106	418	315
	Bottle-nosed dolphin	**2,720**	**76**	1,785	**216**	507	**136**	116	428	315
	Killer whale	**2,814**	**76**	1,879	**216**	507	**136**	116	428	315
	Scammon's minke whale	**2,831**	**76**	1,901	**216**	502	**136**	116	428	315
	Southern white rhinoceros	**2,606**	**76**	1,664	**216**	514	**136**	116	428	315
ELP	Tammar wallaby	**4,268**	**73**	3,522	**216**	311	**146**	126	435	312
	Southern koala	**1,387**	**73**	645	**210**	314	**142**	122	425	306
	Brushtail possum	**3,922**	**73**	3,178	**216**	304	**151**	134	440	309
	Tasmanian devil	**3,117**	**73**	2,360	**219**	311	**154**	134	446	315
	Stripe-faced dunnart	**4,818**	**73**	4,069	**216**	306	**154**	155	443	291
	Fat-tailed dunnart	**4,726**	**73**	3,960	**216**	323	**154**	155	443	291
	South American (Gray short-tailed) opossum^	**4,497**	**73**	2,219	**216**	1,831	**158**	157	445	291
BPTI	Cow (Hereford breed)	**4,789**	**99**	3,500	**191**	450	**547**	539	839	303

#From the start of translation (ATG) to the end of the polyadenylation signal (AATAAA), inclusive; +From the end of the stop codon to the end of the polyadenylation signal (AATAAA) inclusive; *CDS (Coding Sequence); ^Gene contains non-sequenced regions.



Supplementary file 5 - Distribution of the eutherian *CTI* and marsupial *ELP* gene in mammals

Cladogram depicting the relationship between mammals: eutherians, metatherians (marsupials) and protherians (monotremes) based on recent evolutionary studies (Bininda-Emonds et al., 2007; Hassanin et al., 2012; Meredith et al., 2011; Spaulding et al., 2009; Zhou et al., 2012). Species that have a putative functional (protein-coding) *CTI*/*ELP* gene are shown in green text (green ✓) and those with a pseudogene are indicated

by black italicised text (red X). Species for which the status of *CTI*/*ELP* has not been determined are indicated by black text (black ?). Mammalian orders are shown in bold, dark blue text, suborders in bold light blue text, families in bold black text. The status of the *CTI*/*ELP* gene in prototherians is unclear, but all marsupials investigated including both the Australidelphia and Ameridelphia superorders have retained the *ELP* gene. In contrast, *CTI* has become a pseudogene in the Afrotheria, Xenarthra and Euarchontoglires superorders. Within the Laurasiatherian superorder, the Cetartidodactyla have a functional *CTI* gene, but *CTI* is a pseudogene in the Chiroptera and Eulipotyphla. Not all Carnivora and Perissodactyla have a functional *CTI* gene, with the gene disrupted in the Weddell seal (Phocidae) and the horse (family Equidae).

References

Bininda-Emonds, O.R., Cardillo, M., Jones, K.E., MacPhee, R.D., Beck, R.M., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., and Purvis, A. (2007). The delayed rise of present-day mammals. *Nature 446*, 507-512.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen van Vuuren, B., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V., Nguyen, T.T.*, et al.* (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus Biologies 335*, 32-50.

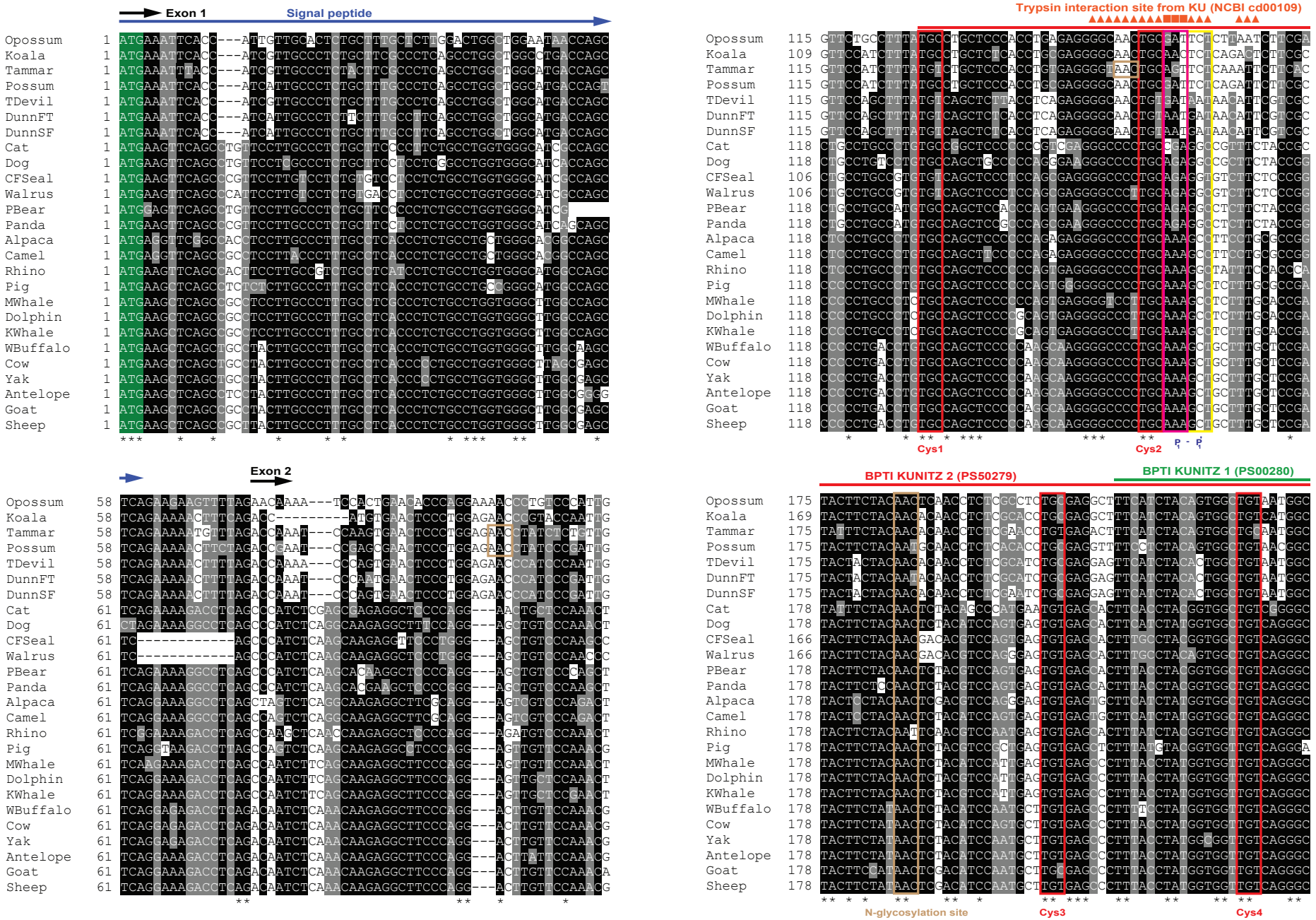
Meredith, R.W., Janecka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simao, T.L., Stadler, T.*, et al.* (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science *334*, 521-524.

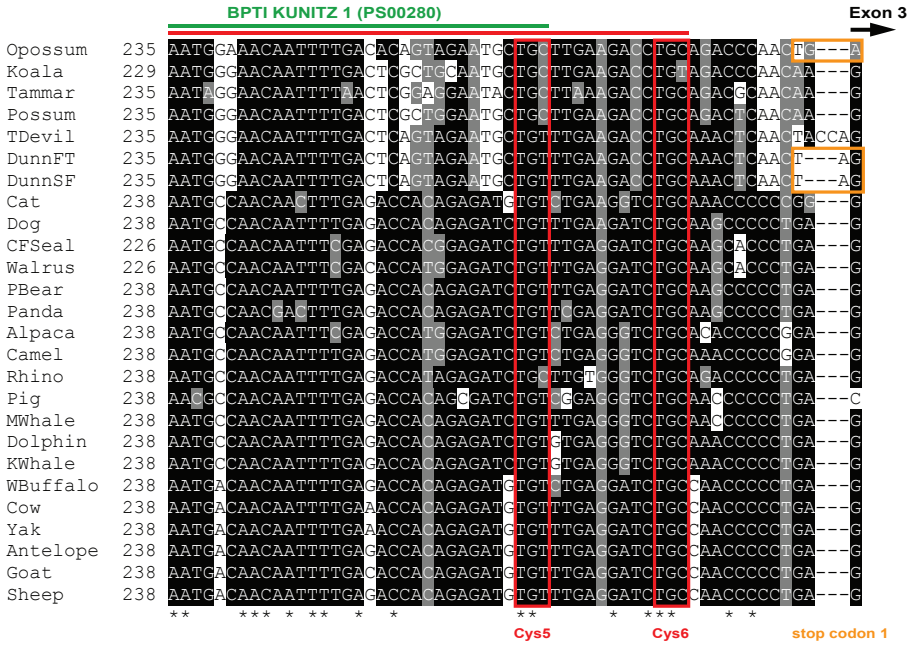
Spaulding, M., O'Leary, M.A., and Gatesy, J. (2009). Relationships of Cetacea (Artiodactyla) among mammals: increased taxon sampling alters interpretations of key fossils and character evolution. *PLoS One 4*, e7062.

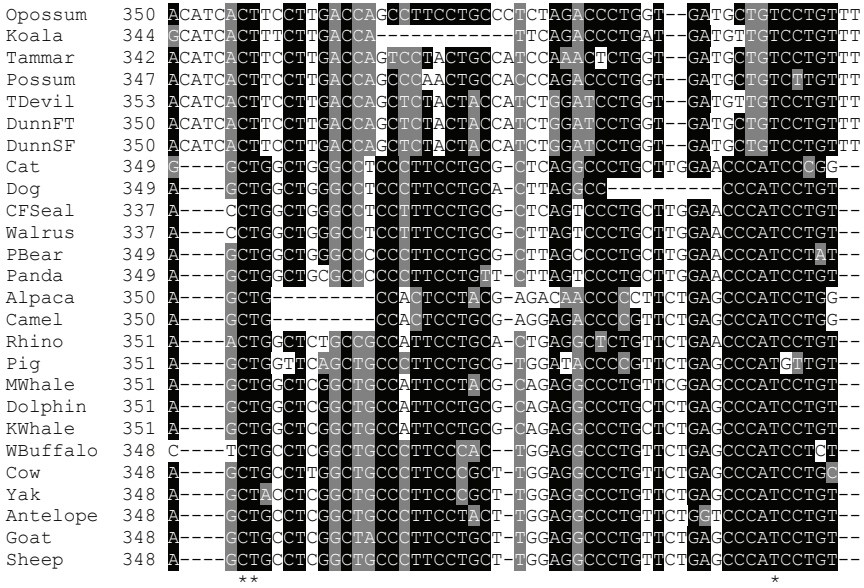
Zhou, X., Xu, S., Xu, J., Chen, B., Zhou, K., and Yang, G. (2012). Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the laurasiatherian mammals. *Syst Biol 61*, 150-164.

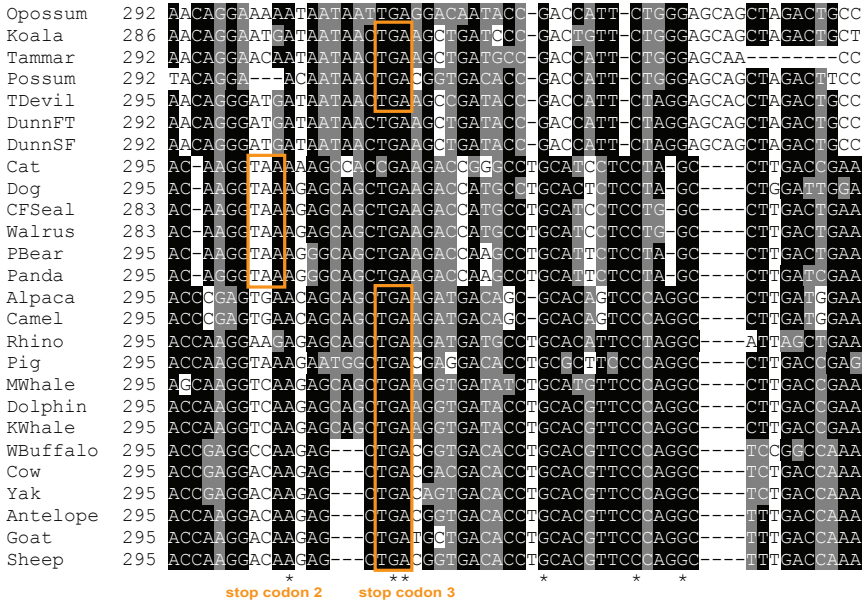
Supplementary file 6 – Alignment of eutherian *CTI* and marsupial *ELP* mRNA transcripts

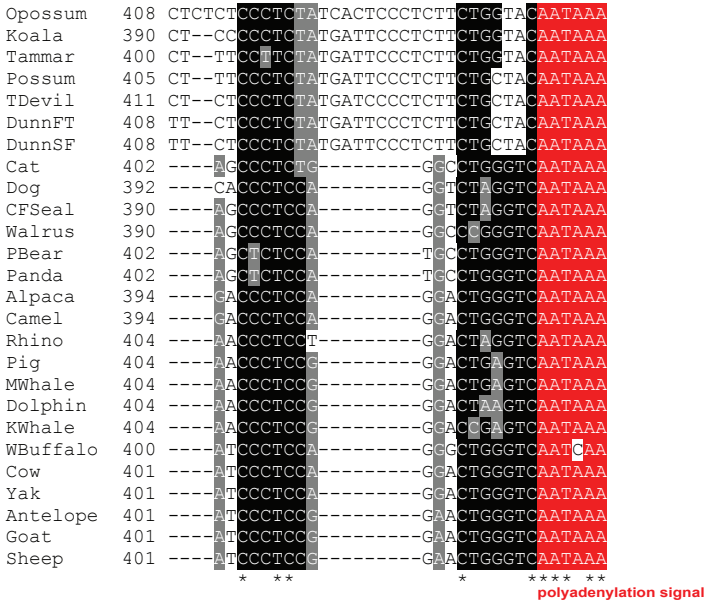
CTI and *ELP* transcripts from the putative translation start to the polyadenylation signal were aligned with Clustal W2 and conserved regions highlighted (BoxShade Server). Nucleotides common to all species are indicated (*). Nucleotides shared by more than 70% of the sequences (at least 19 of 26 sequences) are shaded black, and those of the same type (shaded grey), i.e. purine (A, adenine/G, guanine) or pyrimidine (C, cytosine/T, thymine). Exon start sites are arrowed (black), as is the signal peptide (blue). Codons encoding cysteine residues are boxed in red those encoding the P_1 and P_1' residues boxed in pink and yellow respectively. The BPTI KUNITZ 1 and 2 motifs are indicated by green and red bars, respectively and the trypsin interaction site indicated by orange triangles, with the P_1 encoding nucleotides shown as orange squares. Stop codons are boxed in orange. Putative N-glycosylation sites are boxed in brown.




BPTI KUNITZ 1 (PS00280)
Exon 3
Cys5
Cys6
stop codon 1

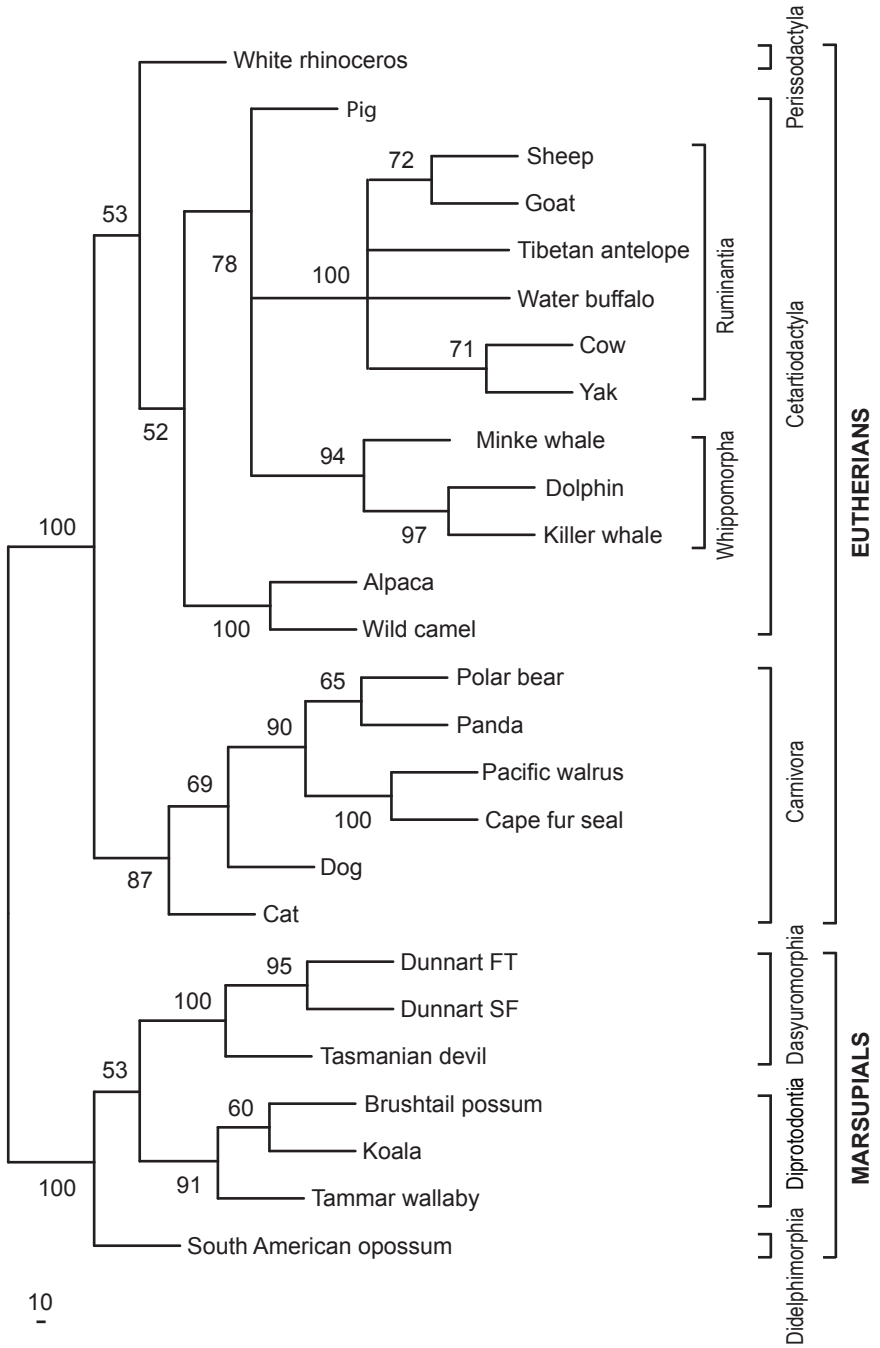



stop codon 2
stop codon 3


polyadenylation signal

Supplementary file 7. Phylogeny of the mature protein-coding *CTI* and *ELP* transcripts

A. The evolutionary relationship between the nucleotides encoding the mature CTI and ELP peptides was determined by maximum likelihood analysis based on the molecular clock assumption, with *SLPI* (secretory leukocyte protease inhibitor) as an outlier (*SLPI* omitted). **B.** Method of phylogenetic analysis and **C.** the *CTI*, *ELP* and *SLPI* sequences used. **A.** Cape fur seal and Pacific walrus *CTI* were most similar, as expected and the evolutionary tree was consistent with a previous phylogeny of *CTI* and *ELP* (Pharo et al., 2012) as well as that of the Carnivora and indeed Laurasiatherian phylogeny (Higdon et al., 2007; Zhou et al., 2012), with the exception of the split between Cetartiodactyla and Perissodactyla. Numbers at branch points indicate confidence levels as determined by bootstrap (100) values and the length of the lines from the nodes indicates evolutionary distances.



B.

The phylogenetic relationships between the mature protein-coding region of the *CTI* and *ELP* transcripts were inferred with MEGA6 (Tamura et al., 2013) and PHYLIP version 3.69 (http://evolution.genetics.washington.edu/phylip.html) using both character-based (maximum likelihood and parsimony) and distance-based methods. Mr Bayes version 3.2 (Ronquist et al., 2012) and Codon PhyML version 1.00 (Gil et al., 2013) were also used to validate bootstrapped trees. Nucleotide sequences encoding the mature ELP and CTI peptides and an outgroup sequence, bovine SLPI were aligned in MEGA6 using the MUSCLE (codon) alignment option. All codon positions containing gaps were removed and 100 bootstrapped alignments generated with seqboot (PHYLIP). Phylogenies were estimated by maximum likelihood analysis using a molecular clock assumption (dnamlk, PHYLIP) based on a hidden Markov Model with 5 categories of rate differences among sites, global rearrangements, a randomized input order jumbled once and a transition/transversion ratio of 1.42. A discrete gamma distribution with a shape of α=1.4027 (MEGA6) and a coefficient of variation of 0.844 were used to model rate differences across sites. An unrooted tree was generated (consense, PHYLIP), plotted (drawgram, PHYLIP) and the SLPI outgroup removed.

C.

```
27 201
DunnFT      GACCAAATCA ACTCCCTGGA GAACCCATCC CGATTGGTTC CAGCTTTATG TCAGCTCTCA
DunnSF      GACCAAATCA ACTCCCTGGA GAACCCATCC CGATTGGTTC CAGCTTTATG TCAGCTCTCA
Koala       GACCATGTGA ACTCCCTGGA GAACCCGTAC CAATTGGTTC CATCTTTATG CCTGCTCTCA
Opossum     GAACAAAATA ACACCCAGGA AAACCCTGTC CCATTGGTTC TGCCTTTATG CCTGCTCCCA
Possum      GACCGAATCA ACTCCCTGGA GAACCTATCC CGATTGGTTC CATCTTTATG CCTGCTCCCA
Tammar      GACCAAATCA ACTCCCTGGA GAACCTATCT CTGTTGGTTC CATCTTTATG TCTGCTCCCA
TDevil      GACCAAAACA ACTCCCTGGA GAACCCATCC CAATTGGTTC CAGCTTTATG TCAGCTCTTA
Cat         GCCCATCTCG AGGCTCCCCA GGAACTGCTC CAAACTCTGC CTGCCCTGTG CCGGCTCCCC
Dog         GCCCATCTCG AGGCTTTCCA GGAGCTGTCC CAAACTCTGC CTGTCCTGTG CCAGCTGCCC
CFSeal      GCCCATCTCG AGGTTCCCTG GGAGCTGTCC CAAGCCCTGC CTGCCGTGTG TCAGCTCCCT
Panda       GCCCATCTCG AAGCTCCCCG GGAGCTGTCC CAAGCTCTGC CTGCCATGTG CCAGCTCCGC
Walrus      GCCCATCTCG AGGCTCCCTG GGAGCTGTCC CAACCCCTGC CTGCCGTGTG TCAGCTCCCT
Alpaca      GCTAGTCTCG AGGCTTCGCA GGAGTCGTCC CAGACTCTCC CTGCCCTGTG CCAGCTCCCC
Antelope    GACAATCTCG AGGCTTCCCA GGACTTATTC CAAACGCCCC CTGACCTGTG CCAGCTCCCC
Camel       GCCAGTCTCG AGGCTTCGCA GGAGTCGTCC CAGACTCTCC CTGCCCTGTG CCAGCTTCCC
Cow         GACAATCTCG AGGCTTCCCA GGACTTGTTC CAAACGCCCC CTGACCTGTG CCAGCTCCCC
Dolphin     GCCAATCTTG AGGCTTCCCA GGAGTTGCTC CAAACTCCCC CTGCCCTCTG CCAGCTCCCC
Goat        GACAATCTCG AGGCTTCCCA GGACTTGTTC CAAACACCCC CTGACCTGTG CCAGCTCCCC
KWhale      GCCAATCTTG AGGCTTCCCA GGAGTTGCTC CGAACTCCCC CTGCCCTCTG CCAGCTCCCC
MWhale      GCCAATCTTG AGGCTTCCCA GGAGTTGTTC CAAACTCCCC CTGCCCTCTG CCAGCTCCCC
Pig         GCCAGTCTCG AGGCCTCCCA GGAGTTGTTC CAAACGCCCC CTGCCCTGTG CCAGCTCCCC
Rhino       GCCAAGCTCG AGGCTCCCCA GGAGATGTCC CAAACTCTCC CTGCCCTGTG CCAGCTCCCC
Sheep       GACAATCTCG AGGCTTCCCA GGACTTGTTC CAAACGCCCC CTGACCTGTG CCAGCTCCCC
WBuffalo    GACAATCTCG AGGCTTCCCA GGACTTGTTC CAAACGCCCC CTGACCTGTG CCAGCTCCCC
Yak         GACAATCTCG AGGCTTCCCA GGACTTGTTC CAAACGCCCC CTGACCTGTG CCAGCTCCCC
PBear       GCCCATCTCA AGGCTCCCCA GGAGCTGTCC CCAGCTCTGC CTGCCATGTG CCAGCTCCAC
bSLPI       GAAAATGCTG CTGGGGCCTG CCCTCCTAGA AAAACTACCC AGTGCCTTGG AGATGAGAAA

            CCTCAGTGTA ATGATAACAT TCGTCGCTAC TACTACAATA CAACCTCTCG CATCTGCGAG
            CCTCAGTGTA ATGATAACAT TCGTCGCTAC TACTACAACA CAACCTCTCG AATCTGCGAG
            CCTGCGTGCA ACTCTCAGAC TCTTCGCTAC TTCTACAACA CAACCTCTCG CACCTGCGAG
            CCTGAGTGCG ATTCTCTTAA TCTTCGATAC TTCTACAACT CAACCTCTCG CCTCTGCGAG
            CCTGCGTGCG ATTCTCAGAT TCTTCGCTAC TTCTACAATG CAACCTCTCA CACCTGCGAG
            CCTGTGTGCA GTTCTCAAAT TCTTCACTAT TTCTACAACA CAACCTCTCG AACCTGTGAG
            CCTCAGTGTG ATAATAACAT TCGTCGCTAC TACTACAACA CAACCTCTCG CATCTGCGAG
            CCCGTCTGCC GAGGCCGTTT CTACCGCTAT TTCTACAACT CTACAGCCCA TGAATGTGAG
            CCAGGGTGCA GAGGCCGCTT CTACCGGTAC TTCTACAACT CTACATCCAG TGAGTGTGAG
            CCAGCGTGCA GAGGTGTCTT CTCCCGGTAC TTCTACAACG ACACGTCCAG TGAGTGTGAG
            CCAGCGTGCA GAGGCCTCTT CTACCGGTAC TTCTCCAACT CTACGTCCAG TGAGTGTGAG
            CCAGCGTGCA GAGGCGTCTT CTCCCGGTAC TTCTACAACG ACACGTCCAG GGAGTGTGAG
            CCAGAGTGCA AAGCCTTCCT GCGCCGGTAC TCCTACAACT CGACGTCCAG GGAGTGTGAG
            CAAGCATGCA AAGCTGCTTT GCTCCGATAC TTCTATAACT CTACATCCAA TGCTTGTGAG
            CCAGAGTGCA AAGCCTTCCT GCGCCGGTAC TCCTACAACT CTACATCCAG TGAGTGTGAG
            CAAGCATGCA AAGCTGCTTT GCTCCGATAC TTCTATAACT CTACATCCAG TGCTTGTGAG
            GCAGTGTGCA AAGCCTCTTT GCACCGATAC TTCTACAACT CTACGTCCAT TGAGTGTGAG
            CAGGCATGCA AAGCTGCTTT GCTCCGATAC TTCCATAACT CGACATCCAA TGCTTGCGAG
            GCAGTGTGCA AAGCCTCTTT GCACCGATAC TTCTACAACT CTACGTCCAT TGAGTGTGAG
            CCAGTGTGCA AAGCCTCTTT GCACCGATAC TTCTACAACT CTACATCCAT TGAGTGTGAG
            CCAGTGTGCA AAGCCTCTTT GCGCCGATAC TTCTACAACT CTACGTCCGC TGAGTGTGAG
            CCAGTGTGCA AAGGCTATTT CCACCCATAC TTCTACAATT CAACGTCCAA TGAGTGTGAG
            CAAGCATGCA AAGCTGCTTT GCTCCGATAC TTCTATAACT CGACATCCAA TGCTTGTGAG
            CAAGCATGCA AAGCTGCTTT GCTCCGATAC TTCTATAACT CTACATCCAA TGCTTGTGAG
            CAAGCATGCA AAGCTGCTTT GCTCCGATAC TTCTATAACT CTACATCCAA TGCTTGTGAG
            CCAGTGTGCA GAGGCCTCTT CTACCGGTAC TTCTACAACT CTACGTCCAG TGAGTGTGAG
            CCCAAGTGCA GAAGTGACTG GAAGAAATGT TGCCTCGACA CTTGCGGAAC TGAATGTCTG
```

```
GAGTTCATCT ACACTGGCTG TAATGGCAAT GGGAACAATT TTGACTCAGT AGAATGCTGT
GAGTTCATCT ACACTGGCTG TAATGGCAAT GGGAACAATT TTGACTCAGT AGAATGCTGT
GCTTTCATCT ACAGTGGCTG TCATGGCAAT GGGAACAATT TTGACTCGCT GCAATGCTGC
GCTTTCATCT ACAGTGGCTG TAATGGCAAT GGAAACAATT TTGACACAGT AGAATGCTGC
GTTTTCCTCT ACAGTGGCTG TAACGGCAAT GGGAACAATT TTGACTCGCT GGAATGCTGC
ACTTTCATCT ACAGTGGCTG CAATGGCAAT AGGAACAATT TTAACTCGGA GGAATACTGC
GAGTTCATCT ACACTGGCTG TAATGGCAAT GGGAACAATT TTGACTCAGT AGAATGCTGT
CACTTCACCT ACGGTGGCTG TCGGGGCAAT GCCAACAACT TTGAGACCAC AGAGATGTGT
CACTTCATCT ATGGTGGCTG TCAGGGCAAT GCCAACAATT TTGAGACCAC AGAGATCTGT
CACTTTGCCT ACGGTGGCTG TCAGGGCAAT GCCAACAATT TCGAGACCAC GGAGATCTGT
CACTTTACCT ACGGTGGCTG TCAGGGCAAT GCCAACGACT TTGAGACCAC AGAGATCTGT
CACTTTGCCT ACAGTGGCTG TCAGGGCAAT GCCAACAATT TCGACACCAT GGAGATCTGT
TGCTTCATCT ATGGTGGCTG TCAGGGCAAT GCCAACAATT TCGAGACCAT GGAGATCTGT
CCCTTTACCT ATGGTGGTTG TCAGGGCAAT GACAACAATT TTGAGACCAC AGAGATGTGT
TGCTTCATCT ATGGTGGCTG TCAGGGCAAT GCCAACAATT TTGAGACCAT GGAGATCTGT
CCCTTTACCT ATGGTGGTTG TCAGGGCAAT GACAACAATT TTGAAACCAC AGAGATGTGT
CCCTTTACCT ATGGTGGTTG TCAGGGCAAT GCCAACAATT TTGAGACCAC AGAGATCTGT
CCCTTTACCT ATGGTGGTTG TCAGGGCAAT GACAACAATT TTGACACCAC AGAGATGTGT
CCCTTTACCT ATGGTGGTTG TCAGGGCAAT GCCAACAATT TTGAGACCAC AGAGATCTGT
CCCTTTACCT ATGGTGGTTG TCAGGGCAAT GCCAACAATT TTGAGACCAC AGAGATCTGT
CTCTTTATGT ACGGTGGTTG TCAGGGAAAC GCCAACAATT TTGAGACCAC AGCGATCTGT
CACTTTATCT ACGGTGGTTG TCAGGGCAAT GCCAACAATT TTGAGACCAT AGAGATCTGC
CCCTTTACCT ATGGTGGTTG TCAGGGCAAT GACAACAATT TTGAGACCAC AGAGATGTGT
CCCTTTTCCT ATGGTGGTTG TCAGGGCAAT GACAACAATT TTGAGACCAC AGAGATGTGT
CCCTTTACCT ATGGCGGTTG TCAGGGCAAT GACAACAATT TTGAAACCAC AGAGATGTGT
CACTTTACCT ACGGTGGCTG TCAGGGCAAT GCCAACAATT TTGAGACCAC AGAGATCTGT
CCATTGGTCC ATGGCCACTG TGAGACAGAC GACCAGTGCG TAGGTACATT AAAGTGCTGC

TTGAAGACCT GCAAACTCAA C
TTGAAGACCT GCAAACTCAA C
TTGAAGACCT GTAGACCCAA C
TTGAAGACCT GCAGACCCAA C
TTGAAGACCT GCAGACTCAA C
TTAAAGACCT GCAGACGCAA C
TTGAAGACCT GCAAACTCAA C
CTGAAGGTCT GCAAACCCCC C
TTGAAGATCT GCAAGCCCCC T
TTGAGGATCT GCAAGCACCC T
TCGAGGATCT GCAAGCCCCC T
TTGAGGATCT GCAAGCACCC T
CTGAGGGTCT GCACACCCCC G
TTGAGGATCT GCCAACCCCC T
CTGAGGGTCT GCAAACCCCC G
TTGAGGATCT GCCAACCCCC T
GTGAGGGTCT GCAAACCCCC T
TTGAGGATCT GCCAACCCCC T
GTGAGGGTCT GCAAACCCCC T
TTGAGGGTCT GCAACCCCCC T
CGGAGGGTCT GCAACCCCCC T
TTGTGGGTCT GCAGACCCCC T
TTGAGGATCT GCCAACCCCC T
CTGAGGATCT GCCAACCCCC T
TTGAGGATCT GCCAACCCCC T
TTGAGGATCT GCAAGCCCCC T
AATGCTGTGT GTGGGAAAGT C
```


Gil, M., Zanetti, M.S., Zoller, S., and Anisimova, M. (2013). CodonPhyML: fast maximum likelihood phylogeny estimation under codon substitution models. *Molecular Biology and Evolution* **30**: 1270-1280.

Higdon, J.W., Bininda-Emonds, O.R., Beck, R.M., and Ferguson, S.H. (2007). Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evolutionary Biology* **7**: 216.

Pharo, E.A., De Leo, A.A., Renfree, M.B., Thomson, P.C., Lefevre, C.M., and Nicholas, K.R. (2012). The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene. *BMC Evolutionary Biology* **12**: 8.

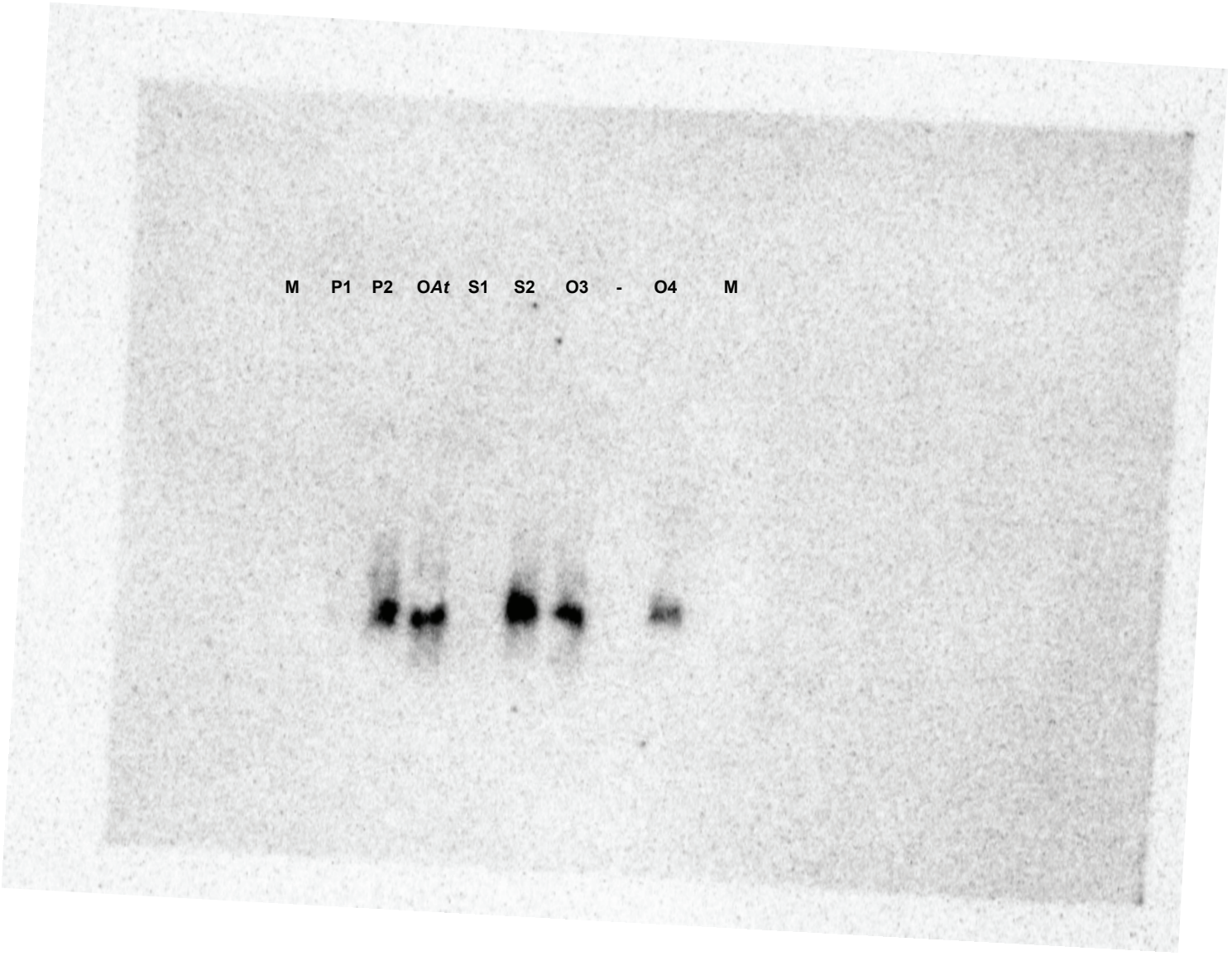
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology* **61**: 539-542.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725-2729.

Zhou, X., Xu, S., Xu, J., Chen, B., Zhou, K., and Yang, G. (2012). Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the laurasiatherian mammals. *Systematic Biology* **61**: 150-164.

Supplementary file 8 - Northern analysis of Cape and subantarctic fur seal mammary gland tissue probed with [α-32P] dCTP-labelled fur seal CTI cDNA.

Original phosphorimager scan of the membrane. M, RNA Marker; P1, pregnant primiparous Cape fur seal; P2, pregnant multiparous Cape fur seal; O*At*, on-shore lactating subantarctic fur seal; S1, at-sea, lactating Cape fur seal, S2, at-sea, lactating Cape fur seal; O3, on-shore lactating Cape fur seal; O4, on-shore lactating Cape fur seal.



Supplementary file 9 - Comparison of the putative physical characteristics of the mature BPTI, CTI and ELP peptides

BPTI is highlighted orange; Order Carnivora CTI pink; Order Cetartiodactyla CTI blue; rhinoceros CTI white and marsupial ELP (Orders Dasyuromorphia, Diprotodontia and Didelphimorphia) green.

Mature ELP/ CTI peptide	No. of amino acids	Mass (kDa)	Iso-electric point (pI)	Peptide charge at pH 7	No. and % of different classes of amino acid residues						Aliphatic Index	Grand Ave. hydropathicity (GRAVY)
					Non-polar (A, C, F, G, I, L, M, P, V, W, Y)	Polar (D, E, H, L, M, P, T, V, W)	Aliphatic (A, I, L, V)	Aromatic (F, H, Y, W)	Basic (K, R, H)	Acidic (D, E)		
bPTI	70	8.75	8.56	5.0	46 - 58.2 %	33 - 41.8 %	13 - 16.5 %	9 - 11.4 %	11 - 13.9 %	6 - 7.60%	71.6	-0.69
Rhinoceros	83	9.39	5.01	-3.0	43 - 51.8 %	40 - 48.2 %	14 - 16.9 %	13 - 13.3 %	8 - 9.64 %	10 - 12.0 %	49.4	-0.69
Cat	79	9.02	7.86	2.5	42 - 53.1 %	37 - 46.8 %	9 - 11.4 %	11 - 13.9 %	13 - 16.5 %	9 - 11.3 %	48.2	-0.54
Dog	79	8.99	7.84	2.0	41 - 51.9 %	38 - 48.1 %	10 - 12.7 %	11 - 13.9 %	11 - 13.9 %	8 - 10.1 %	53.2	-0.67
Panda	79	8.76	7.30	1.5	40 - 50.6 %	39 - 49.4 %	14 - 17.7 %	10 - 12.7 %	12 - 15.2 %	9 - 11.4 %	43.4	-0.72
P.Bear	79	8.91	8.32	1.5	42 - 53.2 %	37 - 46.8 %	15 - 19.0 %	12 - 15.2 %	13 - 16.5 %	7 - 8.86 %	50.4	-0.64
C. F. Seal	75	8.47	6.31	0.5	41 - 54.7 %	34 - 45.3 %	10 - 13.3 %	11 - 14.7 %	10 - 13.3 %	8 - 10.7 %	56.0	-0.54
Walrus	75	8.59	7.01	1.5	41 - 54.7 %	34 - 45.3 %	15 - 20.0 %	11 - 14.7 %	11 - 14.7 %	8 - 10.7 %	52.1	-0.64
Alpaca	83	9.17	7.72	1.0	41 - 49.4 %	42 - 50.6 %	16 - 19.3 %	6 - 7.23 %	9 - 10.8 %	8 - 9.64 %	51.8	-0.61
Antelope	82	9.13	4.89	-2.0	39 - 47.6 %	43 - 52.4 %	13 - 15.9 %	7 - 8.54 %	8 - 9.77 %	10 - 12.2 %	44.2	-0.87
W. Buffalo	82	9.07	4.41	-5.0	40 - 48.8 %	42 - 51.2 %	14 - 17.1 %	7 - 8.54 %	6 - 7.32 %	11 - 13.4 %	45.4	-0.80
Camel	83	9.13	7.72	1.0	41 - 49.4 %	42 - 50.6 %	16 - 19.3 %	6 - 7.23 %	9 - 10.8 %	8 - 9.64 %	51.8	-0.60
Cow	82	9.10	4.31	-6.0	39 - 47.6 %	43 - 52.4 %	8 - 9.76 %	7 - 8.53 %	6 - 7.31 %	12 - 14.6 %	44.2	-0.83
Dolphin	83	9.07	7.76	1.5	42 - 50.6 %	41 - 49.4 %	18 - 21.7 %	7 - 8.43 %	9 - 10.8 %	7 - 8.43 %	58.8	-0.53
Goat	82	9.09	5.15	-1.5	38 - 46.3 %	44 - 53.7 %	13 - 15.9 %	7 - 8.54 %	9 - 11.0 %	10 - 12.2 %	44.2	-0.89
Pig	83	9.02	8.62	3.0	46 - 55.4 %	37 - 44.6 %	18 - 21.7 %	7 - 8.43 %	9 - 10.8 %	6 - 7.23 %	56.5	-0.47
Sheep	82	9.13	4.89	-2.0	39 - 47.6 %	43 - 52.4 %	13 - 15.9 %	7 - 8.54 %	8 - 9.77 %	10 - 12.2 %	44.2	-0.87
K. Whale	83	9.10	8.30	2.5	42 - 50.6 %	41 - 49.4 %	18 - 21.7 %	7 - 8.43 %	10 - 12.0 %	7 - 8.43 %	58.8	-0.54
M. Whale	83	9.21	7.77	1.5	41 - 49.4 %	42 - 50.6 %	16 - 19.3 %	8 - 9.64 %	9 - 10.8 %	7 - 8.43 %	54.1	-0.63
Yak	82	9.13	4.31	-6.0	39 - 47.6 %	43 - 52.4 %	13 - 15.9 %	7 - 8.54 %	6 - 7.31 %	12 - 14.6 %	44.2	-0.86
Dunnart FT	76	8.70	6.41	0.0	37 - 48.7%	39 - 51.3 %	14 - 18.4 %	6 - 7.90 %	8 - 10.5 %	8 - 10.5 %	70.5	-0.63
Dunnart SF	76	8.67	6.41	0.0	37 48.7%	39 - 51.3 %	15 - 19.7 %	6 - 7.90 %	8 - 10.5 %	8 - 10.5 %	74.3	-0.60
Koala	81	9.17	8.27	3.0	36 - 44.4 %	45 - 55.6 %	12 - 14.8 %	9 - 11.1 %	10 - 12.3 %	6 - 7.40%	57.8	-0.79
Opossum	76	8.56	4.48	-4.0	41 - 53.9 %	35 - 46.0 %	16 - 21.1 %	6 - 7.90 %	5 - 6.58 %	9 - 11.8 %	78.2	-0.41
Possum	82	9.33	8.56	3.5	40 - 48.8 %	42 - 51.2 %	17 - 20.7 %	8 - 9.76 %	11 - 13.4 %	7 - 8.54 %	80.9	-0.48
Tas. Devil	84	9.68	4.63	-2.0	38 - 45.2 %	46 - 54.8 %	16 - 19.0 %	7 - 8.33 %	8 - 9.52 %	10 - 11.9 %	67.3	-0.89
Tammar	83	9.58	8.56	3.5	36 - 43.3 %	47 - 56.6 %	16 - 19.3 %	8 - 9.39 %	10 - 12.0 %	6 - 7.23 %	71.6	-0.69

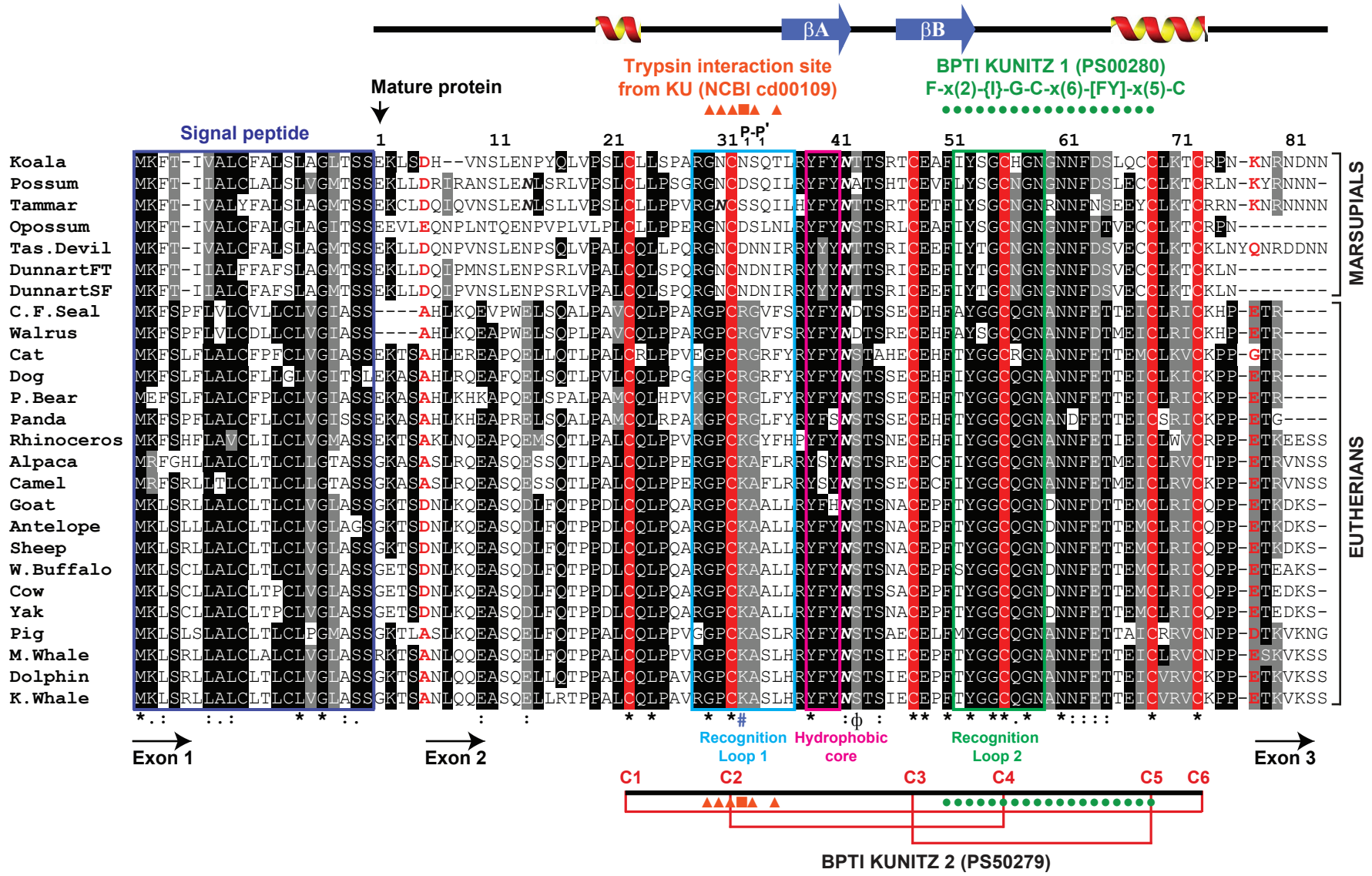
Supplementary file 10 - Percentage similarity and identity between the mature eutherian CTI, marsupial ELP and BPTI peptides

Pairwise similarities (lower left hand section) and identities (upper right hand section) were calculated with MatGATversion 2.03 for selected mature CTI, ELP and BPTI peptides (using the BLOSUM50 matrix). BPTI comparisons are highlighted orange, comparisons within Order Carnivora CTI pink, those within Order Cetartiodactyla blue; rhinoceros CTI white and those within marsupial ELP (Orders Dasyuromorphia, Diprotodontia and Didelphimorphia) green.

				BPTI	EUTHERIAN CTI																		MARSUPIAL ELP						
					Carnivora						Cetartiodactyla											Periss.	Dasyuromorphia			Diprotodontia			Didelph.
			Peptide	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	24
BPTI		1	BPTI		35	33.8	34.9	32.5	37.3	37.3	36.1	36.5	36.1	36.5	34.9	35.3	33.7	36.1	36.1	36.5	36.5	36.5	33.7	29.8	29.8	28.9	32.1	34.5	34.9	33.7
EUTHERIAN CTI	Carnivora	2	Cape fur seal	46.2		92	77.2	75.9	67.1	75.9	62.7	58.5	65.1	59.8	61.4	56.1	60.2	61.4	59	58.5	58.5	58.5	65.1	43	43	42.2	40.5	42.7	38.6	43
		3	Pacific walrus	43.8	93.3		72.2	70.9	64.6	72.2	63.9	56.1	63.9	56.1	59	56.1	57.8	59	56.6	56.1	56.1	56.1	63.9	46.8	46.8	44.6	44	45.1	41	48.1
		4	Polar bear	48.8	83.5	79.7		88.6	75.9	83.5	65.1	59.8	67.5	59.8	65.1	57.3	65.1	65.1	61.4	59.8	57.3	58.5	68.7	41.8	41.8	39.8	41.7	40.2	39.8	41.8
		5	Giant panda	47.5	82.3	78.5	92.4		70.9	79.7	61.4	58.5	63.9	58.5	62.7	57.3	61.4	61.4	60.2	58.5	56.1	57.3	65.1	38	38	37.3	39.3	37.8	33.7	41.5
		6	Cat	50	78.5	77.2	88.6	84.8		79.7	62.7	59.8	63.9	58.5	67.5	57.3	66.3	66.3	62.7	59.8	57.3	58.5	69.9	40.5	40.5	39.8	44	42.7	41	46.3
		7	Dog	55	81	78.5	89.9	87.3	88.6		72.3	63.4	74.7	63.4	67.5	61	66.3	67.5	63.9	63.4	62.2	62.2	73.5	44.3	44.3	43.4	44.6	45.1	43.4	46.3
	Cetartiodactyla	8	Alpaca	53	68.7	71.1	75.9	73.5	72.3	78.3		67.5	97.6	66.3	77.1	65.1	75.9	77.1	73.5	67.5	66.3	66.3	74.7	44.6	44.6	44.6	42.2	42.2	42.2	47.1
		9	Antelope	52.4	69.5	67.1	74.4	73.2	78	76.8	75.9		67.5	96.3	78.3	97.6	77.1	78.3	71.1	100	95.1	97.6	65.1	37.8	37.8	39.8	43.4	42.7	39.8	40.2
		10	Camel	53	71.1	71.1	78.3	75.9	73.5	80.7	97.6	78.3		67.5	78.3	65.1	77.1	77.1	73.5	67.5	66.3	66.3	74.7	44.6	44.6	44.6	41	42.2	41	46
		11	Cow	52.4	68.3	65.9	73.2	73.2	75.6	75.6	74.7	100	77.1		75.9	93.9	74.7	75.9	68.7	96.3	96.3	98.8	62.7	36.6	36.6	38.6	42.2	41.5	38.6	41.5
		12	Dolphin	50.6	71.1	69.9	79.5	75.9	83.1	79.5	81.9	88	83.1	88		75.9	98.8	92.8	80.7	78.3	74.7	75.9	74.7	37.3	37.3	38.6	37.3	37.3	38.6	37.3
		13	Goat	51.2	69.5	67.1	74.4	73.2	78	76.8	75.9	100	78.3	100	88		74.7	75.9	68.7	97.6	92.7	95.1	62.7	37.8	37.8	39.8	43.4	42.7	38.6	40.2
		14	Killer whale	49.4	71.1	69.9	79.5	75.9	83.1	79.5	81.9	88	83.1	88	100	88		91.6	79.5	77.1	73.5	74.7	73.5	38.6	38.6	39.1	38.1	38.6	38.6	40
		15	Minke whale	51.8	71.1	69.9	79.5	75.9	83.1	79.5	81.9	85.5	81.9	85.5	96.4	85.5	96.4		81.9	78.3	74.7	75.9	75.9	39.8	39.8	39.8	39.8	39.8	41	41.9
		16	Pig	50.6	67.5	65.1	72.3	71.1	75.9	74.7	79.5	79.5	80.7	80.7	86.7	79.5	86.7	88		71.1	68.7	68.7	67.5	38.6	38.6	38.6	38.6	38.6	38.6	37.3
		17	Sheep	52.4	69.5	67.1	74.4	73.2	78	76.8	75.9	100	78.3	100	88	100	88	85.5	79.5		95.1	97.6	74.7	37.8	37.8	39.8	43.4	42.7	39.8	40.2
		18	Water buffalo	53.7	69.5	67.1	73.2	73.2	76.8	75.6	73.5	98.8	75.9	98.8	88	98.8	88	85.5	79.5	98.8		97.6	72.3	36.6	36.6	37.3	41	41.5	38.6	41.5
		19	Yak	52.4	68.3	65.9	73.2	73.2	76.8	75.6	74.7	100	77.1	100	88	100	88	85.5	79.5	100	98.8		73.5	36.6	36.6	38.6	42.2	41.5	38.6	41.5
	Perissodactyla	20	White rhinoceros	48.2	73.5	73.5	81.9	78.3	83.1	83.1	79.5	74.7	81.9	73.5	80.7	74.7	80.7	80.7	74.7	65.1	63.9	63.9		41	41	42.2	45.2	41	43.4	48.8
MARSUPIAL ELP	Dasyuromorphia	21	Dunnart FT	43.8	56.6	59.2	57	54.4	55.7	59.5	61.4	53.7	60.2	53.7	53	53.7	53	54.2	53	53.7	53.7	53.7	56.6		98.7	84.3	63.9	67.1	62.7	64.5
		22	Dunnart SF	43.8	56.6	59.2	57	54.4	55.7	59.5	61.4	53.7	60.2	53.7	53	53.7	53	54.2	53	53.7	53.7	53.7	56.6	100		85.5	65.1	67.1	63.9	64.5
		23	Tasmanian Devil	43.4	53	56.6	54.2	51.8	51.8	56.6	59	55.4	59	54.2	51.8	55.4	53	51.8	51.8	55.4	53	54.2	56.6	89.2	89.2		79.5	81.9	73.5	77.1
	Diprotodontia	24	Koala	46.9	56.8	60.5	61.7	59.3	59.3	63	62.7	64.6	62.7	63.4	59	64.6	60.2	61.4	53	64.6	62.2	63.4	62.7	77.8	77.8	66.3		75.9	73.5	61.4
		25	Brushtail possum	51.2	54.9	58.5	57.3	53.7	57.3	62.2	62.7	64.6	63.9	62.2	59	64.6	59	59	56.6	64.6	62.2	63.4	62.7	76.8	76.8	67.5	84.1		77.1	73.2
		26	Tammar wallaby	48.2	49.4	55.4	56.6	50.6	55.4	59	59	57.8	59	56.6	56.6	59	56.6	57.8	53	57.8	55.4	56.6	59	71.1	71.1	63.9	80.7	81.9		67.5
	Didelphimorpha	27	Opossum	50	56.6	61.8	53.2	54.4	57	57	60.2	51.2	60.2	51.2	48.2	51.2	51.8	50.6	44.6	51.2	51.2	51.2	57.8	81.6	81.6	61.4	77.8	62.2	59	

Supplementary file 11 - ClustalW2 alignment of eutherian CTI and marsupial ELP precursor proteins

Notably, the basic P_1 reactive site residue (#, Lys33/Arg33) is conserved in all eutherian CTI peptides. Neither the P_1 residue identity nor type is conserved in marsupial ELP. Putative signal peptides are boxed and red text indicates residues that overlap splice sites. Residues common to all species are indicated (*) and those shared by at least 19 of 27 species shaded black using BoxShade. Grey residues indicate conserved residue type. The, the BPTI_KUNITZ_1 and 2 motifs are shown. The two putative serine protease recognition loops are boxed in blue and green respectively and the putative hydrophobic core is boxed in pink. Predicted N-glycosylation sites (Asn-Xaa-Ser/Thr) are shown in bold, italicised text. Putative serine/threonine phosphorylation sites common to most species are indicated (ϕ) (Ser43/Thr43, alignment numbering). Conservation between groups of amino acids with strongly similar properties [> 0.5 score, Gonnet PAM 250 matrix (:)] and those with weakly similar properties [scoring < 0.5 in the Gonnet PAM 250 matrix) noted (.)]. Gaps within the alignment are indicated (−). A gap was added (position 77) to all sequences apart from Tasmanian devil ELP to assist with the alignment of the codon spanning the exon 2/exon 3 boundary.



Supplementary file 12 - Comparison of the predicted motifs within the mature BPTI, CTI and ELP peptides

BPTI is highlighted orange; Order Carnivora CTI pink; Order Cetartiodactyla CTI blue; rhinoceros CTI white and marsupial ELP (Orders Dasyuromorphia, Diprotodontia and Didelphimorphia) green.

Mature ELP/ CTI peptide	BPTI KUNITZ 1 (PS00280)	BPTI KUNITZ 2 (PS50279)	N-glycosylation (PS00001)	Protein kinase C phosphorylation (PS00005)	Tyrosine phosphorylation (PS00007)	Casein kinase II phosphorylation (PS00006)
bPTI	47-65 FvyGGCrakrnnFkssedC	19-69, 28-52, 44-65	-	-	-	61-64 SseD
Rhinoceros	51-69 FiyGGCqgnannFetieiC	23-73, 32-56, 48-69	42-45 NSTS	4-6 SaK	-	44-47 TsnE, 78-81 TkeE
Cat	51-69 FtyGGCrgnannFettemC	23-73, 32-56, 48-69	42-45 NSTA	-	-	44-47 TahE
Dog	52-70 FiyGGCqgnannFetteiC	24-74, 33-57, 49-70	42-45 NSTS	-	-	44-47 TssE, 46-49 SecE
Panda	51-69 FtyGGCqgnandFetteiC	23-73, 32-56, 48-69	42-45 NSTS	-	-	44-47 TssE, 46-49 SecE
P.Bear	51-69 FtyGGCqgnannFetteiC	23-73, 32-56, 48-69	42-45 NSTS			44-47 TssE, 46-49 SecE
C. F. Seal	47-65 FayGGCqgnannFetteiC	19-69, 28-52, 44-65	38-41 NDTS	-	-	40-43 TssE, 42-45 SecE
Walrus	47-65 FaySGCqgnannFdtmeiC	19-69, 28-52, 44-65	38-41 NDTS	40-42 TsR	42-49 Rec.EhfaY	40-43 TsrE
Alpaca	51-69 FiyGGCqgnannFetmeiC	23-73, 32-56, 48-69	42-45 NSTS	6-8 SIR	46-53 Rec.EcfiY	44-47 TsrE,
Antelope	51-69 FtyGGCqgndnnFettemC	23-73, 32-56, 48-69	42-45 NSTS	-	-	18-21 TppD, 74-77 TppE
W. Buffalo	51-69 FsyGGCqgndnnFettemC	23-73, 32-56, 48-69	42-45 NSTS	-	-	18-21 TppD
Camel	51-69 FiyGGCqgnannFetmeiC	23-73, 32-56, 48-69	42-45 NSTS	6-8 SIR	-	44-47 TssE, 46-49 SecE
Cow	51-69 FtyGGCqgndnnFettemC	23-73, 32-56, 48-69	42-45 NSTS	-	-	-
Dolphin	51-69 FtyGGCqgnannFetteiC	23-73, 32-56, 48-69	42-45 NSTS	-	-	44-47 TsiE
Goat	51-69 FtyGGCqgndnnFdttemC	23-73, 32-56, 48-69	42-45 NSTS	-	-	18-21 TppD
Pig	51-69 FmyGGCqgnannFettaiC	23-73, 32-56, 48-69	42-45 NSTS	6-8 SIK, 35-37 SIR	-	44-47 TsaE
Sheep	51-69 FtyGGCqgndnnFettemC	23-73, 32-56, 48-69	42-45 NSTS	-	-	18-21 TppD
K. Whale	51-69 FtyGGCqgnannFetteiC	23-73, 32-56, 48-69	42-45 NSTS	-	-	44-47 TsiE
M. Whale	51-69 FtyGGCqgnannFetteiC	23-73, 32-56, 48-69	42-45 NSTS	-	-	-
Yak	51-69 FtyGGCqgndnnFettemC	23-73, 32-56, 48-69	42-45 NSTS	-	-	18-21 TppD
Dunnart FT	51-69 FiyTGCngngnnFdsvecC	23-73, 32-56, 48-69*	42-45 NTTS	44-46 TsR, 72-74 TcK	46-53 RiceEfi.Y	-
Dunnart SF	51-69 FiyTGCngngnnFdsvecC	23-73, 32-56, 48-69*	42-45 NTTS	44-46 TsR, 72-74 TcK	46-53 RiceEfi.Y	-
Koala	51-69 FiySGChgngnnFdslqcC	23-73, 32-56, 48-69*	42-45 NTTS	36-38 TIR, 44-46 TsR 72-74 TcR	46-53 Rtc.EafiY	-
Opossum	51-69 FiySGCngngnnFdtvecC	23-73, 32-56, 48-69*	42-45 NSTS	44-46 TsR, 72-74 TcR	46-53 Rlc.EafiY	-
Possum	51-69 FlySGCngngnnFdslecC	23-73, 32-56, 48-69*	14-17 NLSR, 42-45 NATS	27-29 SgR, 72-74 TcR	-	-
Tas. Devil	51-69 FiyTGCngngnnFdsvecC	23-73, 32-56, 48-69*	42-45 NTTS	44-46 TsR, 72-74 TcK	46-53 RiceEfi.Y	-
Tammar	51-69 FiySGCngnrnnFnseeyC	23-73, 32-56, 48-69	14-17 NLSL, 31-34 NCSS, 42-45 NTTS	44-46 TsR, 72-74 TcR	46-53 Rtc.EtfiY	-

Supplementary file 13 – Similarity and identity and between the active trypsin peptides of selected pinnipeds and other eutherians

Pairwise similarities (lower left hand corner) and identities (upper right hand corner) were calculated with MatGAT version 2.03 for active trypsin enzymes of selected species (using the BLOSUM50 matrix). Comparisons within Order Carnivora trypsin are pink, those within Order Cetartiodactyla blue and human trypsin, yellow.

	Active Trypsin	1	2	3	4	5	6	7	8
1	Walrus		89.7	84.8	79.8	74	72.2	69.5	68.3
2	Weddell Seal	93.7		89.2	82.5	77.1	74	71.7	71.4
3	Polar Bear	90.6	92.8		85.7	80.3	78.9	73.1	72.8
4	Dog	88.8	90.6	93.3		80.3	78.9	75.8	75
5	Pig	85.7	88.3	91	91		82.5	81.2	79
6	Cow	83.9	86.5	90.1	89.7	93.3		75.3	75
7	Killer Whale	83.4	84.8	85.7	86.1	89.2	87.4		71.4
8	Human	81.2	83	84.4	84.4	86.6	85.7	84.4	

Chapter Five:

Expression of the mammary gland-specific tammar wallaby *early lactation protein* gene is maintained *in vitro* in the absence of prolactin

Published in Molecular and Cellular Endocrinology: 1st November 2013

Elizabeth A Pharo (2014) Expression of the mammary gland-specific tammar wallaby *early lactation protein* gene is maintained *in vitro* in the absence of prolactin. *Molecular and Cellular Endocrinology*. **382:** 871-880.

Author's personal copy

Molecular and Cellular Endocrinology 382 (2014) 871–880

Contents lists available at ScienceDirect

Molecular and Cellular Endocrinology

journal homepage: www.elsevier.com/locate/mce

Expression of the mammary gland-specific tammar wallaby *early lactation protein* gene is maintained *in vitro* in the absence of prolactin

Elizabeth A. Pharo *

Department of Zoology, The University of Melbourne, Melbourne, Victoria 3010, Australia
Cooperative Research Centre for Innovative Dairy Products, Australia
Victorian Institute of Animal Science, Department of Primary Industries, Attwood, Victoria 3049, Australia

ARTICLE INFO

Article history:
Received 26 November 2012
Received in revised form 23 September 2013
Accepted 25 October 2013
Available online 1 November 2013

Keywords:
Marsupial
Hormone
Explant
Culture

ABSTRACT

Marsupial *ELP* (*early lactation protein*) and its eutherian orthologue, *CTI* (*colostrum trypsin inhibitor*) are expressed in the mammary gland only for the first 100 days postpartum (Phase 2A) in the tammar wallaby and during the bovine and canine colostrogenesis period 24–36 h postpartum respectively. The factors which regulate temporal *ELP* and *CTI* expression are unknown. A tammar mammary gland explant culture model was used to investigate *ELP* gene regulation during pregnancy and early- and mid-lactation (Phase 1, 2A and 2B respectively). Tammar *ELP* expression could only be manipulated in explants *in vitro* if the gene was already expressed *in vivo*. *ELP* expression was maximal in Phase 1 explants treated with lactogenic hormones (insulin, hydrocortisone and prolactin), but unlike *LGB* (*β-lactoglobulin*), *ELP* expression was maintained in insulin or insulin and hydrocortisone over a 12-day culture period. In contrast, *ELP* was down-regulated when cultured without hormones. *ELP* could not be induced in explants cultured from mid-lactation which suggested that transcriptional repressors may prevent *ELP* expression during this period.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The expression of the marsupial *early lactation protein* (*ELP*) gene and the orthologous eutherian *colostrum trypsin inhibitor* (*CTI*) gene (Pharo et al., 2012) is mammary gland-specific and restricted to early lactation in marsupials (De Leo et al., 2006; Demmer et al., 1998; Nicholas et al., 1997; Pharo et al., 2012; Piotte and Grigor, 1996; Simpson, 1998), and to the brief colostrogenesis period in the cow (Laskowski and Laskowski, 1951; Veselsky et al., 1978) and dog (Pharo et al., 2012). This period coincides with the absence of an acquired immune system in the young. However, not all eutherians have a functional *CTI* gene. While the Carnivora (dog, cat, giant panda) and Cetartiodactyla (cow, pig, common bottle-nosed dolphin) have a protein-coding *CTI* gene, it has become a pseudogene in humans and other primates, rodents, the horse, elephant, and sloth (Pharo et al., 2012).

The function(s) of ELP and CTI are unknown, but these small ~10–20 kDa N-glycosylated whey proteins (Joss et al., 2009; Laskowski and Laskowski, 1951; Piotte and Grigor, 1996; Simpson et al., 1998) have a single Kunitz/BPTI (bovine pancreatic trypsin inhibitor) domain indicative of serine protease inhibitor activity (Laskowski and Kato, 1980; Rawlings, 2010). Laskowski and Laskowski (1951) hypothesized that bovine CTI prevents the proteolysis of immunoglobulins transferred from mother to calf in colostrum, whilst Piotte and Grigor (1996) suggested that brushtail possum ELP protects the young against pathogens. Alternatively, the oligosaccharides attached to ELP and CTI may act as soluble receptor analogues for bacterial and viral pathogens, preventing their passage into the gastrointestinal tract of the young (Pharo et al., 2012).

The mechanisms that regulate the temporal expression of *ELP* and *CTI* are yet to be determined, but the induction of tammar wallaby (*Macropus eugenii*) *ELP* at parturition (Pharo et al., 2012) suggests that this gene, like most milk protein genes, responds to a lactogenic complex of hormones (insulin, hydrocortisone and prolactin) at parturition (Topper and Freeman, 1980). In contrast to eutherians (Neville et al., 2002; Pang and Hartmann, 2007), the

Abbreviations: CAL, concurrent asynchronous lactation; CSN1, α-casein; CSN2, β-casein; CTI, colostrum trypsin inhibitor; E_2, 17β-oestradiol; EGF, epidermal growth factor; ELP, early lactation protein; HC, hydrocortisone; GR, glucocorticoid receptor; I, insulin; *INS*, insulin gene; IGF, insulin-like growth factor; IR, insulin receptor; LALBA, α-lactabumin; LGB, β-lactoglobulin; LLP, late lactation protein; M199, Medium 199; MAPK, mitogen-activated protein kinase; MEC, mammary epithelial cell; NS, not significant; Ph., Phase; PI3K, (phosphatidylinositol-3-kinase)/AKT; pp, postpartum; preg, pregnancy; PRL, prolactin; PRLR, prolactin receptor; STAT, signal transducer and activator of transcription; T_3, triiodothyronine; T_4, thyroxine, tetraiodothyronine.

* Corresponding author. Present address: Department of Zoology, The University of Melbourne, Melbourne, Victoria 3010, Australia. Tel.: +61 3 8344 4376; fax: +61 3 8344 7909.

E-mail address: epharo@unimelb.edu.au

0303-7207/$ - see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.mce.2013.10.030

Author's personal copy

withdrawal of progesterone is unnecessary for tammar parturition and lactation to proceed (Hinds and Tyndale-Biscoe, 1982a; Renfree, 1994; Tyndale-Biscoe and Renfree, 1987; Tyndale-Biscoe et al., 1984; Ward and Renfree, 1984).

The tammar and eutherian milk protein genes including α-*casein* (*CSN1*), β-*casein* (*CSN2*), α-*lactalbumin* (*LALBA*) and β-*lactoglobulin* (*LGB*, a pseudogene in primates and rodents) are induced at parturition and expressed throughout lactation (Houdebine, 1999; Nicholas et al., 1997; Rudolph et al., 2003). However, tammar *ELP* (Pharo et al., 2012; Simpson et al., 1998), *whey acidic protein* (*WAP*) (Simpson et al., 2000), and *late lactation protein A* (*LLPA*) and *LLPB* (Trott et al., 2002) are asynchronously expressed at specific times during lactation, correlated with the various developmental milestones in the pouch young (Fig. 1), (Findlay and Renfree, 1984; Trott et al., 2003; Tyndale-Biscoe and Renfree, 1987).

The tammar has a short pregnancy of 26.5 days (Phase 1) and gives birth to an altricial ~440 mg young (Menzies et al., 2012; Renfree et al., 2011, 1989; Tyndale-Biscoe and Renfree, 1987). During the next ~300–350 days of its comparatively long lactation, the mother progressively alters milk production and composition to provide appropriate nutrition for the completion of organogenesis and the neurological and physiological development and growth of the young (Fig. 1) (Deane et al., 1990; Green, 1984; Green et al., 1980; Janssens et al., 1997; Tyndale-Biscoe and Janssens, 1988). Three phases of tammar lactation have been defined, Phase 2A (0–100 days postpartum (pp)), Phase 2B (100–200 days pp) and Phase 3 (200–350 days pp) (Brennan et al., 2007; Nicholas et al., 1997). After birth, the immunologically naive neonate (Edwards et al., 2012; Tyndale-Biscoe and Renfree, 1987) climbs into its mother's pouch where it remains permanently attached to one of the four teats for the next 100 days (Phase 2A, *ELP* gene expressed) (Pharo et al., 2012; Simpson, 1998). Lactation continues in the sucked gland, but ceases in the others (Findlay and Renfree, 1984; Stewart, 1984). From days 100–200 pp (Phase 2B, *WAP* expressed) (Simpson et al., 2000), the young is intermittently attached to the teat, organ development is completed and the young develops fur and homeothermy (Menzies et al., 2012; Tyndale-Biscoe and Renfree, 1987). Around 200 days pp, the young makes its initial exit from the pouch, analogous to birth in precocious eutherians (Menzies et al., 2012) and begins to eat herbage (Phase 3, *LLPA* and *LLPB* expressed) (Collet et al., 1989; Trott et al., 2002). Weaning occurs ~300–350 days pp and the mammary gland is remodelled in preparation for the next lactation cycle (Findlay, 1982a).

The endocrine regulation of some tammar milk protein genes has been characterised using an *in vitro* mammary gland explant culture model system (Brennan et al., 2008b; Simpson, 1998; Simpson et al., 2000; Trott, 1999), similar to those used to characterise milk protein gene regulation in eutherian species, such as the cow, goat, rabbit, guinea pig, rat and mouse (Burditt et al., 1981; Kulski et al., 1983; Puissant and Houdebine, 1991; Skarda et al., 1982). These 3D models are generally superior to primary and immortalised cell lines as the tissue-specific architecture, e.g. basoapical polarity of mammary epithelial cells (MECs) (Brennan et al., 2008b) and signalling between MECs and the extracellular matrix are preserved *in vitro* (Casey et al., 2000; Emerman et al., 1977; Lelievre et al., 1996).

Maximal *in vitro* expression of tammar *CSN1*, *CSN2* (Collet et al., 1992; Nicholas et al., 1995), *LGB* (Nicholas and Tyndale-Biscoe, 1985), *ELP* (Simpson, 1998), *WAP* (Simpson et al., 2000) and *LLPA* and *LLPB* (Trott et al., 2002; Trott, 1999) requires the synergistic

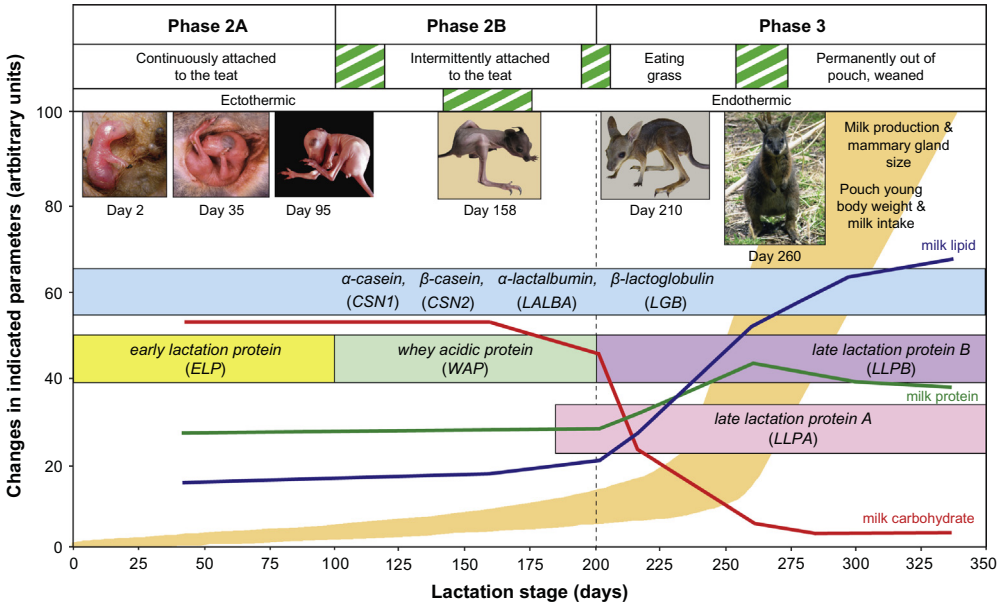


Fig. 1. Expression of the major milk protein genes and development of the pouch young during the reproductive cycle of the tammar wallaby. The tammar has a short pregnancy of ~26.5 days and gives birth to an altricial young. For the next 300–350 days the mother alters milk composition (protein, lipid and carbohydrate) and volume to provide for the growth and development of the pouch young (Findlay, 1982b; Green et al., 1983; Renfree et al., 1989; Shaw and Renfree, 2006; Tyndale-Biscoe and Renfree, 1987). Tammar lactation comprises three phases: Phase 2A, 2B and 3. Each phase is correlated with altered sucking patterns, changes in pouch young physiology and development (green and white striped boxes) (Tyndale-Biscoe and Janssens, 1988; Tyndale-Biscoe and Renfree, 1987) and changes in milk protein gene expression (Nicholas et al., 1997). The young spends the first 200 days in the pouch, continuously (0–100 days pp) and then intermittently attached to the teat (100–200 days pp). Phase 3 commences ~200 days pp and the fully developed young emerges from the pouch and starts to ingest herbage to supplement its milk intake. During this period, there is a dramatic increase in the body weight and milk intake of the young and in maternal mammary gland size and milk production (orange curve) (Findlay, 1982b). Permanent pouch exit occurs ~250 days pp and weaning ~300–350 days pp. As for eutherians, α-casein (*CSN1*), β-casein (*CSN2*), α-lactalbumin (*LALBA*) and β-lactoglobulin (*LGB2*) are induced at parturition and expressed throughout lactation. However, *ELP* is expressed during Phase 2A only, *WAP* during Phase 2B, and Late lactation proteins A and B (*LLPA*, *LLPB*) during late-Phase 2B-Phase 3 and Phase 3 respectively (Collet et al., 1989; Menzies and Nicholas, 2007; Nicholas et al., 1987, 1994, 1995, 1997; Nicholas, 1988a,b; Pharo et al., 2012; Simpson et al., 1998, 2000; Trott et al., 2002; Trott, 1999). Adapted from Green and Merchant (1988) and Green et al. (1983). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Author's personal copy

actions of the combination of "lactogenic" hormones, insulin (I), hydrocortisone (HC) and prolactin (PRL); as for the eutherian *CSN1*, *CSN2* and *WAP* genes (Lockwood et al., 1967; Rosen et al., 1999; Topper and Freeman, 1980). However, tammar *LGB* and *LALBA* are responsive to prolactin alone in explants from pregnancy, whilst *CSN1* and *CSN2* can be induced with either IPRL or HCPRL (Collet et al., 1992; Nicholas et al., 1995).

The phase-specific expression of tammar *ELP*, *WAP*, *LLPA* and *LLPB in vivo* suggests that their regulation is complex. The Phase 2B-specific *WAP* cannot be induced in Phase 3 (late-lactation) mammary explant tissue after the gene has been down-regulated *in vivo* (Simpson et al., 2000). However, *WAP* can be induced with IHCPRL, T_3 (triiodothyronine) and E_2 (17β-oestradiol) *in vitro* in Phase 1 (pregnancy) explants, overriding the repression of *WAP in vivo* (Simpson et al., 2000). Furthermore, the late-Phase 2B/Phase 3-specific *LLPA* cannot be induced with lactogenic hormones in mammary explants *in vitro* unless the gene is expressed *in vivo* prior to culture (Simpson, 1998; Simpson et al., 2000; Trott, 1999). Whether tammar *ELP* can be induced in all phases of lactation is unknown.

The major focus of this study was to investigate the endocrine regulation of tammar *ELP* during the different phases of the reproductive cycle using the tammar mammary explant culture system.

2. Materials and methods

2.1. Explant experiment design

Tammar mammary gland explant experiments were used to investigate both the maintenance and induction requirements of the *ELP*, *WAP* and *LGB* genes during different phases of the reproductive cycle (Table 1). These included: Phase 1 (day 24 of pregnancy, 6 animals); Phase 2A (days 76 and 80 of lactation, i.e., 76 and 80 days pp, 5 animals) and Phase 2B (∼day 150 of lactation, 4 animals).

2.2. Animals

Tammars were euthanised, the pouch cleaned with 70% ethanol and the lactating mammary gland removed using sterile techniques. All four mammary glands were excised from pregnant animals; except for animals for which a pouch young had been removed (RPY) approximately 24 days previously to induce reactivation of the quiescent blastocyst and hence the reactivation of pregnancy (Renfree and Tyndale-Biscoe, 1973). Three glands were removed from these animals. All experiments were approved by the Victorian Institute of Animal Science, Department of Primary Industries or The University of Melbourne Animal Ethics Committees in adherence to the guidelines of the National Health and Medical Research Council, Australia guidelines (NHMRC, 2004).

2.3. Preparation and culture of tammar mammary gland explants

Mammary gland explants (1–2 mg) were cultured on floating siliconised lens paper as described previously (Nicholas and Tyndale-Biscoe, 1985). Two wells were cultured per hormone treatment for each animal and final hormone concentrations (per mL of media) used for explant experiments are listed (Table 2). Explants were snap-frozen and stored at −80 °C for future analysis.

2.4. RNA extraction and Northern blot analysis

Total RNA was extracted from mammary explants with the Qiagen RNeasy Mini Kit (Qiagen, Hilden, Germany) and quantitated by spectrophotometer. Total RNA (10 μg) was electrophoresed through low-formaldehyde (1.1%) 1% agarose gels with 1X MOPS [3(N-Morpholino) Propane Sulphonic Acid] buffer at 4 °C. Nucleic acids were transferred overnight to Zeta-Probe GT Blotting Membrane (Bio-Rad, city, country) in 20X SSC (3.0 M sodium chloride, 0.3 M trisodium citrate, pH 7.0). Membranes were rinsed (2X SSC) and UV crosslinked at 1200 J (Stratagene UV Stratalinker1800) and pre-hybridised [30% deionised formamide, 5X SSC, 50 mM sodium acetate, herring sperm DNA (100 μg/μL), 5 mL Denhardt's 50X stock solution, 0.1% SDS] for 1 h at 42 °C. Blots were sequentially hybridised at 42 °C (from lowest to highest expression) with [α-^{32}P] dCTP labelled [DECAprime II Random Priming DNA Labelling Kit (Ambion, Invitrogen, Carlsbad, CA, USA)] cDNA from tammar *ELP* (Simpson et al., 1998), *WAP* (Simpson et al., 2000) and *Rsa*I digested *LGB* to ensure both *LGB* transcripts were detected (Collet et al., 1991). Membranes were washed twice (0.1X SSC, 0.1% SDS) for 15 min at 60 °C, exposed to a General Purpose Storage Phosphor screen and scanned on a Typhoon 8600 Scanner (GE Healthcare, Pittsburgh, PA, USA). Probes were stripped from membranes by incubation with boiling (100 °C) 1X SSC, 0.1% SDS on a shaking platform for two 15 min periods, then rinsed with RT 1X SSC, 0.1% SDS. Gene expression levels were quantitated by PhosphorImager analysis with ImageQuant software (GE Healthcare) and then normalised to RNA loading with Image J (Schneider et al., 2012).

2.5. Statistical analyses

Statistical analyses were performed using the statistics software packages R 3.0.1 (Team, 2013), JGR 1.7-14 (Helbig and Urbanek, 2013) and Deducer 0.7-6 (Fellows, 2012). For each explant experiment, gene expression data were normalised to RNA loading, log transformed and analysed using a one-way ANOVA (Generalised

Table 1
Summary of tammar mammary gland explant experiments used to investigate hormone(s) required for maintenance and induction of the ELP, WAP and LGB genes at different reproductive stages.

Reproductive stage	Culture experiment design	Culture duration	Number of tammars
Phase 1 (day 24 of pregnancy)	(i) Induce *ELP* and *LGB* gene expression over a 4, 8, and	4, 8 days	3
	(ii) A 12 day period	12 days	3
Phase 2A (76 days pp)	(iii) Maintain milk protein gene expression (*ELP*) over a 4-day culture period	4 days	2
Phase 2A (80 days pp)	(iv) Down-regulate milk protein gene expression with a 4-day insulin treatment and then induce expression (*ELP*, *WAP* and *LGB*) over a 10-day period	14 days	3
Phase 2B (150 days pp)	(v) Maintain milk protein gene expression (*WAP* and *LGB*) and induce *ELP* over a 4-day culture period	4 days	4
	(vi) Down-regulate *WAP* and *LGB* expression by treatment of explants with insulin over 4 days and then induce *ELP*, *WAP* and *LGB* over a 4-day period	8 days	4

Author's personal copy

Table 2
Hormone concentrations used for tammar mammary gland explant culture.

Hormone	Concentration (/mL media)	Supplier
Ovine prolactin (PRL)	200 ng/mL	#NIDDK-oPRL-21 (AFP-10692C), NIH Hormone and Peptide Program, Torrence, CA, USA
Porcine insulin (I)	200 ng/mL	#YQ0193AMQ Eli Lilly & Co., Indianapolis, USA
Hydrocortisone (HC)	50 ng/mL	#H-4001, Sigma, Saint Louis, MD, USA
17β-Oestradiol (E_2)	1 pg/mL	#E-2758, Sigma, Saint Louis, MD, USA
Triiodothyronine (T_3)	650 pg/mL	#T-2877, Sigma, Saint Louis, MD, USA
Murine epidermal growth factor (EGF)	10 ng/mL	#354001, BD Biosciences, Bedford, MA, USA
Human insulin-like growth factor I (IGF-I)	10 ng/mL	#PHG0074, Invitrogen, Carlsbad, CA, USA
Progesterone	5 ng/mL	#P-0130, Sigma, Saint Louis, MD, USA

Linear Model). A value of $p < 0.05$ was considered statistically significant. Post-hoc analysis with Tukey's multiple comparison of means and confidence interval estimates (95% family-wise confidence level) were used to determine treatments for which mean gene expression (log transformed data) levels differed.

3. Results

3.1. Endocrine regulation of the tammar ELP gene in mammary gland explants

The *in vitro ELP* response to endocrine stimulation was dependent upon the reproductive stage from which the mammary tissue explants were cultured (Figs. 2–5 inclusive).

3.1.1. Phase 1 (pregnancy)

Mammary explants from day 24 of pregnancy (Phase 1) were cultured for 4 and 8-day periods with various hormones (Fig. 2). Tammar *ELP* expression increased in the IHCPRL and IHCPRLT_3E_2 8-day treatments compared to t_0 (Day 0) tissue (filled dotted rectangle) ($p < 0.05$) (Fig. 2A and C(i)) in agreement with a preliminary study (Simpson, 1998). Similarly, expression of the prolactin-dependent *LGB* gene (Collet et al., 1991) was maximal in explants treated with either IHCPRL or IHCPRLT_3E_2 after 8 days (Fig. 2B and C(ii)), but *LGB* expression was not significantly different to t_0 tissue levels ($p > 0.05$). Unlike *LGB*, *ELP* expression was maintained in the absence of prolactin in the 8-day I and IHC treatments. In contrast, no *LGB* transcripts were detected by Northern analysis in explants supplemented with insulin and hydrocortisone (IHC) or with insulin (I) alone (Fig. 2C(ii)).

Extension of the culture period to 12 days produced similar results, with *ELP* transcripts present in explants treated with I, IHC and IHCPRL (Fig. 3A(i)). In contrast, *LGB* transcripts were present only in the IHCPRL-treated explants (Fig. 3A(iii)), whilst the Phase 2B-specific *WAP* was not detected (Fig. 3A(ii)). To confirm that *ELP* was responsive to I and IHC, explants were cultured for 5 or 10 days without hormones (Fig. 3B). Neither *ELP* nor *LGB* were detected in the 5- and 10-day no hormone (NH) treatments (Fig. 3B(i) and (ii)). However, both *LGB* (Brennan et al., 2008b) and *ELP* (data not shown) were up-regulated by additional 3-day treatments with IHCPRL after 10-day treatments in NH. This result confirmed that in Phase 1 explants *in vitro* the *ELP* gene was responsive to either insulin or insulin and hydrocortisone and that *ELP* expression was maintained in the absence of prolactin.

3.1.2. Phase 2A (early lactation)

In order to investigate whether *ELP* could be induced *in vitro*, Phase 2A explants were treated with insulin for 4 days to down-regulate milk protein gene expression. This formed the baseline for the experiment (filled dotted rectangle) (Fig. 4A and B). Explants were then treated for an additional 10 days with either I, IHC or IHCPRL. The high levels of both *ELP* and *LGB* expression in Phase 2A tissue (80 days pp) *in vivo* (t_0) could not be maintained in explants *in vitro*. with significant reductions in *ELP* and *LGB* transcripts in the 4-day I treatments ($p < 0.05$; Fig. 4A, C(i), B and C(ii)). However, when compared to the 4-day insulin baseline, Northern

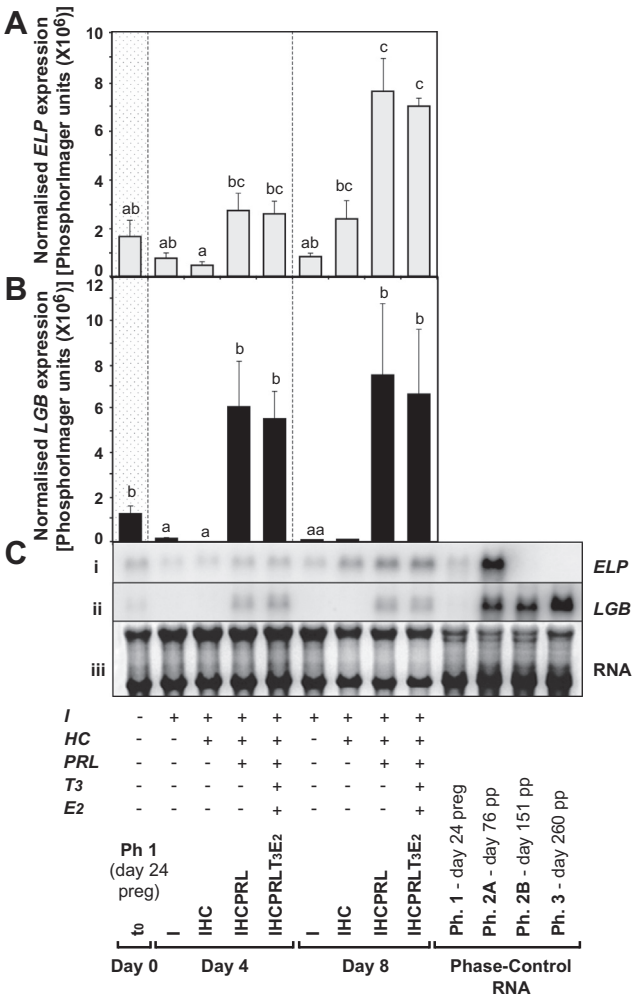


Fig. 2. Endocrine regulation of *ELP* and *LGB* in mammary explants cultured from late-pregnancy (Phase 1). (A–C) Explants from day 24 of pregnancy (t_0; Day 0) were cultured for 4 and 8 days in M199 supplemented with different hormones. (C) Northern analysis indicated that expression of (i) *ELP* and (ii) *LGB* was maximal after 8-day treatments with IHCPRL or IHCPRLT_3E_2. Phase-specific (Ph. 1, Ph. 2A, Ph. 2B, Ph. 3) control tissues are also shown, whilst (iii) ethidium bromide staining of ribosomal bands indicates equal loading and integrity of total RNA (10 μg). Total RNA gels for all samples are provided (Supplementary file 1). (A) PhosphorImager analysis showed that *ELP* expression was prolactin-independent, with expression levels for the 8-day I and IHC treatments similar to t_0 tissue (filled dotted rectangle). (B) In contrast, *LGB* expression was prolactin-dependent, with minimal *LGB* response in the 4- and 8-day I and IHC treatments. *ELP* and *LGB* gene expression levels (PhosphorImager units) were normalised to RNA loading and represent average gene expression ± SEM for 3 animals. Bars with different letters indicate treatments for which gene expression levels (log transformed data) differ significantly ($p < 0.05$) based upon post hoc analysis, Tukey's multiple comparison of means.

Author's personal copy

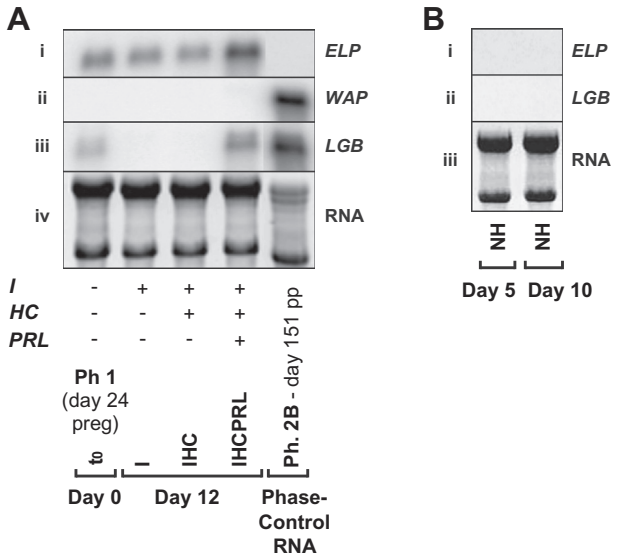


Fig. 3. Endocrine regulation of *ELP*, *WAP* and *LGB* in mammary explants cultured from late-pregnancy (Phase 1). (A) Explants from day 24 of pregnancy (t_0; Day 0) were cultured for 12 days in M199 supplemented with various hormones. Northern analysis showed that (i) *ELP* and (iii) *LGB* expression was maximal after a 12-day treatment with IHCPRL. However, unlike the prolactin-dependent *LGB* gene, *ELP* was expressed in explants supplemented with I and IHC after 12 days in culture. This was consistent for all three animals from which explants were cultured (data not shown). (ii) *WAP* transcripts were not detected in Phase 1 explants, but were present in the Phase 2B (151 days pp) positive control tissue. (iv) Ethidium bromide staining of ribosomal bands indicates equal loading and integrity of total RNA (10 μg). (B) Explants from day 24 of pregnancy were cultured for 5 and 10 days in M199 without hormones (NH). Northern analysis showed that neither (i) *ELP* nor (ii) *LGB* were expressed. (iii) Ethidium bromide stained total RNA (10 μg). Total RNA gels for all samples are provided (Supplementary file 2).

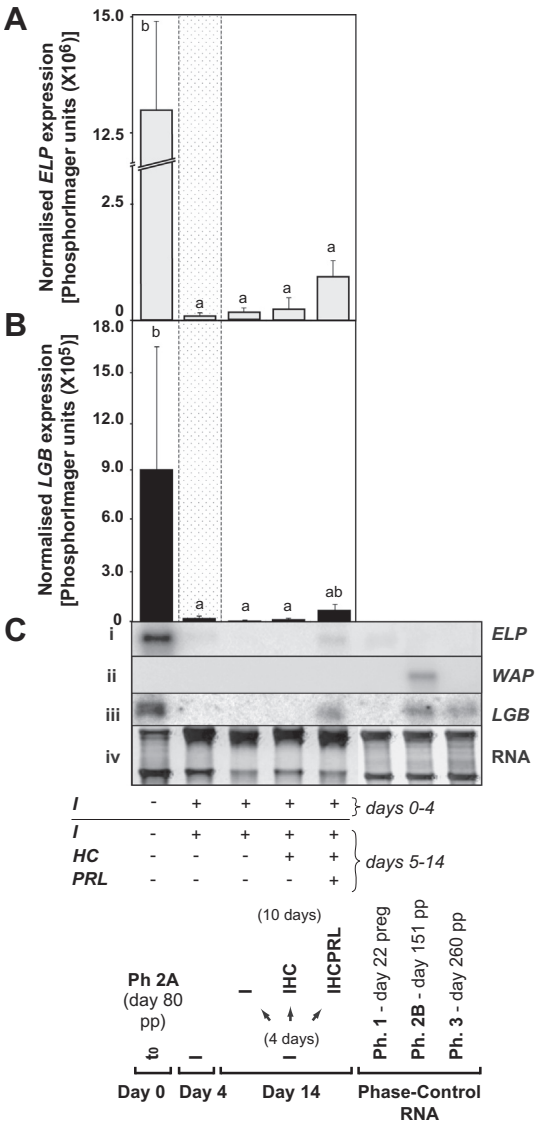


Fig. 4. Endocrine regulation of *ELP*, *WAP* and *LGB* in explants cultured from early lactation (Phase 2A) mammary tissue after down-regulation with insulin *in vitro*. (A–C) Explants from day 80 of lactation (t_0; Day 0) were cultured in M199 supplemented with insulin for 4 days to down-regulate milk protein gene expression (Day 4, experiment baseline, filled dotted rectangle). Explants were then treated with various hormones for an additional 10 days (Day 14). (C) Northern analysis showed that (i) *ELP* was down-regulated in the 4-day I treatment and then only detected in the subsequent IHCPRL 10-day treatment, at a level similar to Phase 1 control tissue. In contrast, *ELP* transcripts were not visible in the I or IHC treatments at Day 14. (ii) *WAP* transcripts were only visible in the Phase 2B control tissue, while (iii) *LGB* was down-regulated after 4 days in insulin and like *ELP*, *LGB* transcripts were only visible in the IHCPRL treatment at Day 14. (iv) Ethidium bromide staining of ribosomal bands indicate equal loading and integrity of total RNA (10 μg). (A) PhosphorImager analysis showed that *ELP* expression was similar for the I, IHC, and IHCPRL 10-day treatments when compared to the Day 4 I baseline (filled dotted rectangle; NS). (B) This was consistent for *LGB* expression (NS). *ELP* and *LGB* gene expression levels (PhosphorImager units) were normalised to RNA loading and represent average gene expression ± SEM for 3 animals. Bars with different letters indicate treatments for which gene expression levels (log transformed data) differ significantly ($p < 0.05$) based upon post hoc analysis, Tukey's multiple comparison of means. Total RNA gels for all samples are provided (Supplementary file 3).

analysis suggested *ELP* expression was higher in explants cultured with IHCPRL compared to day 4 levels (Fig. 4C(i)), similar to *ELP* levels in Phase 1 (day 22 of pregnancy) phase-control tissue, but this was not statistically significant (Fig. 4A). In contrast to Phase 1 explants, *ELP* transcripts were not visible by Northern analysis after 10-day treatments with I or IF (Fig. 4B(i)). *WAP* expression was not detected (Fig. 4C(ii)). Although *LGB* expression was only visible in IHCPRL-supplemented explants (Fig. 4C(iii)), there was no significant difference between hormone treatments ($p > 0.05$).

Like Phase 1 explants, Phase 2A explants cultured without hormones (NH) exhibited reduced *ELP* expression. *ELP* was minimal after 6 days in NH media and undetected after an 8-day NH treatment (personal communication, Amelia Brennan).

An additional Phase 2A (76 days pp) explant experiment in which tissues were treated with hormones for 4 days confirmed that the high levels of *ELP in vivo* could not be maintained *in vitro* (Supplementary file 4). Northern analysis showed that *ELP* expression was highest for the IHCPRL and IHCPRLT_3E_2 treatments, but minimal in I, IHC or PRL, consistent with Phase 1 explants. As for *ELP*, the high level of *LGB* expression *in vivo* was not maintained in culture (data not shown), whilst *WAP* transcripts were not detected. These experiments showed that Phase 2A mammary gland tissue retained its phenotype *in vitro*.

3.1.3. Phase 2B (mid-lactation)

Once *ELP* was down-regulated to minimal levels *in vivo* (~day 150 pp, Phase 2B); a very faint band detected by Northern analysis; the gene could not be induced *in vitro* after 4-day treatments with lactogenic hormones, unlike for *WAP* and *LGB* (Supplementary files 5 and 6). Two attempts to alter the phenotype of Phase 2B mammary explants *in vitro* and induce *ELP* expression were unsuccessful. Milk protein gene expression was down-regulated after a 4-day treatment with insulin (experiment baseline; filled dotted rectangle), but *ELP* could not be induced in any of the subsequent 4-day hormone treatments (Fig. 5C(i)). In contrast, *WAP* was up-reg-

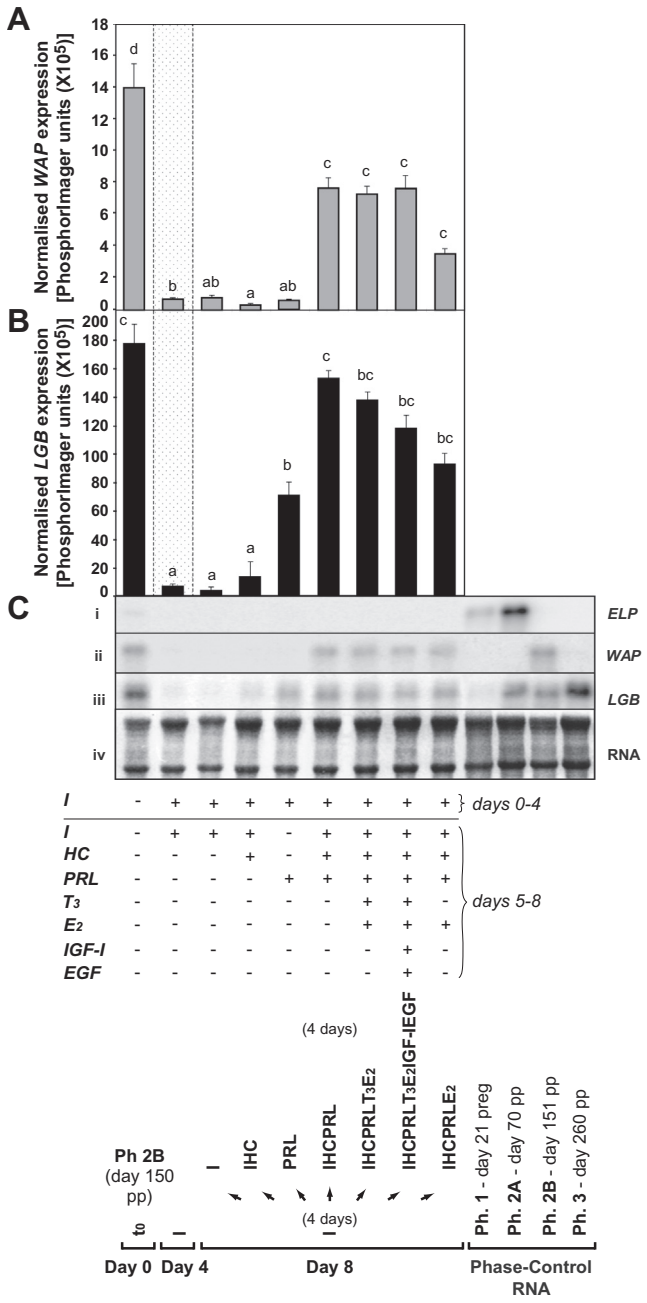


Fig. 5. Endocrine regulation of *ELP*, *WAP* and *LGB* in explants from mid-lactation (Phase 2B) mammary tissue after down-regulation with insulin *in vitro*. (A–C) Explants from ~day 150 of lactation (t_0; Day 0) were cultured in M199 supplemented with insulin for 4 days to down-regulate milk protein gene expression (Day 4, experiment baseline, filled dotted rectangle). Explants were then treated with various hormones for an additional 4 days (Day 8). (C) Northern analysis showed that (i) *ELP* was expressed at minimal levels (a very faint band) in Phase 2B tissue *in vivo*, at lower levels than for Phase 1 tissue. However, *ELP* expression was not detected in explants. (ii) *WAP* was down-regulated to minimal levels after the 4-day I treatment, but *WAP* was only induced in subsequent 4-day treatments which included IHCPRL. (iii) *LGB* was maximal at (t_0; Day 0), down-regulated in the 4-day I treatment and then induced in all treatments containing PRL. (iv) Ethidium bromide staining of ribosomal bands indicates equal loading and integrity of total RNA (10 μg). (A) PhosphorImager analysis showed that IHCPRL was essential for *WAP* expression in Phase 2B explants ($p < 0.05$). (B) *LGB* expression was higher in the IHCPRL-treated explants than for PRL alone ($p < 0.05$). *WAP* and *LGB* gene expression levels (PhosphorImager units) were normalised to RNA loading and represent average gene expression ± SEM for 4 animals. Bars with different letters indicate treatments for which gene expression levels (log transformed data) differ significantly ($p < 0.05$) based upon Tukey's multiple comparison of means. Total RNA gels for all samples are provided (Supplementary file 7).

ulated in all treatments which included IHCPRL, but was minimal in I, IHC or PRL alone ($p < 0.05$, Fig. 5A; 5C(ii)). *LGB* expression was also higher in the IHCPRL treatment than for PRL alone ($p < 0.05$, Fig. 5B and C(iii)).

A second attempt to 'reprogram' Phase 2B tammar explants to a Phase 1 (pregnancy) phenotype was also unsuccessful. Explants were treated for 6 days with hormones present *in vivo* during pregnancy – insulin, progesterone, 17β-oestradiol and EGF, which reduced *WAP* and *LGB* expression. However, whilst *WAP* and *LGB* were induced in all treatments containing prolactin, *ELP* was not expressed (data not shown). Therefore, reprogramming explants with exogenous hormones was unsuccessful because the explants retained their *in vivo* Phase 2B phenotype.

In summary, tammar mammary explants retained the phenotype of mammary tissue from which they were harvested. The hormonal requirements for basal expression and maximal induction of the asynchronously expressed *ELP* gene differed for Phase 1 and Phase 2A tissues based upon Northern analysis. *ELP* responded to I and IHC in explants from pregnancy with gene expression maintained *in vitro*, unlike for explants cultured without hormones (NH). In contrast to Phase 1 explants, the high level of Phase 2A *ELP* expression *in vivo* could not be maintained *in vitro*. *ELP* was down-regulated in early lactation explants by a 4-day insulin treatment, and then *ELP* expression only detected by Northern analysis in the 4-day IHCPRL treatment. *ELP* could not be induced in Phase 2B explants treated with lactogenic hormones. Furthermore, prolactin was essential for *LGB* induction in mammary explants, whilst expression of *ELP*, *WAP* and *LGB* was generally highest in IHCPRL treatments.

4. Discussion

The regulation of the tammar *ELP* gene is complex and involves hormones and factors intrinsic to the mammary gland. *ELP* expression could be manipulated by lactogenic hormones in mammary explants *in vitro* only if the gene was expressed *in vivo* (Phase 1 and Phase 2A; pregnancy and early lactation, respectively). *ELP* did not respond to exogenous hormonal cues *in vitro* after 100–120 days pp suggesting that other mechanisms may suppress *ELP* expression during mid- and late-lactation.

The maximal induction of *ELP* transcription *in vitro* in Phase 1 explants supplemented with IHCPRL over 4-, 8- and 12-day culture periods was consistent with a preliminary experiment (Simpson, 1998) and selected tammar (Collet et al., 1992; Nicholas et al., 1997, 1995; Nicholas and Tyndale-Biscoe, 1985; Simpson, 1998) and eutherian milk protein genes (Rosen et al., 1999; Topper and Freeman, 1980). The induction at parturition of tammar *ELP* and the milk protein genes expressed throughout lactation including *CSN1*, *CSN2*, *LGB*, and *LALBA* coincides with a transient pulse of prolactin, an increase in glucocorticoids (Hinds and Tyndale-Biscoe, 1982a; Renfree et al., 1979; Shaw et al., 1996) and a transient increase in expression of the insulin (*INS*) gene in the tammar mammary gland (Stringer et al., 2012). This suggests that lactogenic hormones play a role *in vivo* in the expression of these genes, consistent with explant experiment results for *CSN1*, *CSN2* (Collet et al., 1992; Nicholas et al., 1995), whilst *LALBA* and *LGB* expression is maximal in prolactin alone (Collet et al., 1991; Nicholas and Tyndale-Biscoe, 1985). However, the temporal expression of *ELP*, *WAP*, *LLPA* and *LLPB* suggests that more complex regulatory mechanisms control these genes.

The maintenance of *ELP* expression in Phase 1 explants supplemented with I or IHC was unexpected, particularly after a 12-day culture period. This was a novel result for a tammar milk protein gene and was validated by the lack of expression of the prolactin-responsive *LGB* gene in these treatments and the absence of

Author's personal copy

ELP transcripts in Phase 1 explants cultured without hormones over 5- and 10-day periods. However, milk protein gene induction in the absence of prolactin has been reported for a eutherian gene. Nicholas and Topper (1980) found that rat *LALBA* was induced in mammary explants from pregnant rats after treatment with insulin and hydrocortisone using lactose synthetase activity to measure *LALBA*. Warner et al. (1993) also induced murine *LALBA* in explants from pregnant mice in the absence of prolactin. In contrast, *LALBA* expression decreased in mammary explants from a lactating mouse treated with insulin and hydrocortisone (Perry and Oka, 1984).

The inability to maintain *ELP* and *LGB* gene expression *in vitro* in Phase 2A mammary explants, even when tissues were supplemented with IHCPRL was unexpected. However, this may be explained by differences in the *in vivo* metabolic activity of the tissues cultured (Nicholas and Tyndale-Biscoe, 1985; Sheehy et al., 2004). Tammar milk protein gene expression *in vivo* is low during pregnancy, but at parturition, there are significant increases in milk protein gene transcription, e.g. *ELP* and *LGB*, as the mammary gland produces milk for the young (Nicholas et al., 1997; Pharo et al., 2012). Therefore, the higher *ELP* and *LGB* expression levels and increased metabolic activity of Phase 2A versus Phase 1 mammary tissue *in vivo* could not be maintained in explants cultured *in vitro*, particularly after a 14-day culture period in insulin alone. Unlike Phase 1 mammary explants (Simpson et al., 2000), *WAP* was not induced in Phase 2A tissues cultured with IHCPRLT_3-E_2, so whilst these hormones can override the inhibition of *WAP in vitro* during pregnancy, this does not occur during early lactation.

For Phase 2B explants, *LGB* responded to PRL alone. In contrast, IHCPRL rather than PRL alone was essential for induction and maximal expression of *WAP in vitro*, consistent with previous Phase 2B explant studies (Simpson et al., 2000; Trott et al., 2005). However, the lack of hormonal induction of *ELP* in explants cultured from Phase 2B mammary tissues was intriguing. When *ELP* is down-regulated *in vivo* at the end of Phase 2A (~100 days pp), there are minimal changes in circulating maternal lactogenic hormones (Hinds, 1988; Hinds and Tyndale-Biscoe, 1982b). Therefore, transcription factors, non-coding RNA(s), changes in chromatin structure, and/or epigenetic mechanisms may suppress *ELP* expression during mid- and late-lactation. Factors that regulate bovine *CTI* have not been determined, but the switch from the secretion of colostrum to mature milk may be analogous to the change from Phase 2A to Phase 2B milk in the tammar, both of which coincide with the development of acquired immunity in the young (Edwards et al., 2012; Laskowski et al., 1952; Laskowski and Laskowski, 1951; Old and Deane, 2000; Tyndale-Biscoe and Renfree, 1987).

The failure to stimulate *in vitro* transcription of the Phase 2A-, 2B-, and 3-specific *ELP* (this study), *WAP* (this study; (Simpson et al., 2000) and *LLPA* (Trott et al., 2005) genes respectively in explants in which the genes are not expressed *in vivo*, with the exception of *WAP* transcription in Phase 1 explants (Simpson et al., 2000), suggests that tammar mammary explants retain the *in vivo* tissue phenotype from which they are harvested. This is consistent with the failure of tammar, bovine and murine mammary explants to enter involution when cultured without hormones *in vitro* (Brennan et al., 2008a,b). These explants retained their lactogenic phenotype and the ability to respond to endocrine stimuli. Brennan et al. (2008b) suggested that involution in explants was uncoupled by the lack of accumulation of luminal contents (milk), which removed the weaning and milk accumulation triggers that cause involution (Watson, 2006). As mammary explants no longer have the *in vivo* systemic feedback mechanisms, we speculate that this causes apparent phenotype stasis *in vitro*.

It is interesting to speculate that whilst explants permit the expression of milk protein genes *in vitro*, a phase-specific MEC architecture and/or chromatin conformation may be responsible for the asynchronous expression of the *ELP*, *WAP*, *LLPA* and *LLPB* genes and that these phase-specific phenotypes cannot be altered by endocrine stimuli alone. Extracellular matrix (ECM), or basement membrane is critical for the expression of eutherian milk protein genes *in vitro* (Aggeler et al., 1991; Barcellos-Hoff et al., 1989; Streuli et al., 1991), as is laminin (Blum et al., 1987; Streuli et al., 1995). In particular, the direct interaction between β1 integrin receptors in ECM and MECs is vital for the development and maintenance of basoapical polarity, correct 3D architecture of MECs and the ability to express milk protein genes (Li et al., 2005; Naylor et al., 2005). A recent *in vitro* study has shown that expression of the Phase 3-specific tammar *LLPB* gene is dependent upon the correct (Phase 3) ECM (Wanyonyi et al., 2013). However, *WAP* was expressed in IFPRL$T_3$$E_2$-treated cells cultured from the Phase 2B tammar mammary gland and plated on either Phase 2A or Phase 2B ECM. In contrast, *ELP* expression was low and similar for both IF- and IFPRL$T_3$$E_2$-treated cells cultured on Phase 2A, 2B and 3 ECM. Epigenetic modifications and global and local spatial organisation of nuclear chromatin may also influence milk protein gene transcription (Lemay et al., 2013; Rijnkels et al., 2013, 2010). Furthermore, growth hormone, growth factors, leptin, serotonin, cytokines, transcription factors and non-coding RNAs may also affect transcription of milk protein genes (Neville et al., 2002; Riley et al., 2010; Vonderhaar and Ziska, 1989).

The maximal response of *ELP* to insulin, hydrocortisone and prolactin suggests their respective signalling pathways are active in the tammar mammary gland. In MECs, prolactin signal transduction occurs via the Jak-STAT pathway (Hughes and Watson, 2012; Watson and Burdon, 1996). In the rodent mammary gland, prolactin binds to the long form of the PRLR, STAT5 dimers are translocated to the nucleus (Binart et al., 2010) and bind to palindromic TTCNNNGAA sequences in milk protein gene promoters (Gouilleux et al., 1994; Groner and Gouilleux, 1995; Wakao et al., 1994). Prolactin can also bind to a short form(s) of the PRLR and activate the MAPK and PI3K/AKT pathways (Binart et al., 2010). However, the prolactin signalling pathway(s) active in tammar MECs are yet to be characterised. Unlike prolactin, glucocorticoids enter MECs by diffusion, form activated glucocorticoid receptor dimers which enter the nucleus and bind to glucocorticoid response elements within gene promoters (Kino et al., 2003; Zhou and Cidlowski, 2005).

Insulin is also essential for milk protein gene transcription (Bolander et al., 1981; Menzies et al., 2010, 2009; Nicholas et al., 1982). During bovine lactation, insulin acts via the PI3K/AKT and Insulin Receptor Signalling pathways (Lemay et al., 2007). In murine MECs, insulin binds to the IR-B insulin receptor isoform and promotes *CSN2* and *WAP* transcription, but it binds neither IR-A, IGF-IR, nor IGF-IIR (Berlato and Doppler, 2009; Prosser et al., 1987). Insulin also has a vital role in tammar lactation and may act via the IR. Expression of the imprinted tammar *INS* gene is minimal during early lactation, but increases from ~100 days pp to the end of lactation (Phases 2B and 3) (Stringer et al., 2012). During late Phase 2B/Phase 3 there is a significant increase in tammar mammary gland size, milk production and milk protein content (Nicholas et al., 1997) which Stringer et al. (2012) suggest are promoted by the increase in *INS* expression. In contrast, *IGF2* transcription is high during pregnancy, but declines from day 9 pp (Stringer et al., 2012).

Thyroid hormones, thyroxine (T_4) and T_3, stimulate milk protein secretion and milk production by enhancing lobuloalveolar growth and the galactopoietic response of the mammary gland to prolactin and growth hormone (Capuco et al., 1999; Song and Oka, 2003; Topper and Freeman, 1980; Vonderhaar and Greco, 1979). Most importantly, the partitioning of metabolic activity is regulated by controlling the conversion of T_4 to the more-active T_3, reducing activity in peripheral tissues and increasing activity in the mam-

Author's personal copy

mary gland and directs maternal resources to the young (Jack et al., 1994). In contrast, the removal of E_2 may facilitate eutherian lactation (Neville et al., 2002). For these tammar explant studies, there was no significant difference between the IHCPRL and IHCPRLT_3E_2 or IHCPRLT_3E_2 and IHCPRLE_2 treatments for either *ELP*, *WAP* or *LGB* expression, suggesting that neither T_3 nor E_2 affect *in vitro* tammar milk protein gene expression, with the exception of *WAP* expression in Phase 1 explants (Simpson et al., 2000). However, thyroid hormones delivered via milk may be critical for the survival of the tammar pouch young which does not have an active thyroid gland and is incapable of thermoregulation until ~d180 pp (Gemmell and Sernia, 1992; Richardson et al., 2002).

During tammar lactation, changes in hormone sensitivity, and autocrine and paracrine factors (Hendry et al., 1998; Nicholas et al., 1995) may affect mammary gland milk protein gene expression. This is highlighted by the concurrent asynchronous lactation (CAL) phenomenon reported in marsupials such as the red kangaroo (Lemon and Bailey, 1966; Sharman and Pilton, 1964) and Agile wallaby (Lincoln and Renfree, 1981) and which can be experimentally induced in the tammar. CAL permits the marsupial mother to simultaneously produce milk of different compositions and different volumes from two different phases, for a newborn in the pouch from one gland (early lactation, Phase 2A) and for a young at foot from the other (late lactation, Phase 3) whilst maintaining a blastocyst *in utero*. Therefore, these two glands function independently of the other (Nicholas et al., 1995; Nicholas, 1988a) which is thought to be due to differences in mesotocin-sensitivity (the marsupial equivalent of oxytocin) between the two glands. Whether there are changes in the sensitivity of the tammar mammary gland to other hormones during pregnancy and throughout lactation is unknown. Furthermore, the CAL phenomenon suggests that the tammar lactation cycle may be 'programmed' by the ageing of the mammary gland so that the young receives nutrition appropriate to its needs throughout its development.

The inability to manipulate the temporal expression of *ELP* with hormones *in vitro* after Phase 2A *in vivo* suggests that expression of this putative protease inhibitor gene may be tightly-controlled and precisely coordinated.

5. Conclusions

The regulation of the mammary-gland specific tammar *ELP* gene is complex and involves both hormone-dependent and hormone-independent factors. Tammar *ELP* could be manipulated *in vitro* provided the gene was expressed *in vivo*. *ELP* expression was maximal in Phase 1 (pregnancy) explants supplemented with lactogenic hormones, insulin, hydrocortisone and prolactin, but *ELP* was also responsive to insulin and insulin and hydrocortisone treatments, contrary to other tammar milk protein genes. Furthermore, like *WAP*, once *ELP* was down-regulated *in vivo*, the gene was unresponsive to lactogenic hormones *in vitro*. Hence, transcriptional repressors may bind to the *ELP* promoter *in vivo* at the end of Phase 2A, ~100 days pp. Future studies will characterise the marsupial *ELP* and eutherian *CTI* promoters and identify transcription factors which activate and repress the *ELP* and *CTI* genes.

Acknowledgements

Jo Argento, Jenni Carfi, Kylie Cane, Amelia Brennan, Coralie Reich, Karensa Menzies and Kevin Nicholas provided assistance with experiment design and cutting explants. Scott Brownlees, Kerry Martin and Jenni Carfi assisted in animal handling. Amelia Brennan provided the Northern blots from the Phase 1 and Phase 2A no hormone explant cultures which were probed for tammar *ELP*. Keng Yih Chew and Professors Marilyn Renfree and Geoff Shaw provided tammar photos. Professors Marilyn Renfree and Kevin Nicholas provided helpful comments on this manuscript. This research was funded by the Cooperative Research Centre for Innovative Dairy Products (CRC-IDP), the Department of Primary Industries, Victoria and the Department of Zoology, The University of Melbourne. Elizabeth A Pharo was supported by an Australian Postgraduate Award and was the recipient of a scholarship top-up from the CRC-IDP.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.mce.2013.10.030.

References

Aggeler, J., Ward, J., Blackie, L.M., Barcellos-Hoff, M.H., Streuli, C.H., Bissell, M.J., 1991. Cytodifferentiation of mouse mammary epithelial cells cultured on a reconstituted basement membrane reveals striking similarities to development in vivo. J. Cell Sci. 99, 407–417.

Barcellos-Hoff, M.H., Aggeler, J., Ram, T.G., Bissell, M.J., 1989. Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. Development 105, 223–235.

Berlato, C., Doppler, W., 2009. Selective response to insulin versus insulin-like growth factor-I and -II and up-regulation of insulin receptor splice variant B in the differentiated mouse mammary epithelium. Endocrinology 150, 2924–2933.

Binart, N., Bachelot, A., Bouilly, J., 2010. Impact of prolactin receptor isoforms on reproduction. Trends Endocrinol. Metab. 21, 362–368.

Blum, J.L., Zeigler, M.E., Wicha, M.S., 1987. Regulation of rat mammary gene expression by extracellular matrix components. Exp. Cell Res. 173, 322–340.

Bolander Jr., F.F., Nicholas, K.R., Van Wyk, J.J., Topper, Y.J., 1981. Insulin is essential for accumulation of casein mRNA in mouse mammary epithelial cells. Proc. Natl. Acad. Sci. 78, 5682–5684.

Brennan, A.J., Sharp, J.A., Digby, M.R., Nicholas, K.R., 2007. The tammar wallaby: a model to examine endocrine and local control of lactation. IUBMB Life 59, 146–150.

Brennan, A.J., Sharp, J.A., Lefevre, C.M., Nicholas, K.R., 2008a. Uncoupling the mechanisms that facilitate cell survival in hormone-deprived bovine mammary explants. J. Mol. Endocrinol. 41, 103–116.

Brennan, A.J., Sharp, J.A., Khalil, E., Digby, M.R., Mailer, S.L., Lefevre, C.M., Nicholas, K.R., 2008b. A population of mammary epithelial cells do not require hormones or growth factors to survive. J. Endocrinol. 196, 483–496.

Burditt, L.J., Parker, D., Craig, R.K., Getova, T., Campbell, P.N., 1981. Differential expression of α-lactalbumin and casein genes during the onset of lactation in the guinea-pig mammary gland. Biochem. J. 194, 999–1006.

Capuco, A.V., Kahl, S., Jack, L.J., Bishop, J.O., Wallace, H., 1999. Prolactin and growth hormone stimulation of lactation in mice requires thyroid hormones. Proc. Soc. Exp. Biol. Med. 221, 345–351.

Casey, T.M., Boecker, A., Chiu, J.F., Plaut, K., 2000. Glucocorticoids maintain the extracellular matrix of differentiated mammary tissue during explant and whole organ culture. Proc. Soc. Exp. Biol. Med. 224, 76–86.

Collet, C., Joseph, R., Nicholas, K., 1989. Molecular cloning and characterization of a novel marsupial milk protein gene. Biochem. Biophys. Res. Commun. 164, 1380–1383.

Collet, C., Joseph, R., Nicholas, K., 1991. A marsupial β-lactoglobulin gene: characterization and prolactin-dependent expression. J. Mol. Endocrinol. 6, 9–16.

Collet, C., Joseph, R., Nicholas, K., 1992. Molecular characterization and *in-vitro* hormonal requirements for expression of two casein genes from a marsupial. J. Mol. Endocrinol. 8, 13–20.

De Leo, A.A., Lefevre, C., Topcic, D., Pharo, E., Cheng, J.F., Frappell, P., Westerman, M., Graves, J.A.M., Nicholas, K.R., 2006. Characterization of two whey protein genes in the Australian dasyurid marsupial, the stripe-faced dunnart (*Sminthopsis macroura*). Cytogenet. Genome Res. 115, 62–69.

Deane, E.M., Cooper, D.W., Renfree, M.B., 1990. Immunoglobulin G levels in fetal and newborn tammar wallabies (*Macropus eugenii*). Reprod. Fertil. Dev. 2, 369–375.

Demmer, J., Ross, I.K., Ginger, M.R., Piotte, C.K., Grigor, M.R., 1998. Differential expression of milk protein genes during lactation in the common brushtail possum (*Trichosurus vulpecula*). J. Mol. Endocrinol. 20, 37–44.

Edwards, M.J., Hinds, L.A., Deane, E.M., Deakin, J.E., 2012. A review of complementary mechanisms which protect the developing marsupial pouch young. Dev. Comp. Immunol. 37, 213–220.

Emerman, J.T., Enami, J., Pitelka, D.R., Nandi, S., 1977. Hormonal effects on intracellular and secreted casein in cultures of mouse mammary epithelial cells on floating collagen membranes. Proc. Natl. Acad. Sci. USA 74, 4466–4470.

Fellows, I., 2012. Deducer: a data analysis GUI for R. J. Stat. Softw. 49, 1–15.

Findlay, L., 1982a. The mammary glands of the tammar wallaby (*Macropus eugenii*) during pregnancy and lactation. J. Reprod. Fertil. 65, 59–66.

Findlay, L., 1982b. Lactation in the Marsupial, *Macropus eugenii*. PhD Thesis. School of Environmental and Life Sciences Perth, Murdoch University.

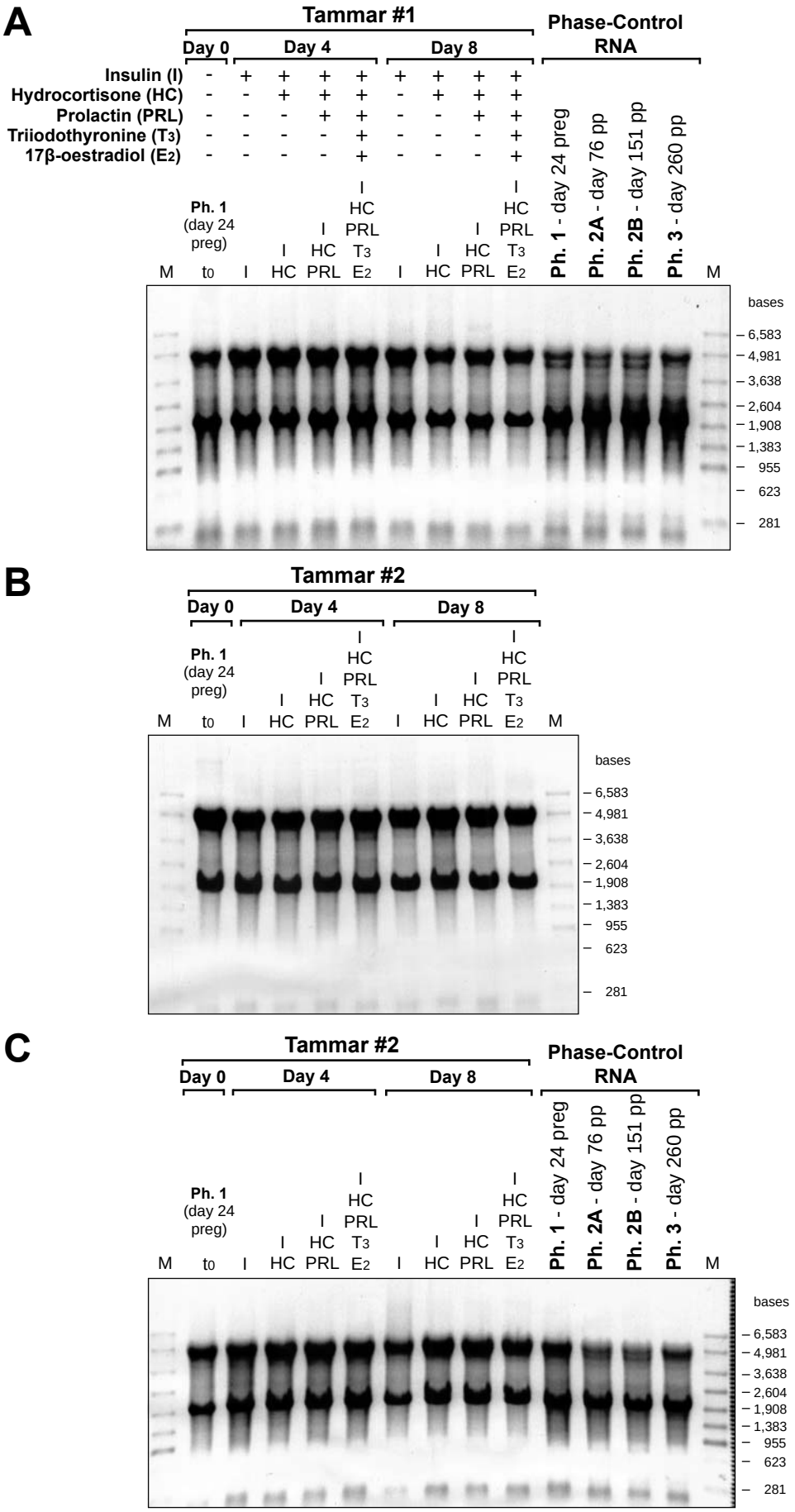
Author's personal copy

Findlay, L., Renfree, M.B., 1984. Growth, development and secretion of the mammary gland of macropodid marsupials. In: Peaker, M., Vernon, R.G., Knight, C.H. (Eds.), Symposia of the Zoological Society of London 51 Physiological Strategies in Lactation. Academic Press, London, pp. 403–432.
Gemmell, R.T., Sernia, C., 1992. The transfer of thyroxine from the mother to the young of the marsupials, the bandicoot (*Isoodon macrourus*) and the possum (*Trichosurus vulpecula*). Comp. Biochem. Physiol. Comp. Physiol. 103, 541–543.
Gouilleux, F., Wakao, H., Mundt, M., Groner, B., 1994. Prolactin induces phosphorylation of Tyr694 of Stat5 (MGF), a prerequisite for DNA binding and induction of transcription. EMBO J. 13, 4361–4369.
Green, B., 1984. Composition of milk and energetics of growth in marsupials. In: Peaker, M., Vernon, R.G., Knight, C.H. (Eds.), Symposia of the Zoological Society of London 51 Physiological Strategies in Lactation. Academic Press, London, pp. 369–387.
Green, B., Merchant, J.C., 1988. The composition of marsupial milk. In: Tyndale-Biscoe, C.H., Janssens, P.A. (Eds.), The Developing Marsupial Models for Biomedical Research. Springer-Verlag, Berlin, Germany, pp. 41–54.
Green, B., Newgrain, K., Merchant, J., 1980. Changes in milk composition during lactation in the tammar wallaby (*Macropus eugenii*). Aust. J. Biol. Sci. 33, 35–42.
Green, B., Griffiths, M., Leckie, R.M., 1983. Qualitative and quantitative changes in milk fat during lactation in the tammar wallaby (*Macropus eugenii*). Aust. J. Biol. Sci. 36, 455–461.
Groner, B., Gouilleux, F., 1995. Prolactin-mediated gene activation in mammary epithelial cells. Curr. Opin. Genet. Dev. 5, 587–594.
Helbig, M., Urbanek, S., 2013. JGR: Java Gui for R Augsburg. RoSuDa, Augsburg University, Augsburg.
Hendry, K.A., Simpson, K.J., Nicholas, K.R., Wilde, C.J., 1998. Autocrine inhibition of milk secretion in the lactating tammar wallaby (*Macropus eugenii*). J. Mol. Endocrinol. 21, 169–177.
Hinds, L.A., 1988. Hormonal control of lactation. In: Tyndale-Biscoe, C.H., Janssens, P.A. (Eds.), The Developing Marsupial Models for Biomedical Research. Springer-Verlag, Berlin, Germany, pp. 55–67.
Hinds, L.A., Tyndale-Biscoe, C.H., 1982a. Plasma progesterone levels in the pregnant and non-pregnant tammar, *Macropus eugenii*. J. Endocrinol. 93, 99–107.
Hinds, L.A., Tyndale-Biscoe, C.H., 1982b. Prolactin in the marsupial *Macropus eugenii*, during the estrous cycle, pregnancy and lactation. Biol. Reprod. 26, 391–398.
Houdebine, L.M., 1999. The regulation of milk protein synthesis. In: Martinet, J., Houdebine, L.M., Head, H.H. (Eds.), Biology of Lactation. INRA, Paris, pp. 401–427.
Hughes, K., Watson, C.J., 2012. The spectrum of STAT functions in mammary gland development. JAK-STAT 1, 151–158.
Jack, L.J., Kahl, S., St Germain, D.L., Capuco, A.V., 1994. Tissue distribution and regulation of 5′-deiodinase processes in lactating rats. J. Endocrinol. 142, 205–215.
Janssens, P.A., Hulbert, A.J., Baudinette, R.V., 1997. Development of the pouch young from birth to pouch vacation. In: Saunders, N.R., Hinds, L. (Eds.), Marsupial Biology Recent Research, New Perspectives. University of New South Wales Press, Sydney, pp. 71–89.
Joss, J.L., Molloy, M.P., Hinds, L., Deane, E., 2009. A longitudinal study of the protein components of marsupial milk from birth to weaning in the tammar wallaby (*Macropus eugenii*). Dev. Comp. Immunol. 33, 152–161.
Kino, T., De Martino, M.U., Charmandari, E., Mirani, M., Chrousos, G.P., 2003. Tissue glucocorticoid resistance/hypersensitivity syndromes. J. Steroid Biochem. Mol. Biol. 85, 457–467.
Kulski, J.K., Topper, Y.J., Chomczynski, P., Qasba, P., 1983. An essential role for glucocorticoid in casein gene expression in rat mammary explants. Biochem. Biophys. Res. Commun. 114, 380–387.
Laskowski, M., Kato, I., 1980. Protein inhibitors of proteinases. Ann. Rev. Biochem. 49, 593–626.
Laskowski Jr., M., Laskowski, M., 1951. Crystalline trypsin inhibitor from colostrum. J. Biol. Chem. 190, 563–573.
Laskowski Jr., M., Mars, P.H., Laskowski, M., 1952. Comparison of trypsin inhibitor from colostrum with other crystalline trypsin inhibitors. J. Biol. Chem. 198, 745–752.
Lelievre, S., Weaver, V.M., Bissell, M.J., 1996. Extracellular matrix signaling from the cellular membrane skeleton to the nuclear skeleton: a model of gene regulation. Recent Prog. Horm. Res. 51, 417–432.
Lemay, D.G., Neville, M.C., Rudolph, M.C., Pollard, K.S., German, J.B., 2007. Gene regulatory networks in lactation: identification of global principles using bioinformatics. BMC Syst. Biol. 1, 56.
Lemay, D.G., Pollard, K.S., Martin, W.F., Freeman Zadrowski, C., Hernandez, J., Korf, I., German, J.B., Rijnkels, M., 2013. From genes to milk: genomic organization and epigenetic regulation of the mammary transcriptome. PLoS One 8, e75030.
Lemon, M., Bailey, L.F., 1966. A specific protein difference in the milk from two mammary glands of a red kangaroo. Aust. J. Exp. Biol. Med. Sci. 44, 705–707.
Li, N., Zhang, Y., Naylor, M.J., Schatzmann, F., Maurer, F., Wintermantel, T., Schuetz, G., Mueller, U., Streuli, C.H., Hynes, N.E., 2005. β1 Integrins regulate mammary gland proliferation and maintain the integrity of mammary alveoli. EMBO J. 24, 1942–1953.
Lincoln, D.W., Renfree, M.B., 1981. Mammary gland growth and milk ejection in the agile wallaby, *Macropus agilis*, displaying concurrent asynchronous lactation. J. Reprod. Fertil. 63, 193–203.
Lockwood, D.H., Stockdale, F.E., Topper, Y.J., 1967. Hormone-dependent differentiation of mammary gland: sequence of action of hormones in relation to cell cycle. Science 156, 945–946.
Menzies, K.K., Nicholas, K.R., 2007. The expression of β-1,3 galactosyltransferase and β-1,4 galactosyltransferase enzymatic activities in the mammary gland of the tammar wallaby (*Macropus eugenii*) during early lactation. Biochim. Biophys. Acta – Gen. Subjects 1770, 115–120.
Menzies, K.K., Lefevre, C., Macmillan, K.L., Nicholas, K.R., 2009. Insulin regulates milk protein synthesis at multiple levels in the bovine mammary gland. Funct. Integr. Genomics 9, 197–217.
Menzies, K.K., Lee, H.J., Lefevre, C., Ormandy, C.J., Macmillan, K.L., Nicholas, K.R., 2010. Insulin, a key regulator of hormone responsive milk protein synthesis during lactogenesis in murine mammary explants. Funct. Integr. Genomics 10, 87–95.
Menzies, B.R., Shaw, G., Fletcher, T.P., Pask, A.J., Renfree, M.B., 2012. Maturation of the growth axis in marsupials occurs gradually during post-natal life and over an equivalent developmental stage relative to eutherian species. Mol. Cell. Endocrinol. 349, 189–194.
Naylor, M.J., Li, N., Cheung, J., Lowe, E.T., Lambert, E., Marlow, R., Wang, P., Schatzmann, F., Wintermantel, T., Schuetz, G., Clarke, A.R., Mueller, U., Hynes, N.E., Streuli, C.H., 2005. Ablation of β1 integrin in mammary epithelium reveals a key role for integrin in glandular morphogenesis and differentiation. J. Cell Biol. 171, 717–728.
Neville, M.C., McFadden, T.B., Forsyth, I., 2002. Hormonal regulation of mammary differentiation and milk secretion. J. Mammary Gland Biol. Neoplasia 7, 49–66.
NHMRC, 2004. Australian Code of Practice for the Care and Use of Animals for Scientific Purposes, seventh ed. Australian Government, Canberra.
Nicholas, K.R., 1988a. Asynchronous dual lactation in a marsupial, the tammar wallaby (*Macropus eugenii*). Biochem. Biophys. Res. Commun. 154, 529–536.
Nicholas, K.R., 1988b. Control of milk protein synthesis in the marsupial *Macropus eugenii*: a model system to study prolactin-dependent development. In: Tyndale-Biscoe, C.H., Janssens, P.A. (Eds.), The Developing Marsupial Models for Biomedical Research. Springer-Verlag, Berlin, Germany, pp. 68–85.
Nicholas, K.R., Topper, Y.J., 1980. Enhancement of a-lactabumin-like activity in mammary explants from pregnant rats in the absence of exogenous prolactin. Biochem. Biophys. Res. Commun. 94, 1424–1431.
Nicholas, K.R., Tyndale-Biscoe, C.H., 1985. Prolactin-dependent accumulation of α-lactalbumin in mammary gland explants from the pregnant tammar wallaby (*Macropus eugenii*). J. Endocrinol. 106, 337–342.
Nicholas, K.R., Sankaran, L., Topper, Y.J., 1982. Insulin: a unique role in the phenotypic expression in the induction of milk proteins of rat mammary epithelial cells unrelated to any function in cell maintenance. J. Cell Biol., 95.
Nicholas, K.R., Messer, M., Elliott, C., Maher, F., Shaw, D.C., 1987. A novel whey protein synthesized only in late lactation by the mammary gland from the tammar (*Macropus eugenii*). Biochem. J. 241, 899–904.
Nicholas, K.R., Wilde, C.J., Bird, P.H., Hendry, K.A.K., 1994. Asynchronous expression of milk protein genes during lactation in the tammar wallaby (*Macropus eugenii*). Trends Comp. Biochem. Physiol. 1, 959–972.
Nicholas, K.R., Wilde, C.J., Bird, P.H., Hendry, K.A.K., Tregenza, K., Warner, B., 1995. Asynchronous concurrent secretion of milk proteins in the tammar wallaby (*Macropus eugenii*). In: Wilde, C.J., Peaker, M., Knight, C.H. (Eds.), Intercellular Signalling in the Mammary Gland. Plenum Press, New York, USA, pp. 153–170.
Nicholas, K., Simpson, K., Wilson, M., Trott, J., Shaw, D., 1997. The tammar wallaby: a model to study putative autocrine-induced changes in milk composition. J. Mammary Gland Biol. Neoplasia 2, 299–310.
Old, J.M., Deane, E.M., 2000. Development of the immune system and immunological protection in marsupial pouch young. Dev. Comp. Immunol. 24, 445–454.
Pang, W.W., Hartmann, P.E., 2007. Initiation of human lactation: secretory differentiation and secretory activation. J. Mammary Gland Biol. Neoplasia 12, 211–221.
Perry, J.W., Oka, T., 1984. The study of differentiative potential of the lactating mouse mammary gland in organ culture. In Vitro 20, 59–65.
Pharo, E.A., De Leo, A.A., Renfree, M.B., Thomson, P.C., Lefevre, C.M., Nicholas, K.R., 2012. The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene. BMC Evol. Biol. 12, 8.
Piotte, C.P., Grigor, M.R., 1996. A novel marsupial protein expressed by the mammary gland only during the early lactation and related to the Kunitz proteinase inhibitors. Arch. Biochem. Biophys. 330, 59–64.
Prosser, C.G., Sankaran, L., Hennighausen, L., Topper, Y.J., 1987. Comparison of the roles of insulin and insulin-like growth factor I in casein gene expression and in the development of α-lactalbumin and glucose transport activities in the mouse mammary epithelial cell. Endocrinology 120, 1411–1416.
Puissant, C., Houdebine, L.M., 1991. Cortisol induces rapid accumulation of whey acid protein mRNA but not of α_s1 and β-casein mRNA in rabbit mammary explants. Cell Biol. Int. Rep. 15, 121–129.
Rawlings, N.D., 2010. Peptidase inhibitors in the MEROPS database. Biochimie 92, 1463–1683.
Renfree, M.B., 1994. Endocrinology of pregnancy, parturition and lactation in marsupials. In: Lamming, G.E. (Ed.), Marshall's Physiology of Reproduction. Chapman & Hall, London, UK, pp. 677–766.
Renfree, M.B., Tyndale-Biscoe, C.H., 1973. Intrauterine development after diapause in marsupial *Macropus-eugenii*. Dev. Biol. 32, 28–40.
Renfree, M.B., Green, S.W., Young, I.R., 1979. Growth of the corpus luteum and its progesterone content during pregnancy in the tammar wallaby, *Macropus eugenii*. J. Reprod. Fertil. 57, 131–136.
Renfree, M.B., Fletcher, T.P., Blanden, D.R., Lewis, D.R., Shaw, G., Gordon, K., Short, R.V., Parer-Cook, E., Parer, D., 1989. Physiological and behavioural events around the time of birth in macropodid marsupials. In: Grigg, G., Jarman, P.J.,

Author's personal copy

Hume (Eds.), Kangaroos, Wallabies and Rat-Kangaroos. Surrey Beatty and Sons Pty. Ltd., Sydney, pp. 323–337.
Renfree, M.B., Papenfuss, A.T., Deakin, J.E., Lindsay, J., Heider, T., Belov, K., Rens, W., Waters, P.D., Pharo, E.A., Shaw, G., Wong, E.S., Lefevre, C.M., Nicholas, K.R., Kuroki, Y., Wakefield, M.J., Zenger, K.R., Wang, C., Ferguson-Smith, M., Nicholas, F.W., Hickford, D., Yu, H., Short, K.R., Siddle, H.V., Frankenberg, S.R., Chew, K.Y., Menzies, B.R., Stringer, J.M., Suzuki, S., Hore, T.A., Delbridge, M.L., Mohammadi, A., Schneider, N.Y., Hu, Y., O'Hara, W., Al Nadaf, S., Wu, C., Feng, Z.P., Cocks, B.G., Wang, J., Flicek, P., Searle, S.M., Fairley, S., Beal, K., Herrero, J., Carone, D.M., Suzuki, Y., Sugano, S., Toyoda, A., Sakaki, Y., Kondo, S., Nishida, Y., Tatsumoto, S., Mandiou, I., Hsu, A., McColl, K.A., Lansdell, B., Weinstock, G., Kuczek, E., McGrath, A., Wilson, P., Men, A., Hazar-Rethinam, M., Hall, A., Davis, J., Wood, D., Williams, S., Sundaravadanam, Y., Muzny, D.M., Jhangiani, S.N., Lewis, L.R., Morgan, M.B., Okwuonu, G.O., Ruiz, S.J., Santibanez, J., Nazareth, L., Cree, A., Fowler, G., Kovar, C.L., Dinh, H.H., Joshi, V., Jing, C., Lara, F., Thornton, R., Chen, L., Deng, J., Liu, Y., Shen, J.Y., Song, X.Z., Edson, J., Troon, C., Thomas, D., Stephens, A., Yapa, L., Levchenko, T., Gibbs, R.A., Cooper, D.W., Speed, T.P., Fujiyama, A., JA, M.G., O'Neill, R.J., Pask, A.J., Forrest, S.M., Worley, K.C., 2011. Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. Genome Biol. 12, R81.
Richardson, S.J., Aldred, A.R., Leng, S.L., Renfree, M.B., Hulbert, A.J., Schreiber, G., 2002. Developmental profile of thyroid hormone distributor proteins in a marsupial, the tammar wallaby *Macropus eugenii*. Gen. Comp. Endocr. 125, 92–103.
Rijnkels, M., Freeman-Zadrowski, C., Hernandez, J., Potluri, V., Wang, L., Li, W., Lemay, D.G., 2013. Epigenetic modifications unlock the milk protein gene loci during mouse mammary gland development and differentiation. PLoS One 8, e53270.
Rijnkels, M., Kabotyanski, E., Montazer-Torbati, M.B., Beauvais, C.H., Vassetzky, Y., Rosen, J.M., Devinoy, E., 2010. The epigenetic landscape of mammary gland development and functional differentiation. J. Mammary Gland Biol. Neoplasia 15, 85–100.
Riley, L.G., Gardiner-Garden, M., Thomson, P.C., Wynn, P.C., Williamson, P., Raadsma, H.W., Sheehy, P.A., 2010. The influence of extracellular matrix and prolactin on global gene expression profiles of primary bovine mammary epithelial cells *in vitro*. Anim. Genet. 41, 55–63.
Rosen, J.M., Wyszomierski, S.L., Hadsell, D., 1999. Regulation of milk protein gene expression. Ann. Rev. Nutr. 19, 407–436.
Rudolph, M.C., McManaman, J.L., Hunter, L., Phang, T., Neville, M.C., 2003. Functional development of the mammary gland: use of expression profiling and trajectory clustering to reveal changes in gene expression during pregnancy, lactation, and involution. J. Mammary Gland Biol. Neoplasia 8, 287–307.
Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675.
Sharman, G.B., Pilton, P.E., 1964. The life history and reproduction of the red kangaroo, *Megaleia rufa*. Proc. Zool. Soc. London 142, 29–48.
Shaw, G., Renfree, M.B., 2006. Parturition and perfect prematurity: birth in marsupials. Aust. J. Zool. 54, 139–149.
Shaw, G., Renfree, M.B., Fletcher, T.P., 1996. A role for glucocorticoids in parturition in a marsupial, *Macropus eugenii*. Biol. Reprod. 54, 728–733.
Sheehy, P.A., Della-Vedova, J.J., Nicholas, K.R., Wynn, P.C., 2004. Hormone-dependent milk protein gene expression in bovine mammary explants from biopsies at different stages of pregnancy. J. Dairy Res. 71, 135–140.
Simpson, K., 1998. The Tammar Wallaby, *Macropus eugenii*, A Model to Study Autocrine and Endocrine Regulation of Lactation. PhD Thesis. Department of Agriculture Melbourne, Latrobe University.
Simpson, K., Shaw, D., Nicholas, K., 1998. Developmentally-regulated expression of a putative protease inhibitor gene in the lactating mammary gland of the tammar wallaby, *Macropus eugenii*. Comp. Biochem. Phys. B 120, 535–541.
Simpson, K.J., Ranganathan, S., Fisher, J.A., Janssens, P.A., Shaw, D.C., Nicholas, K.R., 2000. The gene for a novel member of the whey acidic protein family encodes three four-disulfide core domains and is asynchronously expressed during lactation. J. Biol. Chem. 275, 23074–23081.
Skarda, J., Urbanova, E., Houdebine, L.M., Delouis, C., Bilek, J., 1982. Effects of insulin, cortisol and prolactin on lipid, protein and casein syntheses in goat mammary tissue in organ culture. Reprod. Nutr. Dev. 22, 379–386.
Song, S., Oka, T., 2003. Regulation of type II deiodinase expression by EGF and glucocorticoid in HC11 mouse mammary epithelium. Am. J. Physiol. Endocrinol. Metab. 284, E1119–E1124.
Stewart, F., 1984. Mammogenesis and changing prolactin receptor concentrations in the mammary glands of the tammar wallaby (*Macropus eugenii*). J. Reprod. Fertil. 71, 141–148.
Streuli, C.H., Bailey, N., Bissell, M.J., 1991. Control of mammary epithelial differentiation: basement membrane induces tissue-specific gene expression in the absence of cell–cell interaction and morphological polarity. J. Cell Biol. 115, 1383–1395.
Streuli, C.H., Schmidhauser, C., Bailey, N., Yurchenco, P., Skubitz, A.P., Roskelley, C., Bissell, M.J., 1995. Laminin mediates tissue-specific gene expression in mammary epithelia. J. Cell Biol. 129, 591–603.
Stringer, J.M., Suzuki, S., Pask, A.J., Shaw, G., Renfree, M.B., 2012. Selected imprinting of *INS* in the marsupial. Epigenetics Chromatin 5, 14.
R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
Topper, Y.J., Freeman, C.S., 1980. Multiple hormone interactions in the developmental biology of the mammary gland. Physiol. Rev. 60, 1049–1106.
Trott, J.F., 1999. Endocrine and Local Control of Milk Protein Gene Expression in the Tammar Wallaby, *Macropus eugenii*. PhD Thesis. Department of Molecular Biology, Genetics & Zoology Melbourne, University of Melbourne.
Trott, J.F., Wilson, M.J., Hovey, R.C., Shaw, D.C., Nicholas, K.R., 2002. Expression of novel lipocalin-like milk protein gene is developmentally-regulated during lactation in the tammar wallaby, *Macropus eugenii*. Gene 283, 287–297.
Trott, J.F., Simpson, K.J., Moyle, R.L., Hearn, C.M., Shaw, G., Nicholas, K.R., Renfree, M.B., 2003. Maternal regulation of milk composition, milk production, and pouch young development during lactation in the tammar wallaby (*Macropus eugenii*). Biol. Reprod. 68, 929–936.
Trott, J.F., Adams, T.E., Wilson, M., Nicholas, K.R., 2005. Positive and negative regulatory elements in the late lactation protein-A gene promoter from the tammar wallaby (*Macropus eugenii*). Biochim. Biophys. Acta – Gen Subjects – Gene Struct. Expr. 1728, 65–76.
Tyndale-Biscoe, C.H., Janssens, P.A. (Eds.), 1988. The Developing Marsupial. Models for Biomedical Research. Springer-Verlag, Berlin, Germany.
Tyndale-Biscoe, C.H., Renfree, M.B., 1987. Reproductive Physiology of Marsupials Cambridge. Cambridge University Press, UK.
Tyndale-Biscoe, C.H., Stewart, F., Hinds, L.A., 1984. Some factors in the initiation and control of lactation in the tammar wallaby. In: Symposia of the Zoological Society of London 51 Physiological Strategies in Lactation, pp. 389–401.
Veselsky, L., Cechova, D., Jonakova, V., 1978. Secretion and immunochemical properties of the trypsin inhibitor from bovine colostrum. Hoppe Seylers Z. Physiol. Chem. 359, 873–878.
Vonderhaar, B.K., Greco, A.E., 1979. Lobulo-alveolar development of mouse mammary glands is regulated by thyroid hormones. Endocrinology 104, 409–418.
Vonderhaar, B.K., Ziska, S.E., 1989. Hormonal regulation of milk protein gene expression. Annu. Rev. Physiol. 51, 641–652.
Wakao, H., Gouilleux, F., Groner, B., 1994. Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response. EMBO J. 13, 2182–2191.
Wanyonyi, S.S., Lefevre, C., Sharp, J.A., Nicholas, K.R., 2013. The extracellular matrix locally regulates asynchronous concurrent lactation in tammar wallaby (*Macropus eugenii*). Matrix Biol. 32, 342–351.
Ward, K.L., Renfree, M.B., 1984. Effects of progesterone on parturition in the tammar, *Macropus eugenii*. J. Reprod. Fertil. 72, 21–28.
Warner, B., Janssens, P., Nicholas, K., 1993. Prolactin-independent induction of alpha-lactalbumin gene expression in mammary gland explants from pregnant Balb/c mice. Biochem. Biophys. Res. Commun. 194, 987–991.
Watson, C., 2006. Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. Breast Cancer Res. 8, 203.
Watson, C., Burdon, T., 1996. Prolactin signal transduction mechanisms in the mammary gland: the role of the Jak/Stat pathway. Rev. Reprod. 1, 1–5.
Zhou, J., Cidlowski, J.A., 2005. The human glucocorticoid receptor: one gene, multiple proteins and diverse responses. Steroids 70, 407–417.

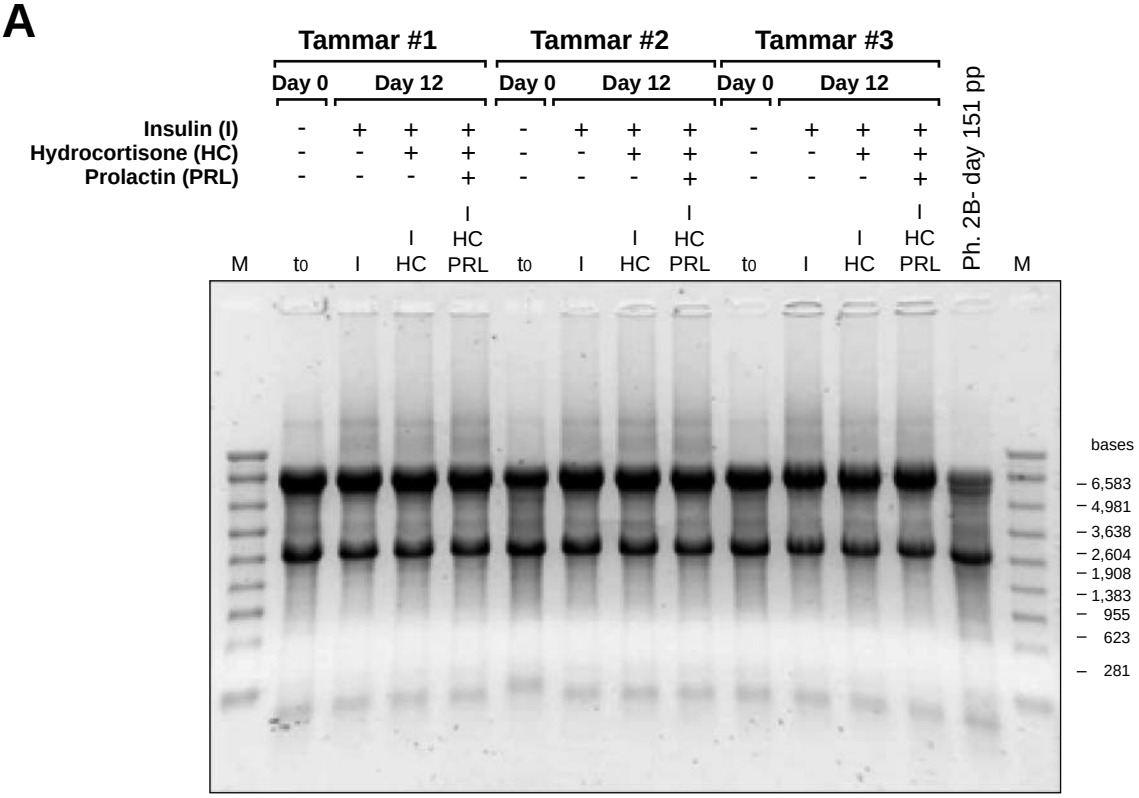
Supplementary file 1



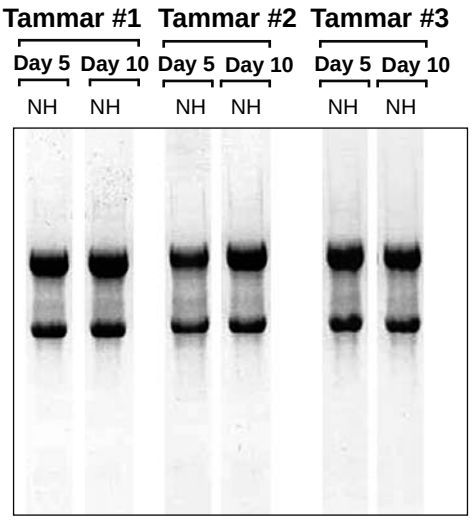
A-C. Phase 1 (day 24 of pregnancy) explant experiment. Mammary explants were cultured from three different animals for 4 and 8 days in M199 supplemented with various hormones (I, insulin; HC, hydrocortisone; PRL, prolactin; T_3, triiodothyronine; E_2, 17β-oestradiol). Total RNA was extracted from t0 tissue and from explants and 10 μg per sample electrophoresed through low-formaldehyde, 1% agarose gels with 1X MOPS [3(N-Morpholino) Propane Sulfonic Acid] buffer at 4°C. **A.** Tammar #1, **B.** Tammar #2, and **C.** Tammar #3. Phase control samples are also shown. RNA markers indicate transcript sizes (#G319A, Promega Corporation).

Supplementary file 2

A

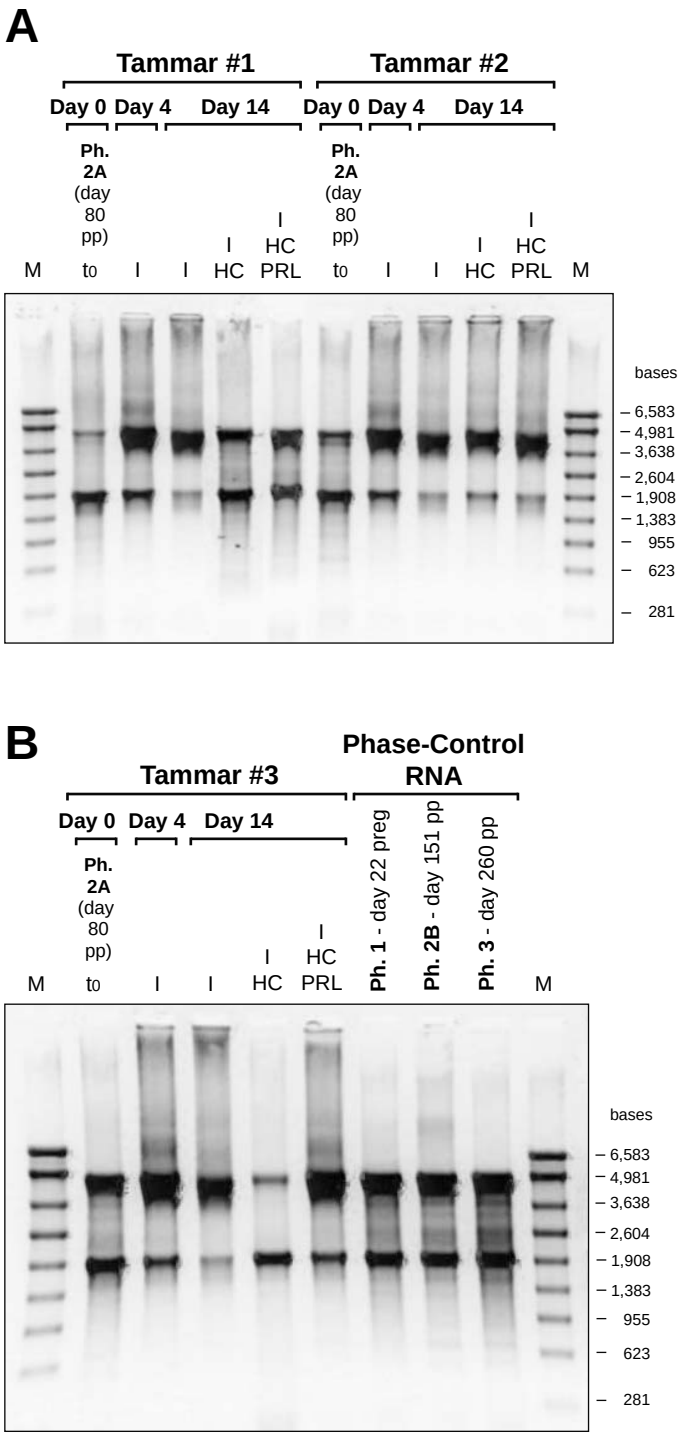


B



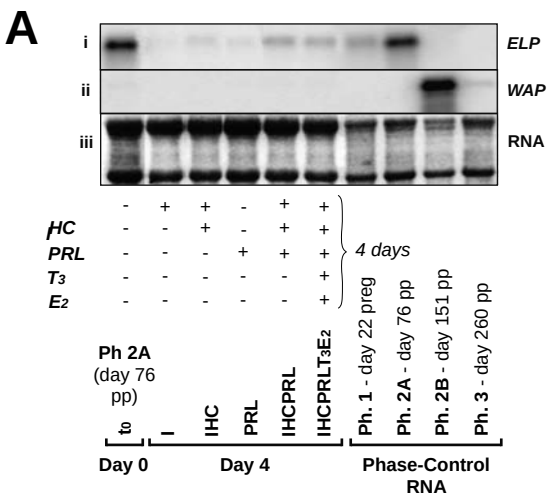
A & B. Phase 1 (day 24 of pregnancy) explant experiment. **A.** Mammary explants were cultured from three different animals for 12 days in M199 supplemented with various hormones (I, insulin; HC, hydrocortisone; PRL, prolactin). Total RNA was extracted from t_0 tissue and from explants and 10 μg per sample electrophoresed through low-formaldehyde, 1% agarose gels with 1X MOPS [3(N-Morpholino) Propane Sulfonic Acid] buffer at 4°C. Total RNA for Tammar #1, Tammar #2, and Tammar #3, as well as for a day 151 pp phase control sample are shown. RNA markers indicate transcript sizes (#G319A, Promega Corporation). **B.** Total RNA for tammar mammary explants cultured in no hormone (NH) Medium 199 for 5- and 10-day periods.

Supplementary file 3



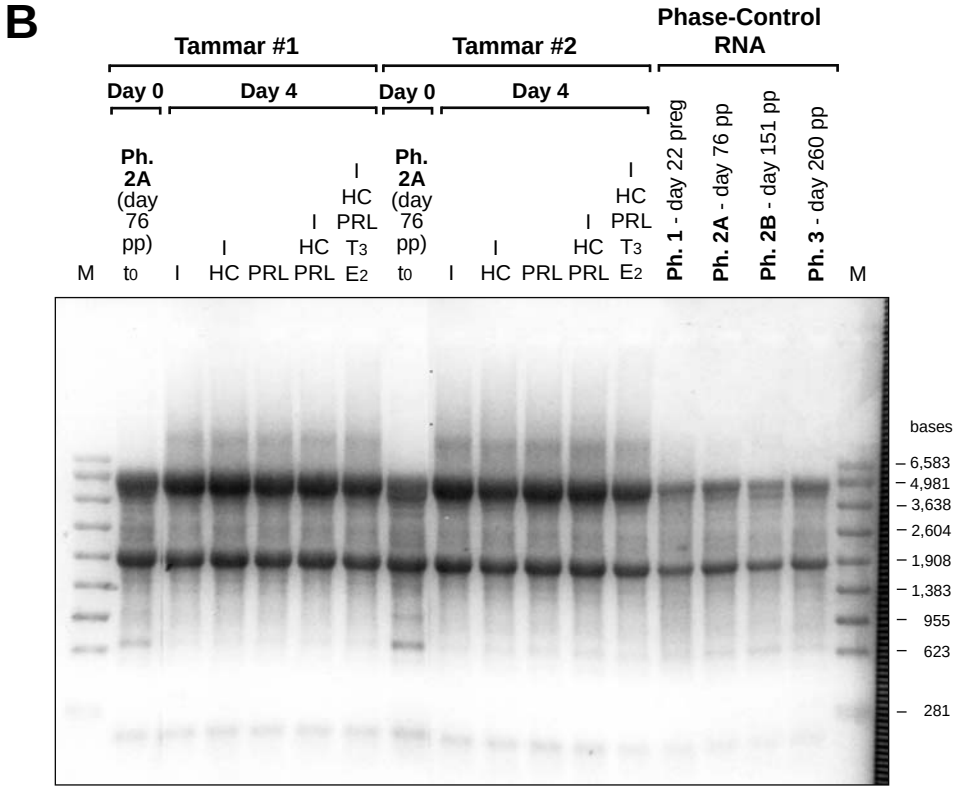
A & B. Phase 2A (day 80 pp, early lactation) explant experiment. **A.** Mammary explants were cultured from three different animals for 12 days in M199 supplemented with various hormones (I, insulin; HC, hydrocortisone; PRL, prolactin). Total RNA was extracted from t_0; Day 0 tissue and from explants and 10 µg per sample electrophoresed through low-formaldehyde, 1% agarose gels with 1X MOPS [3(N-Morpholino) Propane Sulfonic Acid] buffer at 4°C. **A.** Total RNA for Tammar #1 & Tammar #2, and **B.** Tammar #3 and phase control sample are shown. RNA markers indicate transcript sizes (#G319A, Promega Corporation).

Supplementary file 4



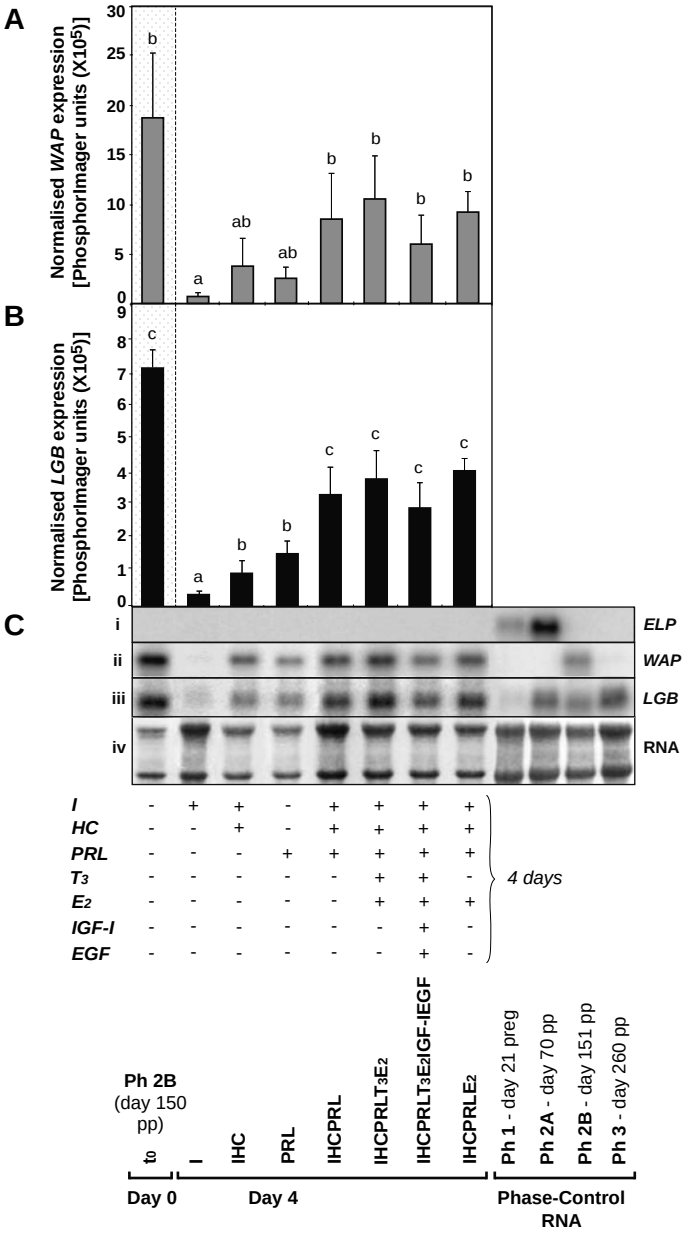
A. Endocrine regulation of *ELP* and *WAP* in mammary explants cultured from early lactation (Phase 2A)

Explants from two tammars at day 76 pp (t_0; Day 0) were cultured for 4 days in M199 supplemented with different hormones (I, insulin; HC, hydrocortisone; PRL, prolactin; T_3, triiodothyronine; E_2, 17β-oestradiol). Northern analysis indicated that (i) *ELP* expression was maximal in treatments containing IHCPRL, at similar levels to the Phase 1 control tissue. (ii) No *WAP* expression was detected in explants, whilst *LGB* expression was highest in the IHCPRL treated explants (data not shown). Phase-specific (Ph. 1, Ph. 2A, Ph. 2B, Ph. 3) control tissues are shown, whilst (iii) ethidium bromide staining of ribosomal bands indicates equal loading and integrity of total RNA (10 μg). Total RNA gels for all samples are provided below (B).



B. Phase 2A (day 76 pp) explant experiment. Mammary explants were cultured from two different animals for 4 days in M199 supplemented with various hormones (I, insulin; HC, hydrocortisone; PRL, prolactin; T_3, triiodothyronine; E_2, 17β-oestradiol). Total RNA was extracted from t_0 tissue and from explants and 10 μg per sample electrophoresed through low-formaldehyde, 1% agarose gels with 1X MOPS [3(N-Morpholino) Propane Sulfonic Acid] buffer at 4°C. Total RNA for Tammar #1, and Tammar #2, as well as for Phase 1, 2A, 2B and 3 control samples are shown. RNA markers indicate transcript sizes (#G319A, Promega Corporation).

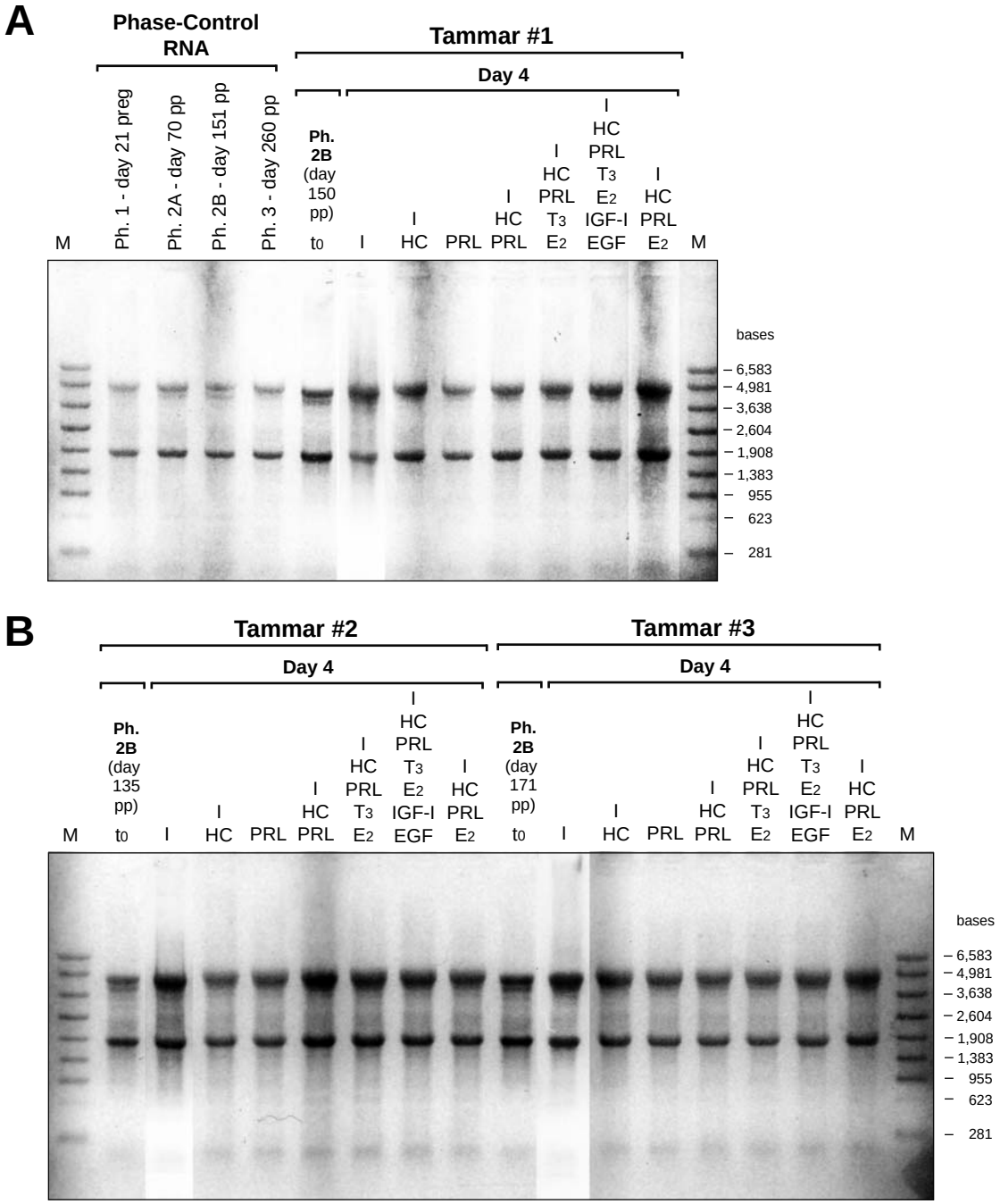
Supplementary file 5



Endocrine regulation of *ELP, WAP* and *LGB* in mammary explants cultured from mid-lactation (Phase 2B)

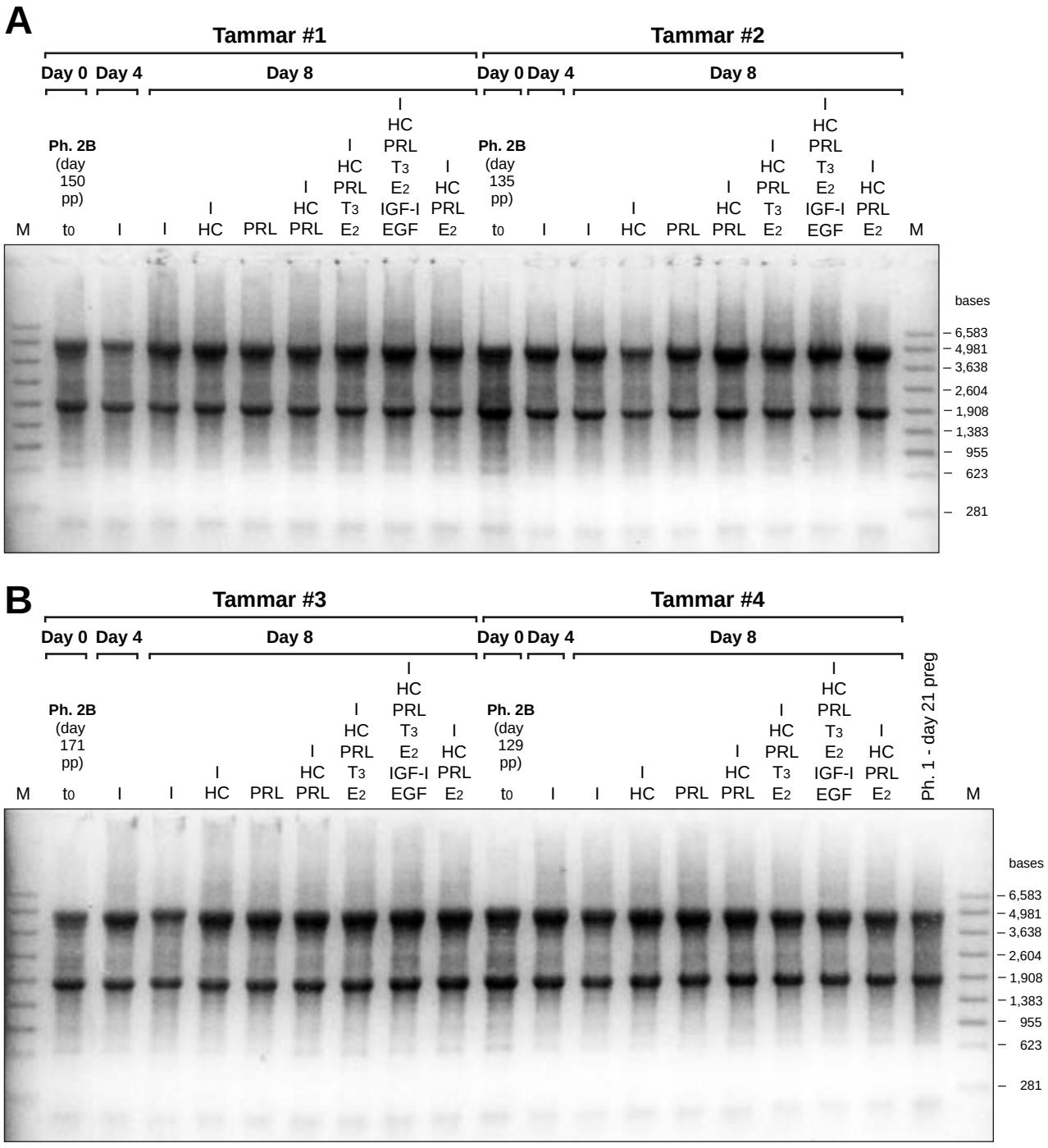
A-C. Explants from ~day 150 pp (t_0; Day 0) were cultured for 4 days in M199 supplemented with different hormones (I, insulin; HC, hydrocortisone; PRL, prolactin; T_3, triiodothyronine; E_2, 17β-oestradiol ; IGF-I, insulin-like growth factor I; EGF, epidermal growth factor). **C.** Northern analysis showed that (i) *ELP* was not detected in explants, but was present in Phase 1 and 2A control tissues. (ii) *WAP* transcript levels were minimal in 4-day insulin-treated explants and high in explants treated with IHCPRL. (iii) *LGB* expression followed similar trends to *WAP*, with minimal expression in the 4-day I treatments and maximal expression in all treatments supplemented with IHCPRL. Phase-specific (Ph. 1, Ph. 2A, Ph. 2B, Ph. 3) control tissues are shown, whilst (iv) ethidium bromide staining of ribosomal bands indicates equal loading and integrity of total RNA (10 μg). **A.** PhosphorImager analysis showed that *WAP* expression was higher in the IHCPRL-treated explants at day 8 than those treated with I for 4 days ($p<0.05$), but NS compared to IF and PRL at day 8. **B.** *LGB* expression was highest in IHCPRL-treated explants compared to the I, IHC and PRL treatments. *WAP* and *LGB* gene expression levels (PhosphorImager units) were normalised to RNA loading and represent average gene expression ± SEM for 4 animals. Bars with different letters indicate treatments for which gene expression levels (log transformed data) differ significantly ($p<0.05$) based upon Tukey's multiple comparison of means. Total RNA gels for all samples are provided (Supplementary file 6).

Supplementary file 6



A-C. Phase 2B explant experiment. Mammary explants were cultured from four different animals for 4 days in M199 supplemented with various hormones (I, insulin; HC, hydrocortisone; PRL, prolactin; T_3, triiodothyronine; E_2, 17β-oestradiol; IGF-I, insulin-like growth factor I; EGF, epidermal growth factor). Total RNA was extracted from (t_0; Day 0) tissue and from explants and 10 μg per sample electrophoresed through low-formaldehyde, 1% agarose gels with 1X MOPS [3(N-Morpholino) Propane Sulfonic Acid] buffer at 4°C. **A.** Phase control total RNA samples (Ph. 1, Ph. 2A, Ph. 2B, Ph. 3) & Tammar #1 **B.** Tammar #2 & Tammar#3, and **C.** Tammar #4. RNA markers indicate transcript sizes (#G319A, Promega Corporation).

Supplementary file 7



Phase 2B explant experiment. Mammary explants were cultured from four different animals for 4 and 8 days in M199 supplemented with various hormones (I, insulin; HC, hydrocortisone; PRL, prolactin; T_3, triiodothyronine; E_2, 17β-oestradiol ; IGF-I, insulin-like growth factor I; EGF, epidermal growth factor). Total RNA was extracted from (t_0; Day 0) tissue and from explants and 10 µg per sample electrophoresed through low-formaldehyde, 1% agarose gels with 1X MOPS [3(N-Morpholino) Propane Sulfonic Acid] buffer at 4°C. **A.** Tammar #1 & Tammar #2, **B.** Tammar #3 & Tammar #4. **C.** Phase control total RNA samples (Ph. 1, Ph. 2A, Ph. 2B, Ph. 3). RNA markers indicate transcript sizes (#G319A, Promega Corporation).

Chapter Six:

Mammary cell-activating factor regulates the hormone-independent transcription of the *Early Lactation Protein* (*ELP*) gene in a marsupial

Submitted to Molecular and Cellular Endocrinology

Mammary cell-activating factor regulates the hormone-independent transcription of the *Early Lactation Protein* (*ELP*) gene in a marsupial

Elizabeth A Pharo[a,b]†*, Marilyn B Renfree[a] and Kylie N Cane[a,b]

[a] *School of BioSciences, The University of Melbourne, Melbourne, VIC, 3010, Australia*

[b] *Cooperative Research Centre for Innovative Dairy Products, Australia*

* Corresponding author

Email: e.pharo@unimelb.edu.au

Tel: +61-3-8344-4376

Fax: +61-3-8344-7909

† Present address

Email addresses of other authors:

Marilyn B Renfree: m.renfree@unimelb.edu.au

Kylie N Cane: kylie.cane@gmail.com

Abstract

The regulation of the tammar wallaby *early lactation protein* (*ELP*) gene is complex. *ELP* is responsive to the lactogenic hormones; insulin (I), hydrocortisone (HC) and prolactin (PRL) in mammary gland explants, but in primary cells isolated from the tammar mammary gland it could not be induced. Tammar wallaby *ELP* promoter constructs transiently-transfected into KIM-2 cells were also unresponsive to IHCPRL, unlike the rat and mouse *β-casein* (*CSN2*) promoter constructs. Similarly, although transiently-transfected rodent *CSN2* promoter constructs were PRL-responsive in HEK293Ts constitutively expressing the prolactin receptor and signal transducer and activator of transcription 5A (STAT5A), tammar *ELP* promoter constructs were not. Identification of the minimal promoter required for the hormone-independent transcription of tammar *ELP* in human embryonic kidney cells (HEK293Ts) and comparative analysis of marsupial *ELP* and orthologous eutherian colostrum trypsin inhibitor (*CTI*) proximal promoters suggested that mammary cell-activating factor (MAF), an E26 transformation-specific (ETS) factor, binds to the CTTCCT(-) motif and activates tammar *ELP*.

Keywords

MAF, ETS, C/EBP, mammary gland, Kunitz, Colostrum trypsin inhibitor

Abbreviations

AP, activating protein; BCE1, bovine casein enhancer 1; BPTI, bovine pancreatic trypsin inhibitor; CAMP, cathelicidin antimicrobial peptide; CAT, chloramphenicol acetyltransferase; C/EBP, CCAAT/enhancer binding protein; CREB, cAMP response element binding protein; CMV, cytomegalovirus; CoRE, composite response element; *CSN1*, *α-casein*; *CSN2*, *β-casein*; CTF, CCAAT box transcription factor; CTI, colostrum trypsin inhibitor; D, dexamethasone; d2EGFP, destabilised enhanced green fluorescent protein; DHS, DNase I hypersensitivity site; DRE, distal regulatory element; ECM, extracellular matrix; ELF, E47-like factor; ELP, early lactation protein; ERE, estrogen response element; EST, expressed sequence tag; ERRα/β, estrogen-related receptor alpha/beta; ETS, E26 transformation-specific; EVI1, ecotopic virus integration site 1; GAS, interferon-gamma activation site; GR, glucocorticoid receptor; ½GRE, glucocorticoid response element; H3Ac, histone H3 acetylation; H3K4me2, histone H3 lysine 4 di-methylation; HC, hydrocortisone; I, insulin; IFNɣ, interferon gamma; Inr, initiator element; IRE, insulin response element; IRF, Interferon regulatory factor; JAK, Janus kinase; L, lactation; *LALBA*; α-lactalbumin; *LGB*, β-lactoglobulin; *LLP*, late lactation protein; MaeuCath, tammar wallaby cathelicidin; MAF, mammary cell-activating factor; MEC, mammary epithelial cell; MGF, mammary gland factor; MMTV, mouse mammary tumour virus; MZF, Myeloid zinc finger; ncRNA, non-coding RNA; NFAT, Nuclear activator of T-cells; NFI, nuclear factor I; NF-IL6, nuclear factor-interleukin 6; NF-κB, nuclear factor kappa-light-chain-enhancer of activated B cells; NFY, nuclear factor Y; OCT, octamer; P, pregnancy; PMF, pregnancy-specific mammary nuclear factor; PMGC, primary mammary gland cells; pp, postpartum; PRL, prolactin; PRLR, prolactin receptor; SP, specificity protein; STAT, signal transducer and activator of transcription; TBP, TATA binding protein; TFBS, transcription factor binding site; TSS, transcription start site; UTR, untranslated region; WAP, whey acidic protein; WPPCS, whey protein promoter core sequence; YY1, Yin-Yang 1.

Introduction

Marsupials have a long and complex lactation, during which, the composition of milk proteins, carbohydrates and lipids and mammary gland gene expression change dynamically (Green, 1984; Green *et al.*, 1980; Nicholas *et al.*, 1997; Sharp *et al.*, 2014). In contrast, apart from the secretion of colostrum for the first 24-36 h postpartum (pp), the changes in mature eutherian milk are generally less dramatic (Jenness, 1974; Kruse, 1983).

The mammary gland-specific marsupial *early lactation protein (ELP*) gene encodes one of the major tammar wallaby milk proteins, which is secreted only in the first third of lactation after birth (Nicholas *et al.*, 1997; Pharo *et al.*, 2012; Simpson *et al.*, 1998). *ELP*, like its orthologous eutherian *colostrum trypsin inhibitor* (*CTI)* gene (Pharo *et al.*, 2012) is expressed at low levels during mid-late pregnancy (Phase 1; preparation for lactation) and then is upregulated at parturition (beginning of Phase 2A) (Laskowski and Laskowski, 1951; Nicholas *et al.*, 1997; Pharo *et al.*, 2012; Veselsky *et al.*, 1978). Marsupial *ELP* expression is prolonged relative to the duration of lactation in three unrelated marsupial species; for at least 37 days pp of the 70 day lactation in the stripe-faced dunnart (De Leo *et al.*, 2006; Frigo and Woolley, 1997), for up to 100 of 200 days pp in the brushtail possum (Demmer *et al.*, 1998; Demmer *et al.*, 2001) and for ~100-125 of 350 days pp during early lactation (Phase 2A) to early Phase 2B in the tammar wallaby (Nicholas *et al.*, 1997; Simpson, 1998; Elizabeth Pharo, unpublished data). Thereafter, *ELP* is down-regulated for the remainder of lactation (Phases 2B and 3) (Nicholas *et al.*, 1997; Simpson, 1998).

Unlike *ELP*, in the only three eutherian species examined so far, bovine *CTI* expression is brief: ~24-48 h pp (Veselsky *et al.*, 1978), with CTI protein highest in colostrum on day 1 pp (Laskowski and Laskowski, 1951). In contrast, Cape fur seal *CTI* is expressed for at least 2-3 months pp and expression may be intermittent, i.e. up-regulated when the mother suckles her pup on-shore and down-regulated as the mother forages at-sea (Pharo *et al.* 2016). This is supported by the lack of involution in the mammary glands of the lactating Cape fur seals that were foraging at sea (Cane, 2005).

ELP encodes a small (~9-21 kDa), N-glycosylated, secreted milk protein (Joss *et al.*, 2007; Joss *et al.*, 2009; Piotte and Grigor, 1996; Simpson *et al.*, 1998) with a single Kunitz/(bovine pancreatic trypsin inhibitor) BPTI domain indicative of serine protease inhibition (Rawlings *et al.*, 2014). However, the precise function and target protease(s) of ELP are yet to be determined. The secretion of ELP during a period when the gut of

the young is immature (Kwek *et al.*, 2009) and therefore permeable to macromolecules and intact immunoglobulins (Brambell, 1970) suggests that ELP is essential for the survival of the marsupial young. Furthermore, ELP expression coincides with many key stages of pouch young development including a period of disproportionate brain growth, the transitioning from mesonephric to metanephric kidneys and from hetero- to homeothermic temperature regulation, plus the lack of an acquired immune system (Edwards *et al.*, 2012; Tyndale-Biscoe and Renfree, 1987). Similarly, the temporal secretion of CTI and its inhibition of trypsin (Laskowski and Laskowski, 1951) suggests that it may inhibit the degradation of immunoglobulins transferred from mother to young via milk (Laskowski and Laskowski, 1951).

The endocrine regulation of tammar *ELP* has been investigated using an *in vitro* mammary gland explant culture system (Pharo, 2014; Simpson, 1998) which retains a 3D tissue structure typical of the lactation phenotype (Brennan *et al.*, 2008). The maximal response of *ELP* to the lactogenic hormones: insulin (I), hydrocortisone (HC) and prolactin (PRL) *in vitro* (Pharo, 2014; Simpson, 1998) is similar to that observed during the *in vivo* induction of *ELP* at parturition when these hormones are elevated (Hinds, 1988; Shaw and Renfree, 2006). However, *ELP* can only be induced in explants *in vitro* if the gene is already expressed *in vivo*, i.e. in tissues from Phase 1 (pregnancy) and Phase 2A (early lactation) tissue, and not in mid- (Phase 2B) (Pharo, 2014) or late lactation (Phase 3) tissue (Elizabeth Pharo, unpublished data). Interestingly, *ELP* expression is maintained in Phase 1 explants treated with either I or I+HC (Pharo, 2014), unlike many other milk protein genes which require lactogenic hormones, or PRL at least, to initiate gene transcription (Brisken and Ataca, 2015; Rosen *et al.*, 1999; Topper and Freeman, 1980).

The inability to induce *ELP* expression in mid-and late lactation explants with lactogenic hormones (Pharo, 2014) suggests that *ELP* regulation is complex. Furthermore, the identification of ELP as a putative Kunitz serine protease inhibitor (Cechova *et al.*, 1969; Nicholas *et al.*, 1997; Piotte and Grigor, 1996) suggests that like its potential target protease(s), it may be tightly-regulated in order to maintain cellular homeostasis and to prevent pathological conditions such as inflammation and cancer (Laskowski and Qasim, 2000; Scott and Taggart, 2010). *ELP* regulation may involve novel mechanisms including those intrinsic to the mammary gland. These may include *cis*- and *trans*-acting transcription factors (those that bind directly to a consensus sequence and those that bind to other factors, but not directly to the DNA, respectively),

non-coding RNAs (ncRNA), distal and/or intragenic enhancers, extracellular matrix (ECM) factors, cell-cell interactions, global and/or local changes in chromatin structure and post-transcriptional regulation (Qian and Zhao, 2014; Rijnkels *et al.*, 2010; Riley *et al.*, 2010; Rosen *et al.*, 1999). Furthermore, recent research has challenged our understanding of the regulation of gene transcription (Biggin, 2011; Murray *et al.*, 2015; Pai *et al.*, 2015; Spivakov, 2014).

The transcriptional and endocrine mechanisms that regulate several of the major marsupial milk protein genes have been investigated e.g. whey protein genes: *β-lactoglobulin* (*LGB*), *whey acidic protein* (*WAP*), *α-lactalbumin* (*LALBA*), *late lactation protein (LLP) A* and *B* and caseins: *α-casein* (*CSN2*), *β-casein* (*CSN2*) (Brennan *et al.*, 2007; Collet *et al.*, 1991, 1992; Nicholas and Tyndale-Biscoe, 1985; Simpson *et al.*, 2000; Topcic *et al.*, 2009; Trott *et al.*, 2005; Trott *et al.*, 2002). As for eutherians (see reviews) (Qian and Zhao, 2014; Rosen *et al.*, 1999; Rosen *et al.*, 1998)), many of these genes are prolactin-responsive. Hence, signal transducer and activator of transcription 5A (STAT5A)-binding to the promoter is essential for gene transcription (Hennighausen and Robinson, 2005; Qian and Zhao, 2014; Rosen *et al.*, 1999). The binding of PRL to the long form of its receptor (PRLR) on the mammary epithelial cell (MEC) surface induces the Janus kinase (JAK)-STAT/PRL-signalling pathway (see reviews) (Hughes and Watson, 2012; Watson and Burdon, 1996). The PRLR dimerises and JAK2 phosphorylates tyrosine residues of the PRLR, STAT5A and JAK2 itself. STAT5A dimerises and is translocated into the nucleus where it binds to palindromic TTCNNNGAA sites within milk protein gene promoters and activates gene transcription (Gouilleux *et al.*, 1994; Horseman and Gregerson, 2014; Wakao *et al.*, 1994). Insulin signalling and the diffusion of glucocorticoids into mammary epithelial cells are also critical for gene activation (Menzies *et al.*, 2009; Zhou and Cidlowski, 2005).

Although the mechanisms that regulate the major milk protein genes have been investigated (Qian and Zhao, 2014; Rijnkels *et al.*, 2010; Rosen *et al.*, 1999; Rosen *et al.*, 1998), those that control the temporal, mammary-specific expression of *ELP* are unknown. We identified the tammar *ELP* promoter by primer extension analysis and examined its regulation using *in vitro* cell culture models and promoter reporter assays. An *in silico* comparative genomics strategy was used to identify conserved putative transcription factor binding sites (TFBSs) within the marsupial *ELP* promoter that may regulate the early lactation-specific transcription of the gene.

Materials and Methods

Animals

Tammar wallabies (*Macropus eugenii*) were kept in grass paddocks with access to food, water and shelter in accordance with the National Health and Medical Research Council guidelines (NH&MRC, 2013). All experiments were approved by The University of Melbourne Animal Experimentation Ethics Committee.

Primer extension analysis of tammar *ELP*

The transcription start site (TSS) of the tammar *ELP* promoter at three different time points during pregnancy and lactation: day 24P (Phase 1, pregnancy), day 80L (Phase 2A, early lactation) and day 260L (Phase 3, late lactation) was identified using the Primer Extension System AMV Reverse Transcriptase kit (Promega Corporation) according to the manufacturer's instructions. A 30 bp tammar *ELP* reverse primer: 5'-GCGAAGTAGAGGGCAACGATGGTAAATTTC-3' was designed to bind within the coding region of tammar *ELP* exon 1 (+31 bp to +2 bp relative to the start of translation, ATG) and extend transcription to the 5'-end (beginning) of the tammar *ELP* transcript. A 6% polyacrylamide/7 M urea gel in 1x TBE gel was pre-run at 250 V for 1 h at 4°C and the primer extension samples loaded and electrophoresed for 2.5 h at 4°C. The gel was transferred onto Chromatography Paper (Whatman), exposed to a Storage Phosphor screen, scanned (Typhoon 8600 Scanner, GE Healthcare) and analysed with ImageQuant software.

Creation of tammar *ELP* and murine *CSN2* d2EGFP and luciferase promoter reporter constructs

Tammar *ELP* promoter constructs were generated by PCR amplification of a modified pBeloBAC11-14.7kbtELP vector (Supplementary file 1) using proofreading Platinum *Taq* DNA Polymerase High Fidelity (11304-011, Invitrogen/Life Technologies) and gene-specific forward and reverse primers with *Xho*I and *Sac*II restriction enzyme sites incorporated respectively for subcloning purposes (Table 1). The vector insert comprised a 14,704 bp region of tammar *ELP* genomic DNA (-7,865/+6,839, +1 denotes the TSS of the short transcript). This included the *ELP* promoter (7,865 bp), the complete coding region of the 3-exon gene and a 2,531 bp 3' non-coding region. This 14.7 kb fragment was also used to create tammar *ELP* stable cell lines (Supplementary file 2). A lactogenic hormone-responsive murine *β-casein* (*CSN2*) promoter construct (-259/+7, numbering relative to the start site of transcription) (Kanai *et al.*, 1993) was also created by PCR amplification of murine genomic DNA (Table 1). All PCR products were cloned into the pGEM-T Easy vector

(Promega Corporation), plasmid DNA prepared (QIAprep Spin Miniprep Kit, Qiagen) and sequenced to verify promoter integrity. Promoter fragments were double-digested from pGEM-T Easy with *Xho*I/*Sac*II (NEB, New England Biolabs), with the exception of the 8 kb *ELP* construct. This plasmid was sequentially digested with *Sal*I (NEB) and then *Sac*II, creating a 7.949 kb *ELP* fragment with a *Xho*I-compatible 5'-end. All *ELP*/*CSN2* promoter fragments were then sub-cloned into a *Xho*I/*Sac*II digested promoterless pd2EGFP-1 vector (Clontech) (Supplementary file 3A). A 'functional' promoter results in the production of d2EGFP (destabilised enhanced green fluorescent protein) reporter.

The pGL3-Basic vector system (Promega Corporation), which has a firefly (*Photinus pyralis*) luciferase (*Luc*) reporter, was used to quantify tammar *ELP* and rodent *CSN2* promoter activity. All *ELP* constructs, with the exception of the 8.0 kb promoter, were sub-cloned from pd2EGFP-1 into the *Xho*I cut (5'-end), *Hind*III-digested and blunt-ended (3'-end) promoterless pGL3-Basic vector (Supplementary file 3B). Briefly, pGL3-Basic was *Hind*III digested (NEB), gel extracted (QIAquick Gel Extraction Kit, Qiagen) and the 5'-protruding-end filled in with *E. coli* DNA Polymerase I Large (Klenow) Fragment (NEB). The vector was dephosphorylated with shrimp alkaline phosphatase (NEB), column purified (QIAquick PCR Purification Kit, Qiagen), *Xho*I digested gel extracted and column purified again. Tammar *ELP* and murine *CSN2* promoter fragments were double-digested from their respective pd2EGFP-1 vectors with *Xho*I and *Sma*I and gel purified. However, the pd2EGFP-1 1.7 and 1.5 kb tammar *ELP* constructs, both of which had internal *Sma*I sites, were linearised with *Sac*II, the 3'-overhang blunt-ended with *E. coli* DNA Polymerase I Large (Klenow) Fragment and digested with *Xho*I. The 8 kb *ELP* promoter fragment in pGEM-T Easy was also linearised with *Sac*II, the 3'-overhang blunt-ended and then *Sal*I digested to create a *Xho*I-compatible 5'-end. All eleven promoter constructs were ligated at 16°C overnight into the 5'-*Xho*I cut and 3'-*Hind*III digested and blunt-ended pGL3-Basic vector and endotoxin-free plasmid DNA prepared (EndoFree Plasmid Maxi Kit, Qiagen). Two additional short tammar *ELP* promoter constructs (-85/+20 and -35/+20) were PCR-amplified from the -161/+20 tammar *ELP* pd2EGFP-1 plasmid (Table 1). PCR products were cloned into pGEMT-Easy, sequenced, *Xho*I/*Hind*III digested and cloned into the *Xho*I/*Hind*III digested pGL3-Basic vector. The *ELP* and *CSN2* pGL3-Basic promoter constructs were verified by *Xho*I/*Nco*I double digests and sequencing with RVprimer3: 5'-CTAGCAAAATAGGCTGTCCC-3', and/or GLprimer2: 5'-CTTTATGTTTTTGGCGTCTTCCA-3', Promega Corporation).

Table 1. Primers and conditions used to create the tammar *ELP* and murine *CSN2* promoter constructs

Promoter Construct (genomic region)*	Primer F (5' … 3')	Primer F (5' … 3')	PCR Product Size (bp)	PCR Conditions
8.0 kb t*ELP* (-7981/+20)*	**T7** (20 bp) TAATACGACTCACTATAGGG	**+23^_*Sac*II_tELP_R** (30 bp) GTG*CCGCGG*GTAGACTACACTTGATTAGTG	7981	94°C for 2 min; 35 cycles of [94°C for 30 s; 57°C for 30 s; 68°C for 8 min]; 68°C for 10 min.
1.7 kb t*ELP* (-1668/+20)*	**-1668^_*Xho*I_tELP_F** (30 bp) TGGT*CTCGAG*AGTCACAGTCAGATGCATGA	"	1706	94°C for 2 min; 35 cycles of [94°C for 30 s; 50°C for 30 s; 68°C for 2 min]; 68°C for 10 min.
1.5 kb t*ELP* (-1455/+20)*	**-1466^_*Xho*I_tELP_F** (30bp) GGA*CTCGAG*GTGCACTCAGAGAGATACAGC	"	1492	94°C for 2 min; 35 cycles of [94°C for 30 s; 50°C for 30 s; 68°C for 2 min]; 68°C for 10 min.
1.3 kb t*ELP* (-1255/+20)*	**-1270^_*Xho*I_tELP_F** (30bp) TGCA*CTCGAG*ACATTGCAGACTTTTTGTAA	"	1293	94°C for 2 min; 35 cycles of [94°C for 30 s; 50°C for 30 s; 68°C for 2 min]; 68°C for 10 min.
1.1 kb t*ELP* (-1072/+20)*	**-1084^_*Xho*I_tELP_F** (30bp) GTGC*CTCGAG*CCAAGGGTCTGGCAGAGTGC	"	1110	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 68°C for 1 min]; 68°C for 10 min.
0.9 kb t*ELP* (-876/+20)*	**-903^_*Xho*I_tELP_F** (30bp) GCATCCTG*CTCGAG*TTTAGACAAGAAGAGG	"	929	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 68°C for 1 min]; 68°C for 10 min.
0.7 kb t*ELP* (-712/+20)*	**-724^_*Xho*I_tELP_F** (30bp) GAGC*CTCGAG*GGGTTAGTAGGGCTGGGGCC	"	750	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 68°C for 1 min]; 68°C for 10 min.
0.5 kb t*ELP* (-542/+20)*	**-553^_*Xho*I_tELP_F** (30bp) CTG*CTCGAG*ACTTAGGCATCAGAGTGTCTT	"	579	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 68°C for 1 min]; 68°C for 10 min.
0.3 kb t*ELP* (-323/+20)*	**-341^_*Xho*I_tELP_F** (30bp) CCTTCAGAAG*CTCGAG*AGATGATGCCAGCT	"	367	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 68°C for 1 min]; 68°C for 10 min.
0.16 kb t*ELP* (-161/+20)*	**-176^_*Xho*I_tELP_F** (30bp) GGTTCCT*CTCGAG*AAAAACTTTTTGCCTGG	"	202	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 68°C for 1 min]; 68°C for 10 min.
0.085 kb t*ELP* (-85/+20)*	**-91^_*Xho*I_tELP_F** (23bp) TGACCA*CTCGAG*CACCTACCTTG	**+23^_*HindIII*_pGL3_tELP_R** (25 bp) GCC*AAGCTT*GGGTAGACTACACTTG	127	94°C for 2 min; 35 cycles of [94°C for 30 s; 56°C for 30 s; 68°C for 1 min]; 68°C for 10 min.
0.035 kb t*ELP* (-35/+20)*	**-38^_*Xho*I_tELP_F** (25bp) TTG*CTCGAG*GATATAAAACAAACCC	**+23^_*HindIII*_pGL3_tELP_R** (25 bp) GCC*AAGCTT*GGGTAGACTACACTTG	71	94°C for 2 min; 35 cycles of [94°C for 30 s; 56°C for 30 s; 68°C for 1 min]; 68°C for 10 min.
0.26 kb m*β-CSN* (-259/+7)*	**-265^_*Xho*I_mβ-CSN_F** (30bp) *CTCGAG*CTCTCACTTGGCTGGAGGAACATG	**+11_*Sac*II_mβ-CSN_R** (30bp) GCT*CCGCGG*ATGATGTCCTATCAGACTCTG	279	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 68°C for 1 min]; 68°C for 10 min.

* Gene promoter region of the construct (relative to the transcription start site, +1)
^ Position of the restriction enzyme site incorporated into the promoter construct

Other plasmids used for transfection experiments

The pCMV-d2EGFP-1 plasmid that constitutively expresses d2EGFP (under the control of the CMV cytomegalovirus promoter) was used as a positive transfection control and was a gift from Denijal Topcic (The University of Melbourne, Australia). The prCSN2(-344/-1)-GL3 rat *CSN2* promoter vector, which was derived from the pβc(-344/-1)CAT vector (Doppler *et al.*, 1990) was used as a positive control for the lactogenic hormone-responsiveness of transfected cells. The pGL3-Control vector was the positive control for luciferase activity and pSV-β-Galactosidase (Promega Corporation) the internal transfection control, while pRL-CMV (*Renilla* luciferase, Promega Corporation) was the internal transfection control for dual luciferase assays. The pECE-Nb2PRLR plasmid which constitutively expresses the Nb2 (intermediate) form of the prolactin receptor under the control of the SV40 early promoter (Ali *et al.*, 1992), was a gift from Dr Li-yuan Yu-Lee (Baylor College of Medicine, Houston, Texas, USA). The pXM-MGF (STAT5A) plasmid (Wakao *et al.*, 1994) which constitutively expresses ovine STAT5A, was a gift from Dr Bernd Groner (Institute for Biomedical Research, Frankfurt am Main, Germany).

Cell Culture, stable and transient transfections

Cell culture conditions and transfection protocols for the immortalised murine HC11 and KIM-2 mammary epithelial cell lines and the MDA-MB-231 breast cancer cell line, plus the method of extraction, culture and transfection of tammar wallaby primary mammary gland cells (PMGCs), a mixed population of cells isolated from mammary gland tissue at day 16 of pregnancy (Phase 1), are provided in Supplementary file 2.

HEK293T Cell Culture

Human embryonic kidney, or 293T cells (HEK293T), which stably express the SV40 large T antigen, were cultured in Dulbecco's Modified Eagle Medium (DMEM, 11-016-0500V, Thermo Electron Corporation, Australia), 10% heat-inactivated fetal bovine serum (FBS, 716717, CSL Biosciences, Australia), 2 mM L-glutamine (25030-081, Gibco), Penicillin-Streptomycin (100 IU/mL-100 µg/mL, 15040-066, Gibco) and 400 µg/mL Geneticin Selective Antibiotic (10131-027, Gibco) at 37°C in 5% CO_2.

Creation of prolactin-responsive HEK293T cells

The prolactin signalling pathway was created in 293T cells similar to the methods described for COS-7 and MCF-7 cells (Yamashita *et al.*, 2001). Culture dishes (24-well) were seeded 24 h prior to transfection at 2 x 10^5 cells per well, in 293T media (500 µL) without antibiotics. Cells (~95% confluent) were transfected with 1 µg DNA and 2 µL

Lipofectamine 2000 (Invitrogen) in Opti-MEM I reduced serum medium (Invitrogen) as per the manufacturer's instructions. Plasmid DNA transfected into cells included: pECE-Nb2PRLR (0.23 μg), pXM-MGF (0.23 μg), the tammar *ELP* or rat or mouse *CSN2* promoter reporter construct (0.46 μg) and the pSV-β-Galactosidase plasmid (0.09 μg) as an internal transfection control. The pGL3-Control and pGL3-Basic vectors were used as positive and negative controls of luciferase activity respectively. Complexes were removed 4 h post-transfection and replaced with fresh 293T media with reduced FBS (2%) and ovine prolactin (+/-5 μg/mL media, NIDDK-oPRL-21). Transgene expression was assayed 48 h post-transfection.

Minimal tammar ELP promoter required for transcription in HEK293T cells

Three separate experiments were carried out in triplicate to identify the minimal tammar *ELP* promoter required for gene transcription. Culture dishes (96-well) were seeded 24 h prior to transfection at 2 x 10^4 cells per well in 293T media (100 μL) without antibiotics. Cells were transfected with 50 ng plasmid DNA (*ELP* promoter construct) and ViaFect (0.2 μL, Promega) in Opti-MEM I according to the manufacturer's protocol. The pRL-CMV *Renilla* luciferase plasmid (0.5 ng) was co-transfected as an internal transfection control. Transgene expression was analysed 24 h post-transfection by dual luciferase assays.

Reporter assays: analysis of transgene expression

Detection of d2EGFP

Media was removed from adherent cells transfected with the pd2EGFP gene promoter constructs. Plates were inverted and viewed using a Leitz Wetzlar Dialux 20 microscope with a GFP filter and photographed with a Nikon Digital Sight DS-5M camera and imaging system (DS-L1), using variable exposure times.

Luciferase, β-Galactosidase and dual luciferase assays

Lysates of cells transfected with both the pGL3 and pSV-β-Galactosidase reporter constructs were prepared as described (Luciferase Assay System, Promega Corporation). However, 1X Passive lysis buffer [80 μL/400 μL per 24-well/6-well plate well respectively] was used to lyse cells. Luciferase activity was measured with the Lumat LB 9501 (Berthold Australia Berthold Pty Ltd) using a 2 s delay and then a 10 s measurement of luciferase activity. β-Galactosidase assays were performed in 96-well plates using the β-Galactosidase Enzyme Assay System (Promega Corporation) as described by the manufacturer's instructions and the absorbance (A_{420nm}) and hence β-Galactosidase activity of cell lysates measured with a microtitre plate reader

(Multiskan Ex, Thermo Electron Corporation, Finland).

The Dual Luciferase Reporter Assay System (Promega Corporation) was used to determine *ELP* and *CSN2* promoter activity (firefly luciferase) normalised to the *Renilla* luciferase internal transfection control as per the manufacturer's protocol, but with several modifications made. Media was aspirated from transfected cells under vacuum, 1X Passive lysis buffer (30 µL) was added and the plate incubated on an orbital shaker at 4°C for 30 min. Cell lysates (20 µL) were transferred to a 96-well LUMITRAC 200 white microtitre plate (Greiner), Luciferase assay reagent II (40 µL) added and firefly luciferase activity measured (1 s delay after injection, 5 s integration time) using the GloMax-96 Microplate Luminometer (Promega Corporation). Immediately after the 96-well plate was assayed for firefly luciferase, the *Renilla* luciferase assay was commenced. Stop & Glo Reagent (40 µL) was added to the Luciferase assay reagent II/cell lysate mix and *Renilla* activity measured (1 s delay after injection, 5 s integration time).

Statistical analyses

The statistical software packages R 3.2.1 (http://www.r-project.org/), JGR 1.7-16 (http://www.rforge.net/JGR/) and Deducer 0.7.7 (http://www.deducer.org) were used to analyse luciferase reporter assays using a one-way ANOVA (Generalised Linear Model), with a p-value < 0.05 considered to be statistically significant. Post-hoc analysis using Tukey's multiple comparison of means and confidence interval estimates (95% level) were used to determine luciferase activities and hence *ELP* promoter constructs that differed.

Cloning and *in silico* identification of marsupial *ELP* and eutherian *CTI* promoters

The fat-tailed dunnart and koala *ELP* and Cape fur seal *CTI* promoters were amplified by PCR from ~200 ng of genomic DNA (Sambrook and Russell, 2001) proofreading Platinum *Taq* DNA Polymerase High Fidelity, cloned into pGEM-T Easy, sequenced and submitted to GenBank (accession numbers are listed in Supplementary file 4). Primers and PCR conditions used are shown in Supplementary file 5. Promoter regions of *ELP* and *CTI* genes previously identified (tammar, dog and opossum (Pharo *et al.*, 2012)), plus *CTI* of the camel, killer whale and white rhinoceros were identified by BLAST searches of the NCBI GenBank nr database.

Prediction of putative *ELP* and *CTI* transcription factor binding sites

ELP and *CTI* promoters and genes were screened for conserved regions and consensus TFBSs using multiple online tools. These included the JASPAR transcription

factor binding profile database v5.0 alpha (http://jaspar.genereg.net/) (Mathelier *et al.*, 2014), TRANSFAC Professional Release 2015.1 and TRANSFAC Public v7 (http://www.gene-regulation.com/pub/databases.html). A minimum of 80% for relative profile score threshold (JASPAR) and core matrix similarity (TRANSFAC) were used. Conserved regions within the *ELP* and *CTI* promoters and genes were identified with mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml) and rVISTA (http://genome.lbl.gov/vista/rvista/submit.shtml) was used to predict conserved TFBSs. Putative binding sites were also identified by manual searches of sequences for published TFBSs and comparison of *ELP* and *CTI* genomic sequences and *ELP* and *CTI*-specific matrices to TFBS databases. PAZAR (http://www.pazar.info/), HOCOMOCO (http://autosome.ru/HOCOMOCO/), TOMTOM (http://meme-suite.org/tools/tomtom), FIMO (http://meme-suite.org/tools/fimo), P-MATCH v1.0 (http://www.gene-regulation.com/cgi-bin/pub/programs/pmatch/bin/p-match.cgi) and AliBaba v2.1 (http://www.gene-regulation.com/pub/programs/alibaba2/index.html) were also used to predict TFBSs. Putative TFBS sequence logos were generated with WebLogo v2.8.2 (http://weblogo.berkeley.edu/logo.cgi).

Putative TFBSs within genes with a similar expression profile to tammar *ELP*

Tammar genes with a similar microarray expression profile to *ELP*, and hence potentially similar regulatory mechanisms, were identified by k-means and hierarchical clustering (Cluster 3.0; http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm). The tammar wallaby microarray experiment characterised the expression of ~10,000 mammary ESTs in the virgin mammary gland and during pregnancy, lactation and involution and has been described previously (Pharo *et al.*, 2012; Sharp *et al.*, 2009). Promoter sequences of genes with an *ELP*-like expression profile were obtained by BLAST searches of a new draft of the tammar wallaby genome (version 1.3) and scanned for putative TFBS as previously described.

Results

Tammar *ELP* has two transcription start sites

In order to identify the tammar *ELP* TSS and hence design *ELP* promoter constructs to examine gene regulation *in vitro*, we performed primer extension analysis on mRNA (Fig. 1). Two different *ELP* TSSs were identified in transcripts from both Phase 1 and Phase 2A mammary gland total RNA (positive controls; Fig. 1A; Supplementary file 6). In contrast, *ELP* transcripts were not detected in the minus RNA or Phase 3 total RNA (negative controls). The long *ELP* TSS (transcript 1) commenced at guanine 96

(numbering based upon the start site of the *ELP* reverse primer) and was initiated by TATA box 1 (TATAATTT; Fig. 1B). In contrast, the short *ELP* TSS (transcript 2) started at cytosine 61 and was initiated by TATA box 2 (TATAAAAC; Fig. 1B). These results were consistent with ~95 sequenced tammar *ELP* mammary gland ESTs (Lefèvre *et al.*, 2007). Furthermore, the location of TATA boxes 1 and 2, -31 and -30 bp respectively upstream of the TSS (+1) within the initiator (Inr) element (i.e., YYANWYY, A=+1, Y=C/T, W=A/T) was characteristic to eutherian gene promoters (Juven-Gershon *et al.*, 2008). PhosphorImager analysis of the end-labelled *ELP* primer extension products indicated that reproductive stage did not alter *ELP* transcript abundance.

The short transcript was more abundant (74%) than the long transcript (26%) in both Phase 1 and Phase 2A (76% and 24% respectively) mammary tissue. Although tammar *ELP* has two different TSSs, translation of the long and short transcripts showed that they both encode ELP (data not shown).

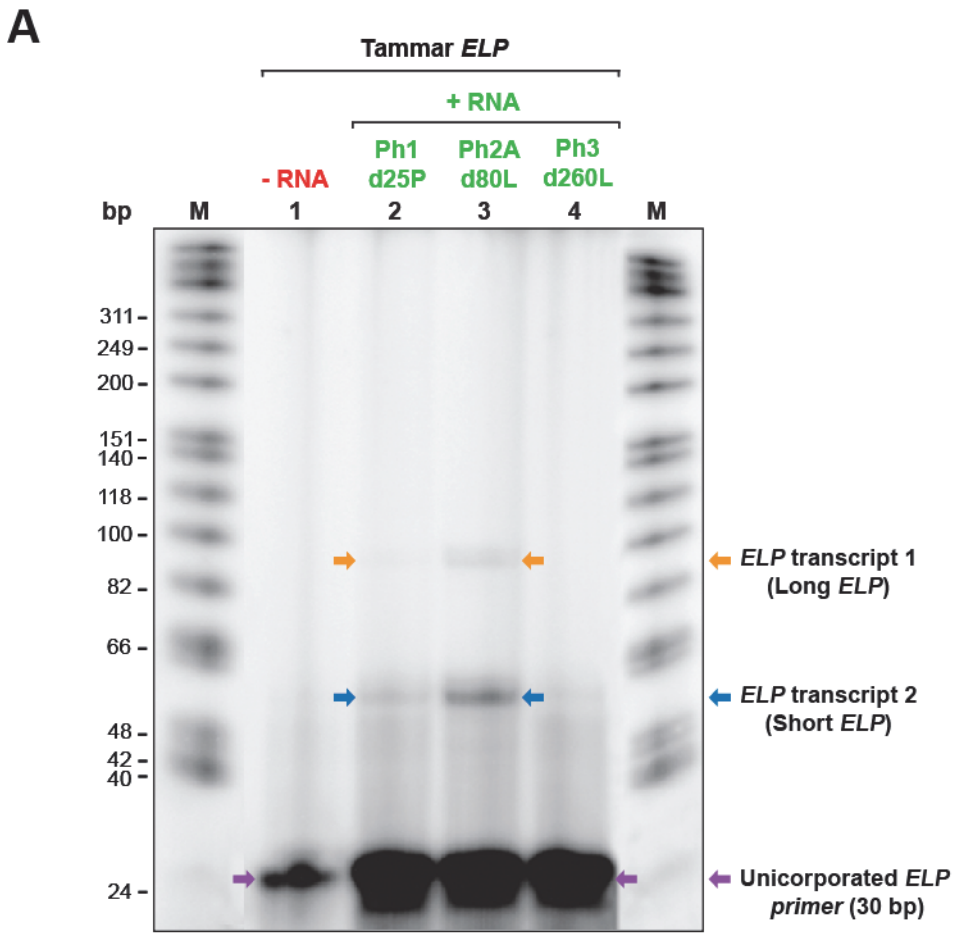
Identification of a putative milk protein gene-like composite response element (CoRE) within the tammar *ELP* promoter

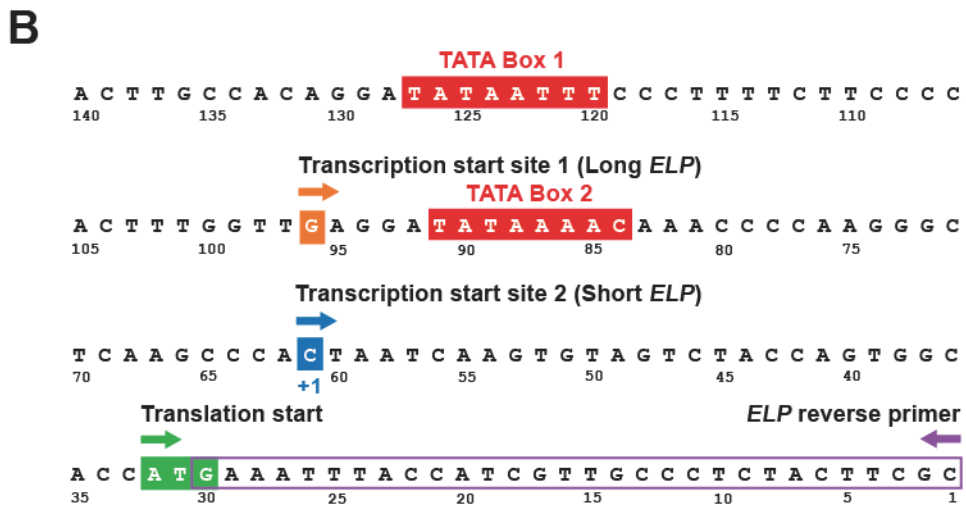
A bioinformatics approach was used to identify putative TFBSs within the -1678/+32 region of tammar *ELP*, i.e., 1678 bp upstream from the TSS (short transcript; Fig 2) and to design promoter constructs to investigate the mechanism(s) that regulate the temporal, early lactation-specific expression of tammar *ELP*. TFBSs associated with the mammary gland-specific activation and repression of milk protein genes (Malewski, 1998) as well as other regulatory elements, were identified within the *ELP* promoter (Fig 2) plus intragenic and flanking regions (Supplementary file 7). Although the transcriptional machinery, i.e. sequence recognition domains, usually scan genomic DNA from 5' to 3', putative TFBS on both the sense (+) and antisense (-) strands are shown because transcription factors that bind to genomic DNA usually have a dsDNA binding domain. Furthermore, many, but not all TFBS have a palindromic sequence motif and hence a strong match to the antisense strand suggests that the corresponding TFBS on the sense strand may be a site of lower binding affinity.

Interestingly, the proximal *ELP* promoter (-131/-113) included a putative 19 bp 'milk box'-like CoRE (Rosen *et al.*, 1999) with consensus binding sites for STAT/interferon gamma (IFNγ) activation site (GAS), CCAAT/enhancer binding protein (C/EBP), a glucocorticoid response element half site (½GRE), Yin-Yang 1 (YY1) and mammary cell-activating factor (MAF) (Fig. 2). The first four factors form the rat *CSN2* promoter lactation associated complex (-140/-110) that regulates the lactogenic hormone response

Fig. 1. Identification of two transcription start sites in the tammar *ELP* promoter

A. Primer extension analysis of the tammar *ELP* gene. A 30 bp [γ-^{32}P]-ATP end-labelled *ELP* reverse primer was used to prime first strand *ELP* cDNA synthesis from tammar mammary gland total RNA (10 µg) during pregnancy (Phase 1, d25P, day 25 of pregnancy) and lactation (Phase 2A, d80L, day 80 of lactation, and Phase 3, d260L). The negative control is also shown (Lane 2 –RNA). Two different tammar *ELP* primer extension products were detected, suggesting the use of two different TATA boxes. Transcript sizes were calculated based on the [γ-^{32}P]-ATP end-labeled φX174 HinfI DNA Markers (M). The original gel, which includes the kanamycin control, is provided (Supplementary file 6). **B.** The tammar *ELP* proximal promoter sequence and part of exon 1 are shown, including the *ELP* reverse primer and the two different *ELP* transcription start sites. Numbering below the sequence indicates primer extension product and hence transcript size.


A
Tammar ELP
+ RNA
- RNA
Ph1 d25P
Ph2A d80L
Ph3 d260L
bp
M
1
2
3
4
M
311
249
200
151
140
118
100
82
66
48
42
40
24
ELP transcript 1 (Long ELP)
ELP transcript 2 (Short ELP)
Unicorporated ELP primer (30 bp)


B
TATA Box 1
A C T T G C C A C A G G A T A T A A T T T C C C T T T T C T T C C C C
140 135 130 125 120 115 110
Transcription start site 1 (Long ELP)
TATA Box 2
A C T T T G G T T G A G G A T A T A A A A C A A A C C C C A A G G G C
105 100 95 90 85 80 75
Transcription start site 2 (Short ELP)
T C A A G C C C A C T A A T C A A G T G T A G T C T A C C A G T G G C
70 65 +1 60 55 50 45 40
Translation start
ELP reverse primer
A C C A T G A A A T T T A C C A T C G T T G C C C T C T A C T T C G C
35 30 25 20 15 10 5 1

of *CSN2* (Raught *et al.*, 1994). Although tammar *ELP* (-1678/+32) has four consensus STAT binding sites (Fig. 2), only one of these (-560/-552) is identical to the STAT5A consensus motif (TTCNNNGAA; N=any base) which is associated with milk protein gene expression (Rosen *et al.*, 1999). However, two consensus STAT5A TFBSs were located within the large 3,522 bp first intron.

Three additional promoter MAF consensus sites, GRRGSAAGK (R=A/G, S=G/C, K=G/T), plus two shorter CTTCCT(-, anti-sense strand)/AGGAAG(+, sense strand) MAF motifs were predicted, with 16 sites in intron 1, 2 in intron 2 and 1 in the 3'-UTR of exon 3. Notably, MAF directs the hormone-independent, mammary gland-specific expression of *WAP*, (Welte *et al.*, 1994b), mouse mammary tumour virus (*MMTV*) (Mink *et al.*, 1992) and *CSN2* (distal enhancer) (Winklehner-Jennewein *et al.*, 1998). MAF is a member of the E26 transformation-specific (ETS) family, 24 members of which are expressed in the normal murine mammary gland and bind to a GGA(A/T) core sequence (Coletta *et al.*, 2004; Galang *et al.*, 2004; Wei *et al.*, 2010).

Aside from MAF, multiple ETS TFBSs were predicted throughout the *ELP* gene and flanking regions, TFBS for the contactin/F11/ACAAAG gene enhancer, which acts synergistically with MAF and nuclear factor I (NFI) to activate mammary-specific gene expression (Mink *et al.*, 1992) were also located in the *ELP* promoter.

Four motifs for the ubiquitous C/EBP family were also predicted in the promoter, as well as 15 sites in intron 1. C/EBP α, β, and δ; the latter two factors also known as nuclear factor-interleukin 6 (NF-IL6) and NF-IL6β respectively, are encoded by different genes, but bind to the same consensus sequence (RTTGCGYAAY) (Osada *et al.*, 1996). C/EBP family members are associated with tissue-specific gene expression, proliferation and differentiation in the mammary gland and hence cell cycle control (Doppler *et al.*, 1995; Gigliotti and DeWille, 1998; Robinson *et al.*, 1998). Gene activation is dependent upon the formation of homo- and/or heterodimers. Alternate splicing of the *C/EBPβ* gene in the mammary gland produces three C/EBPβ isoforms, the ratio of which alters proliferation and differentiation (Zahnow *et al.*, 2001). C/EBPβ LIP (liver inhibitory protein) lacks the N-terminal trans-activation domain is associated with proliferation (Zahnow *et al.*, 2001). In contrast, C/EBPβ LAP1 (Liver-enriched Activator Protein) and C/EBPβ LAP2 are associated with differentiation and expression of milk protein genes.

Additional promoter TFBSs of note included one full glucocorticoid receptor (GR)-binding site (GGTACA$(N)_3$TGTTCT) within intron 1 and an additional six ½GRE sites

(TGTTCT, TGTNCY) within the promoter (Fig. 2). Ten whey protein promoter core sequence (WPPCS)-like elements similar to those in the *LALBA* (Lubon and Hennighausen, 1988) and *WAP* promoters (Kolb *et al.*, 1995; Kolb *et al.*, 1994) were also predicted in tammar *ELP*; five in the promoter, four in intron 1 and, one in the 3'-UTR of exon 3. Other putative TFBSs associated with milk protein gene expression included: specificity proteins 1 and 3 (SP1 and SP3), octamer 1 (OCT1), activating proteins 1 and 2 (AP1 and AP2), half sites for CCAAT-box-transcription factor (CTF)/NFI, estrogen response element (ERE), interferon response factor (IRF) and insulin response element (IRE). TFBS for nuclear factor of activated T-cells (NFAT), cAMP response element binding protein (CREB) and the MZF1 (myeloid zinc finger) were also predicted (Fig. 2, Supplementary file 7).

ELP repressor sites predicted included pregnancy-specific mammary nuclear factor (PMF) YY1 and nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) (Fig. 2). During pregnancy, *CSN2* transcription in the rodent mammary gland is suppressed by progesterone (mediated by PMF), YY1 and NF-κB (Doppler *et al.*, 2000; Raught *et al.*, 1994). At parturition, these repressors are removed and *CSN2* is expressed, while PMF mediates progesterone inhibition of murine *CSN2* transcription during pregnancy (Lee and Oka, 1992a, b), YY1 can act as either an enhancer or a repressor. During pregnancy, both suppress *CSN2* transcription in the rat and mouse mammary glands respectively.

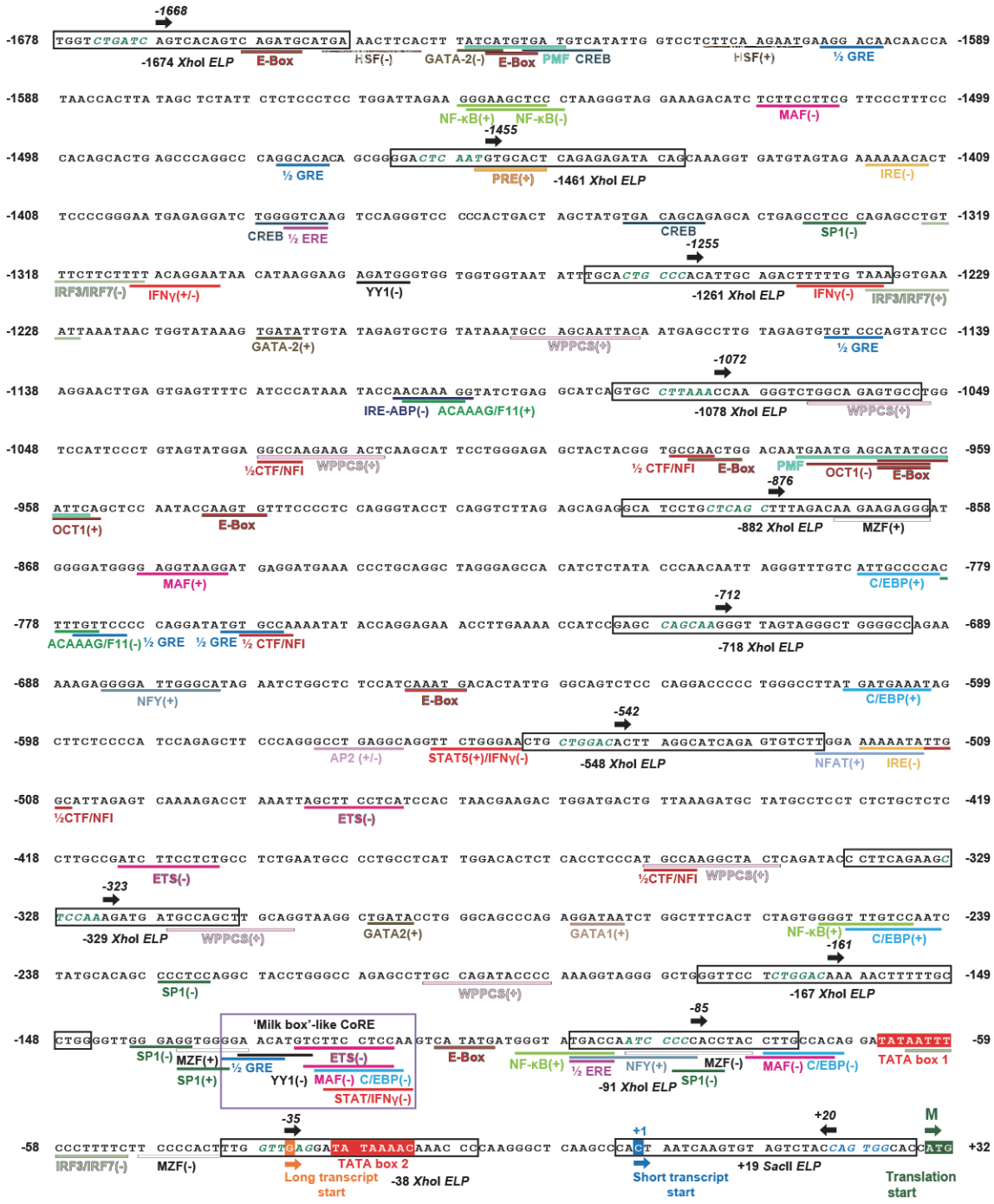
Tammar *ELP* is unresponsive to lactogenic hormones in 2D cell culture models

In order to investigate the transcriptional regulation of tammar *ELP in vitro*, twelve different tammar *ELP* promoter constructs were successfully created using the pd2EGFP-1 promoterless reporter vector (Fig. 2, Table 1). However, due to the lack of sensitivity and difficulty in measuring d2EGFP protein levels (and hence promoter activity), tammar *ELP* and rodent *CSN2* promoter constructs were subcloned into the pGL3-Basic promoterless reporter vector. Transient transfection of these constructs into murine HC11s and KIM-2s and treatment with lactogenic hormones showed that HC11s in our lab were not IDPRL-responsive unlike KIM-2s (Supplementary files 8 and 9). Although the murine *CSN2* promoter was responsive to lactogenic hormones in KIM-2s, unexpectedly, tammar *ELP* promoter constructs were not (Supplementary file 9).

Additionally, HC11s stably-transfected with a 14.7 kb (-7,865/+6839) genomic DNA fragment containing the tammar *ELP* gene (Supplementary file 2) were successfully produced, but neither *ELP*, nor the endogenous murine *CSN2* gene could be induced

Fig. 2. Putative transcription factor binding sites within the tammar wallaby *ELP* promoter and *ELP* promoter constructs

Putative consensus TFBS sites located on the sense (+) and antisense (-) strands of the -1,678/+32 bp region of the tammar *ELP* gene are shown (sequence numbering is relative to the short *ELP* TSS (+1)). The location of the 'milk box'-like CoRE (-131/-113) and other TFBS are also indicated. The locations of forward primers, plus the reverse primer used to amplify tammar *ELP* promoter constructs for reporter assays (Table 1), are shown (black outlined boxes). *Xho*I and *Sac*II restriction enzyme sites incorporated into the forward primers and the reverse primer are shown in italicised green and blue text respectively. Numbered black arrows indicate the start site of each *ELP* promoter construct. The two TATA boxes (red boxes) that initiate transcription of the long and short *ELP* transcripts are also shown.


-1668
-1678 TGGTCTGATC AGTCACAGTC AGATGCATGA AACTTCACTT TATCATGTGA TGTCATATTG GTCCTCTTCA AGAATGAAGG ACAACAACCA -1589
-1674 XhoI ELP
E-Box
HSF(-)
GATA-2(-)
E-Box
PMF
CREB
HSF(+)
½ GRE
-1588 TAACCACTTA TAGC TCTATT CTCTCCCTCC TGGATTAGAA GGGAAGCTCC CTAAGGGTAG GAAAGACATC TCTTCCTTCG TTCCCTTTCC -1499
NF-κB(+)
NF-κB(-)
MAF(-)
-1455
-1498 CACAGCACTG AGCCCAGGCC CAGGCACACA GCGGGGACTC AATGTGCACT CAGAGAGATA CAGCAAAGGT GATGTAGTAG AAAAAACACT -1409
½ GRE
PRE(+)
-1461 XhoI ELP
IRE(-)
-1408 TCCCCGGGAA TGAGAGGATC TGGGGTCAAG TCCAGGGTCC CCCACTGACT AGCTATGTGA CAGCAGAGCA CTGAGCCTCC CAGAGCCTGT -1319
CREB ½ ERE
CREB
SP1(-)
-1255
-1318 TTCTTCTTTT ACAGGAATAA CATAAGGAAG AGATGGGTGG TGGTGGTAAT ATTTGCACTG CCCACATTGC AGACTTTTTG TAAAGGTGAA -1229
IRF3/IRF7(-)
IFNγ(+/-)
YY1(-)
-1261 XhoI ELP
IFNγ(-)
IRF3/IRF7(+)
-1228 ATTAAATAAC TGGTATAAAG TGATATTGTA TAGAGTGCTG TATAAATGCC AGCAATTACA ATGAGCCTTG TAGAGTGTGT CCCAGTATCC -1139
GATA-2(+)
WPPCS(+)
½ GRE
-1072
-1138 AGGAACTTGA GTGAGTTTTC ATCCCATAAA TACCAACAAA GGTATCTGAG GCATCAGTGC CTTAAACCAA GGGTCTGGCA GAGTGCCTGG -1049
IRE-ABP(-)
ACAAAG/F11(+)
-1078 XhoI ELP
WPPCS(+)
-1048 TCCATTCCCT GTAGTATGGA GGCCAAGAAG ACTCAAGCAT TCCTGGGAGA GCTACTACGG TGCCAACTGG ACAATGAATG AGCATATGCC -959
½CTF/NFI
WPPCS(+)
½ CTF/NFI
E-Box
PMF
OCT1(-)
E-Box
-876
-958 ATTCAGCTCC AATACCAAGT GTTTCCCCTC CAGGGTACCT CAGGTCTTAG AGCAGAGGCA TCCTGCTCAG CTTTAGACAA GAAGAGGGAT -858
OCT1(+)
E-Box
-882 XhoI ELP
MZF(+)
-868 GGGGATGGGG AGGTAAGGAT GAGGATGAAA CCCTGCAGGC TAGGGAGCCA CATCTCTATA CCCAACAATT AGGGTTTGTC ATTGCCCCAC -779
MAF(+)
C/EBP(+)
-712
-778 TTTGTTCCCC CAGGATATGT GCCAAAATAT ACCAGGAGAA ACCTTGAAAA CCATCCGAGC CAGCAAGGGT TAGTAGGGCT GGGGCCAGAA -689
ACAAAG/F11(-)
½ GRE
½ GRE
½ CTF/NFI
-718 XhoI ELP
-688 AAAGAGGGGA TTGGGCATAG AATCTGGCTC TCCATCAAAT GACACTATTG GGCAGTCTCC CAGGACCCCC TGGGCCTTAT GATGAAATAG -599
NFY(+)
E-Box
C/EBP(+)
-542
-598 CTTCTCCCCA TCCAGAGCTT CCCAGGGCCT GAGGCAGGTT CTGGGAACTG CTGGACACTT AGGCATCAGA GTGTCTTGGA AAAAATATTG -509
AP2 (+/-)
STAT5(+)/IFNγ(-)
-548 XhoI ELP
NFAT(+)
IRE(-)
-508 GCATTAGAGT CAAAAGACCT AAATTAGCTT CCTCATCCAC TAACGAAGAC TGGATGACTG TTAAAGATGC TATGCCTCCT CTCTGCTCTC -419
½CTF/NFI
ETS(-)
-418 CTTGCCGATC TTCCTCTGCC TCTGAATGCC CCTGCCTCAT TGGACACTCT CACCTCCCAT GCCAAGGCTA CTCAGATACC CTTCAGAAGC -329
ETS(-)
½CTF/NFI
WPPCS(+)
-323
-328 TCCAAAAGATG ATGCCAGCTT GCAGGTAAGG CTGATACCTG GGCAGCCCAG AGGATAATCT GGCTTTCACT CTAGTGGGGT TTGTCCAATC -239
-329 XhoI ELP
WPPCS(+)
GATA2(+)
GATA1(+)
NF-κB(+)
C/EBP(+)
-161
-238 TATGCACAGC CCCTCCAGGC TACCTGGGCC AGAGCCTTGC CAGATACCCC AAAGGTAGGG GCTGGGTTCC TCTGGACAAA AACTTTTTGC -149
SP1(-)
WPPCS(+)
-167 XhoI ELP
-85
'Milk box'-like CoRE
-148 CTGGGGTTGG GAGGTGGGGA ACATGTCTTC CTCCAAGTCA TATGATGGGT ATGACCAATC CCCCACCTAC CTTGCCACAG GATATAATTT -59
SP1(-)
MZF(+)
SP1(+)
½ GRE
YY1(-)
ETS(-)
MAF(-) C/EBP(-)
STAT/IFNγ(-)
E-Box
NF-κB(+)
½ ERE
NFY(+)
MZF(-)
SP1(-)
MAF(-) C/EBP(-)
TATA box 1
-91 XhoI ELP
-35
+1
+20
M
-58 CCCTTTTCTT CCCCACTTTG GTTGAGGATA TAAAACAAAC CCCAAGGGCT CAAGCCCACT AATCAAGTGT AGTCTACCAG TGGCACCATG +32
IRF3/IRF7(-)
MZF(-)
TATA box 2
Long transcript start
-38 XhoI ELP
Short transcript start
+19 SacII ELP
Translation start

with IDPRL *in vitro* (data not shown), indicative of the loss of prolactin-responsiveness and hence the lactation phenotype. Furthermore, although tammar primary mammary gland cells (PMGCs) harvested on day 16 of pregnancy (Supplementary file 10) were responsive to IDPRL (e.g. expression of endogenous tammar *LGB*), endogenous tammar *ELP* was not (Supplementary file 11). This result was contradictory to endogenous *ELP* expression *in vitro* in Phase 1 (day 24 of a 26.5 day pregnancy) tammar mammary gland explants cultured for 8 days in media supplemented with IHCPRL versus insulin alone ($p<0.05$) (Pharo, 2014).

Tammar *ELP* is not responsive to PRL alone

As milk protein genes are generally responsive to PRL, i.e. the binding of STAT5A to the promoter initiates transcription (Wakao *et al.*, 1994; Welte *et al.*, 1994a), we tested whether tammar *ELP* promoter constructs are activated by the PRL-signalling pathway (and hence STAT5A binding) alone *in vitro*. We used HEK293Ts as they have excellent transfection efficiency, as high as 99% (Roy *et al.*, 1999).

Firstly, we established that HEK293Ts were not PRL-responsive (Fig. 3A). This was confirmed by the lack of response of transiently transfected rodent *CSN2* promoter constructs treated with prolactin (+PRL) compared to the no hormone treatment (-PRL). Similarly, there was no difference in luciferase activity of the PRL-treated or untreated tammar *ELP* promoter constructs tested (Fig. 3A).

To confirm whether the tammar *ELP* promoter constructs were PRL-responsive, we successfully created the JAK/STAT pathway in 293Ts by transient co-transfection with the plasmids constitutively expressing pECE-Nb2PRLR and pXM-MGF (STAT5A) and treatment with prolactin (Fig. 3B). Transient co-transfection with the rat or mouse *CSN2* promoter construct and treatment with PRL produced an average increase in luciferase activity of ~26-fold (rat) and ~37-fold (mouse) compared to the minus PRL treatments, $p<0.05$ (Fig. 3B). In contrast, neither the -1668/+20, -876/+20, nor the -323/+20 tammar *ELP*-pGL3 promoter constructs were PRL-responsive (inset, Fig. 3B).

Identification of the minimal tammar *ELP* promoter required for hormone-independent gene transcription

Without a suitable *in vitro* cell culture model to reproduce the IDPRL-responsiveness of tammar *ELP* in mammary gland explants, we used dual luciferase assays to determine the minimal promoter necessary to initiate *ELP* gene transcription in 293Ts (Fig. 4). The only promoter construct with a greater than one-fold change in luciferase activity relative to the promoterless pGL3-Basic vector was ptELP(-161/+20)-GL3,

$p<0.05$ (Fig. 4). In contrast, luciferase activity was reduced for the shorter ptELP(-85/+20)-GL3 and ptELP(-35/+20)-GL3 constructs, which contained both TATA boxes and TATA box 1 only, respectively. Increasing the *ELP* promoter length to -323/+20 and -542/+20 produced sequential reductions in luciferase activity. Likewise, as *ELP* promoter length increased (from -712/+20 to -1688/+20), promoter activity decreased. Similar trends in tammar *ELP* promoter activity were observed in duplicate experiments in the MDA-MB-231 immortalised breast cancer cell line (Supplementary file 12).

Identification of highly-conserved putative MAF/ETS, AP2γ and overlapping C/EBP and NF-KB sites within the proximal *ELP* and *CTI* promoters

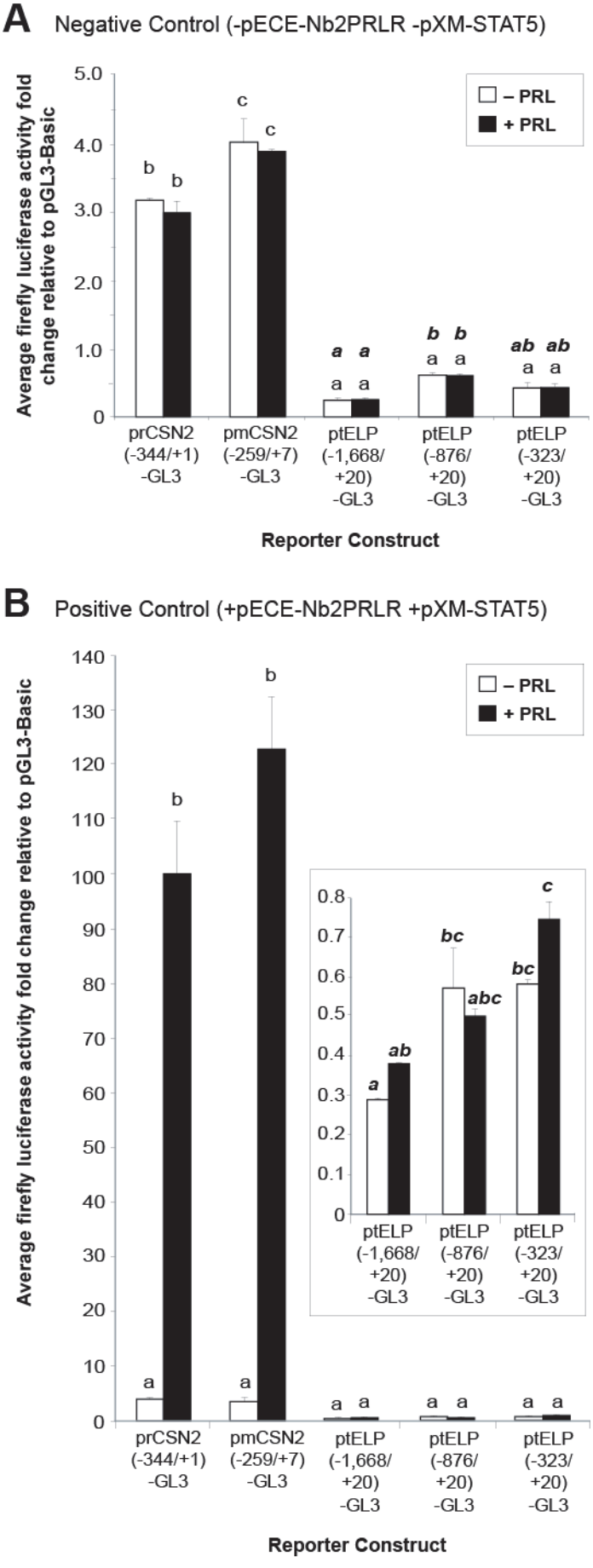
Tammar *ELP* promoter reporter assays indicated several regions within the proximal promoter where positive and negative regulatory elements may bind to and regulate gene transcription. However, alignment of 6 marsupial *ELP* proximal promoters showed the high homology (Supplementary file 13) between the sequences and thus, the inability to determine conserved regions.

It was only through a comparative genomics approach and the alignment of the proximal promoters of both of marsupial *ELP* (Supplementary file 13) and the orthologous eutherian *CTI* gene (Supplementary file 14) that eight conserved regions were identified (i-viii inclusive (red bars), selected promoters shown; Fig. 5). The complete alignment of 27 *ELP* and *CTI* proximal promoters, including the orthologous region of the human *CTI* pseudogene is provided (Supplementary file 15).

In silico searches of the JASPAR and TRANSFAC databases and published TFBS suggested these 8 conserved segments included: (i) estrogen-related receptor α/β (ERRα/β(+)); (ii) a low-affinity ecotopic virus integration site 1 (EVI(+)) site; (iii) highly-conserved, overlapping NF-κB(+) and C/EBP(+) sites; (iv) SP1/SP3(-); (v) an ETS-like(-) site; (vi) AP2γ(+); (vii) a highly-conserved MAF/ETS(-) motif with an overlapping low-affinity ERRα/β-like site (+), and (viii) a canonical TATA box (TATAAAAC) in marsupial *ELP*. In contrast, most eutherian *CTI* promoters (with the exception of the cow, yak, dog, Cape fur seal and walrus) had a low-affinity TATA motif (TATAAGT) (Fig. 5). However, an alternate *CTI* TATA Box (TAATGA(A/G)A) was predicted ~60 bp upstream and adjacent to the MAF/ETS site. While the TATA box, CCAAT box and GC box are characteristic to many promoters (Sandelin *et al.*, 2007), only the marsupial *ELP* promoters had a CCAAT box (NFY) ~55-65 bp upstream of the TSS and most promoters lacked a canonical GC box (SP1: GGGCGG), but did have low-affinity SP1 TFBSs. A highly-conserved marsupial *ELP*-specific motif

Fig. 3. Luciferase activity of tammar *ELP* promoter constructs in prolactin-responsive HEK293T cells

A. The PRL-signalling pathway is inactive in normal HEK293T cells, as there was no difference between the activity of -/+ PRL-treated (5 µg/mL) rat and mouse *CSN2* promoter constructs, prCSN2(-344/+1)-GL3 and pmCSN2(-259/+7)-GL3 respectively. There was also no differential expression of the -/+ PRL-treated tammar *ELP* promoter constructs ptELP(-1,668/+20)-GL3, ptELP(-876/+20)-GL3 and ptELP(-323/+20)-GL3. Data depicted shows the fold change of average luciferase activity for each promoter construct compared to the pGL3-Basic promoterless vector ± SEM for duplicate wells. Different letters indicate luciferase (promoter) activities that differ significantly ($p<0.05$) between the *CSN2* and *ELP* promoter constructs. Different italicised letters indicate significantly different *ELP* promoter activity ($p<0.05$). **B.** The PRL-signalling pathway was successfully created in HEK293T cells by transient co-transfection of plasmids constitutively expressing the Nb2-PRLR (pECE-Nb2PRLR) and STAT5 (pXM-MGF). The PRL-responsiveness of these cells was confirmed by the significantly higher fold change in luciferase activity for the PRL-treated (5 µg/mL) prCSN2(-344/+1)-GL3 and pmCSN2(-259/+7)-GL3 constructs compared to those not treated with PRL ($p<0.05$). In contrast, none of the *ELP* promoter constructs were responsive to PRL. Inset: Fold change in luciferase activity of the tammar *ELP* promoter constructs only.


A
Negative Control (-pECE-Nb2PRLR -pXM-STAT5)
Average firefly luciferase activity fold change relative to pGL3-Basic
5.0
4.0
3.0
2.0
1.0
0
– PRL
+ PRL
b b
c c
a a
a a
b b
a a
ab ab
a a
prCSN2 (-344/+1) -GL3
pmCSN2 (-259/+7) -GL3
ptELP (-1,668/ +20) -GL3
ptELP (-876/ +20) -GL3
ptELP (-323/ +20) -GL3
Reporter Construct
B
Positive Control (+pECE-Nb2PRLR +pXM-STAT5)
Average firefly luciferase activity fold change relative to pGL3-Basic
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
– PRL
+ PRL
b
b
a
a
a a
a a
a a
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
a
ab
bc
abc
bc
c
ptELP (-1,668/ +20) -GL3
ptELP (-876/ +20) -GL3
ptELP (-323/ +20) -GL3
prCSN2 (-344/+1) -GL3
pmCSN2 (-259/+7) -GL3
ptELP (-1,668/ +20) -GL3
ptELP (-876/ +20) -GL3
ptELP (-323/ +20) -GL3
Reporter Construct

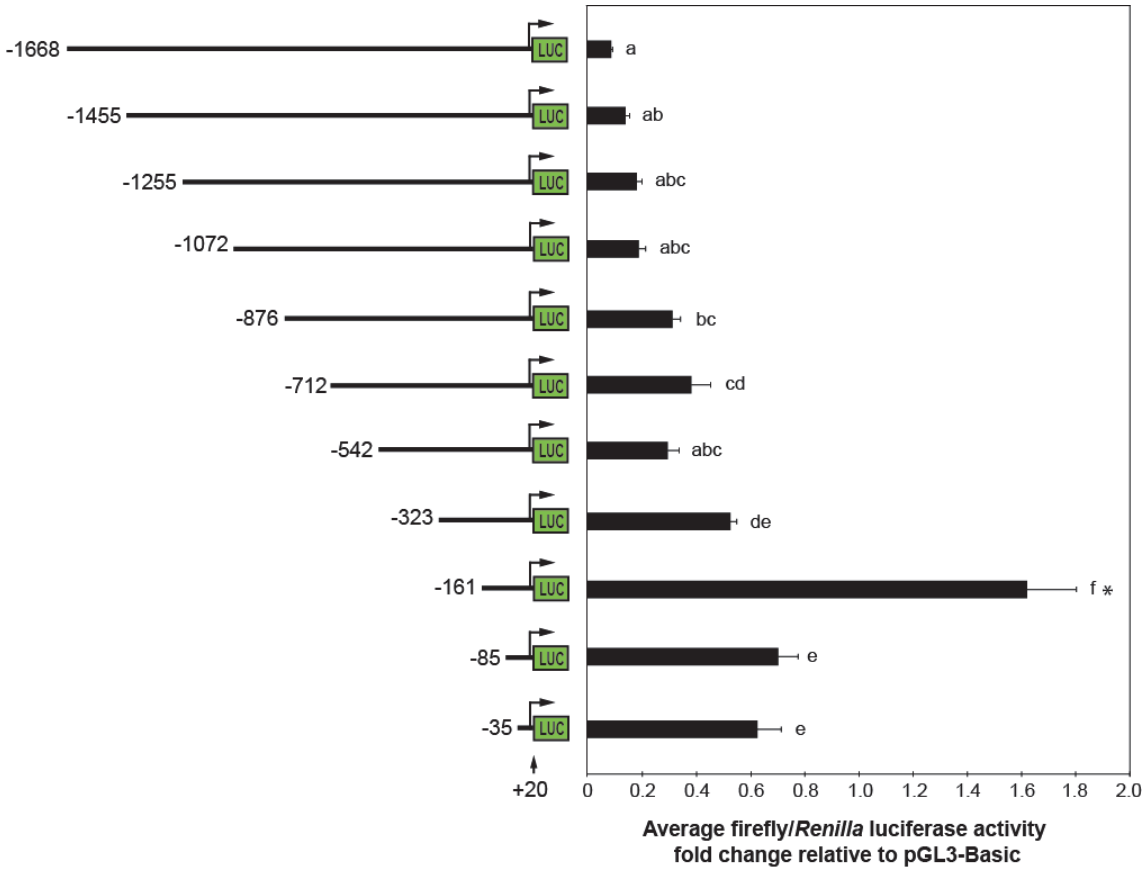


Fig. 4. Hormone-independent activity of tammar *ELP* promoter constructs in HEK293T cells

293Ts were transiently co-transfected with tammar *ELP* promoter constructs in the firefly luciferase pGL3-Basic plasmid, plus the *Renilla* luciferase pRL-CMV plasmid as an internal transfection control. Normalised firefly luciferase activity was highest for the ptELP(-161/+20)-GL3 construct ($p<0.05$), the only vector with a fold change greater than one. The significant differences between the activities of the -161/+20 and -323/+20 and -323/+20 and -542/+20 *ELP* promoter constructs suggested that putative repressor element(s) are located within both the -323/-162 and -542/-323 regions of tammar *ELP*. Data shown represents the average fold change of firefly/*Renilla* luciferase activity relative to pGL3-Basic activity ± SEM (n=3) for three identical experiments.

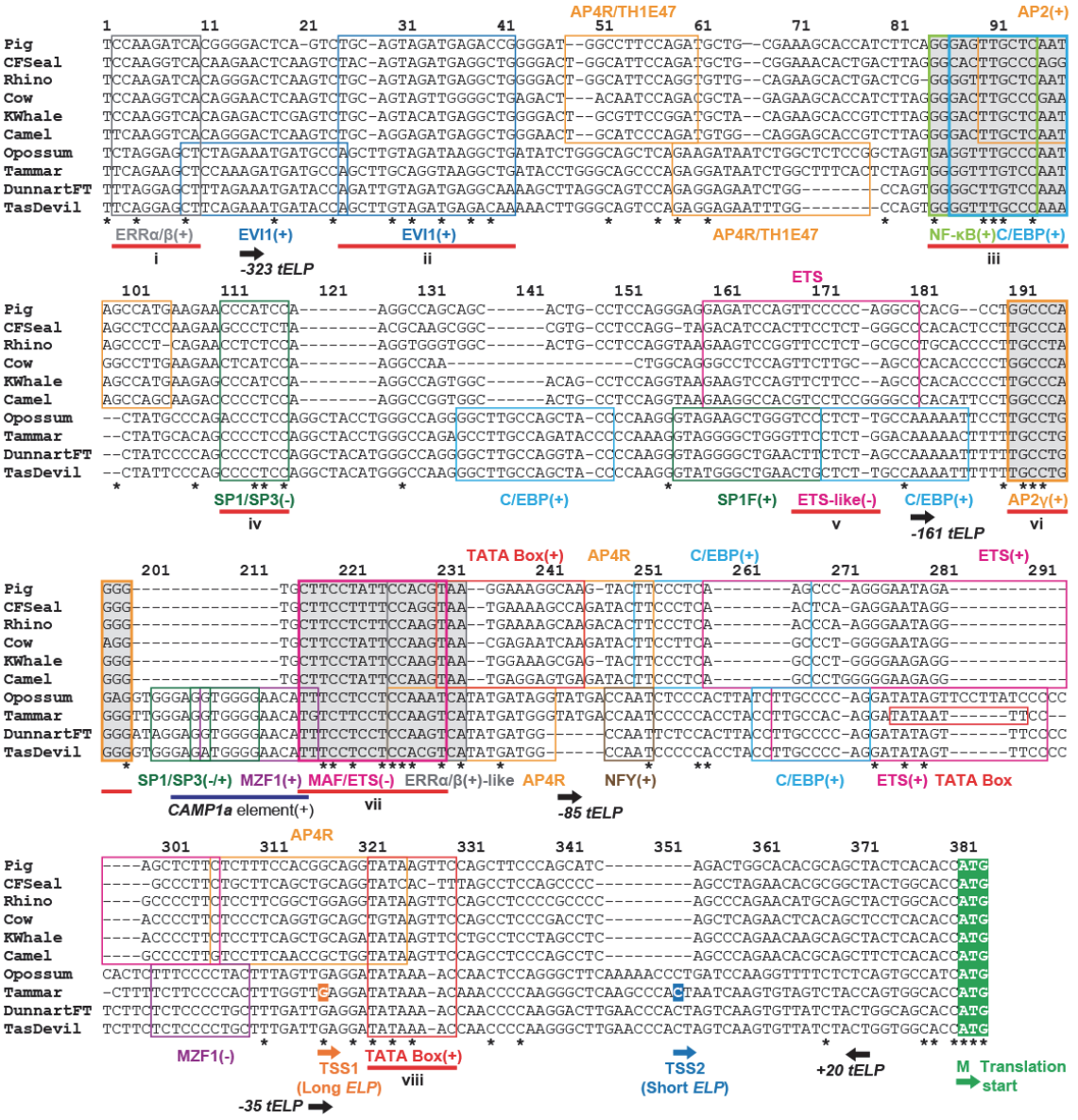


Fig. 5. Conserved transcription factor binding sites within the marsupial *ELP* and eutherian *CTI* proximal promoters

Eight conserved regions (i-viii, red bars) within the tammar *ELP* proximal promoter were identified by comparing selected marsupial *ELP* and eutherian *CTI* promoters. Three of these regions were highly-conserved (shaded grey) and included (iii) overlapping NF-κB and C/EBP sites, (vi) AP2γ and (vii) a MAF/ETS site which is partially overlapped by an ERRα/β-like motif. The tammar *ELP* TATA box 1 (long transcript), TATA box 2 and a low-affinity eutherian *CTI* TATA binding protein site (nt 231-245) are also noted. *ELP*-specific sites include MZF1 and NFY, plus a *CAMP1a*, SP1/SP3(-/+)-like element conserved in both the tammar *CAMP1a* and *ELP* promoters (dark blue bar). Tammar *ELP* TSS1 and TSS2 are indicated, and numbered black arrows show the respective start sites of the tammar *ELP* promoter constructs. Asterisks indicate nucleotides common to all sequences and alignment numbering is shown above sequences.

containing overlapping SP1/SP3(-/+) TFBS was also identified (5'-(A/G)GAG(G/A)TGGGGAACAT-3', nt 214-228, alignment numbering, Fig. 5). This sequence is also present in the tammar *CAMP1a* promoter (-325/-315 upstream from the translation start site, data not shown).

Comparison of several highly-conserved motifs identified within the *ELP* and *CTI* promoters (overlapping NF-κB and C/EBP sites, AP2γ and MAF/ETS, grey shaded boxes, Fig. 5) to publically available databases using Tomtom confirmed their identities ($p<0.001$, $p<0.05$) (Supplementary file 16). WebLogo representations of these motifs shows their high conservation within the marsupial *ELP* and eutherian *CTI* promoters (Fig. 6).

Conserved distal and intragenic elements in the *ELP* and *CTI* promoters

Regulatory elements are not only located within the proximal promoters of milk protein genes but are also found in distal and intragenic regions (Lemay *et al.*, 2013; Qian and Zhao, 2014; Rijnkels *et al.*, 2013). Therefore, we compared the genes and flanking regions of *ELP* and *CTI* in six species (tammar, camel, dog, killer whale, opossum and white rhinoceros) using mVISTA and rVISTA (Fig. 7). Conserved ETS, C/EBP, NF-κB, and EVI1 TFBS were predicted within intragenic and 5'- and 3'-flanking regions of all 6 species.

Table 2. Genes with a similar expression profile to *ELP* as identified by cluster analysis of microarray expression data of the tammar wallaby mammary gland during the reproductive cycle*

Gene	Abbreviation
Cathelicidin antimicrobial peptide 2	*CAMP2*
Ganglioside activator	*GM2A*
Cerebral endothelial cell adhesion molecule 1	*CERCAM1*
Platelet-derived growth factor alpha	*PDGFA*
Thioredoxin	*TXN*
Stearoyl-CoA desaturase	*SCD*
Six transmembrane epithelial antigen of the prostate 1	*STEAP1*
Dual specificity phosphatase 5	*DUSP5*
Tumour protein P53 inducible nuclear protein 1	*TP53INP1*

*(Pharo *et al.,* 2012; Sharp *et al.,* 2009)

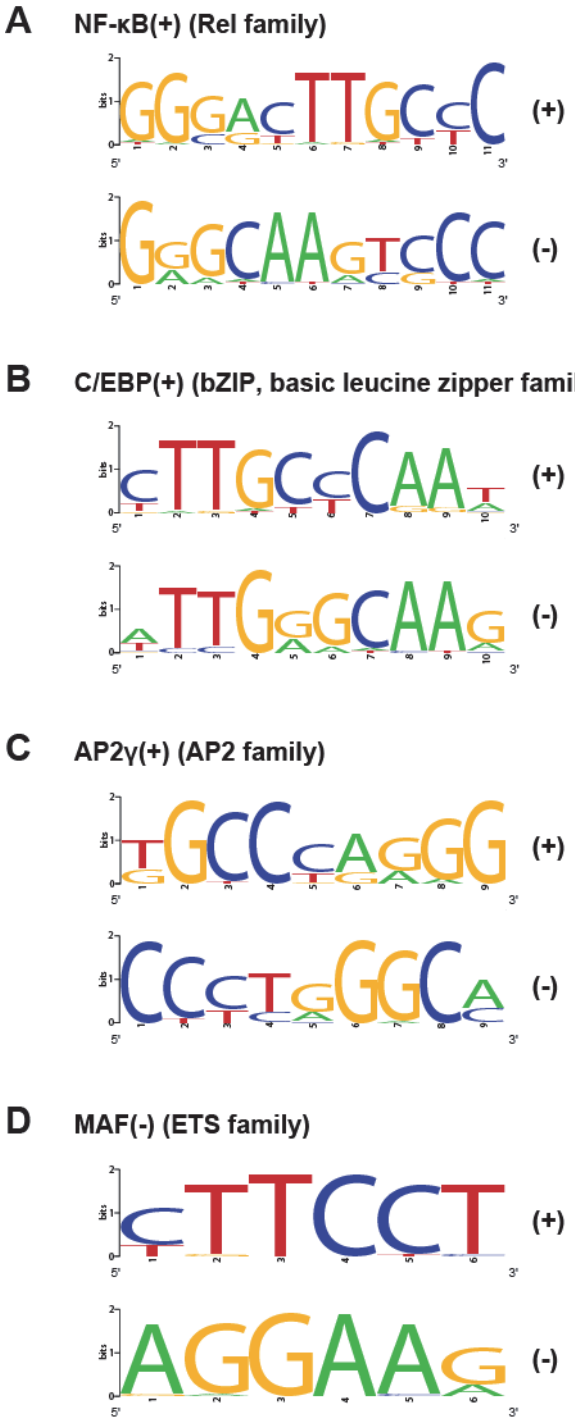


Fig. 6. Highly conserved putative *ELP*- and *CTI*-specific proximal promoter TFBS

Motifs for the highly-conserved putative **A.** NF-κB(+), **B.** C/EBP(+), **C.** AP2γ(+) and **D.** MAF/ETS(-) TFBS were generated with WebLogo. The motifs for both the sense (+) and antisense (-) strands are shown.

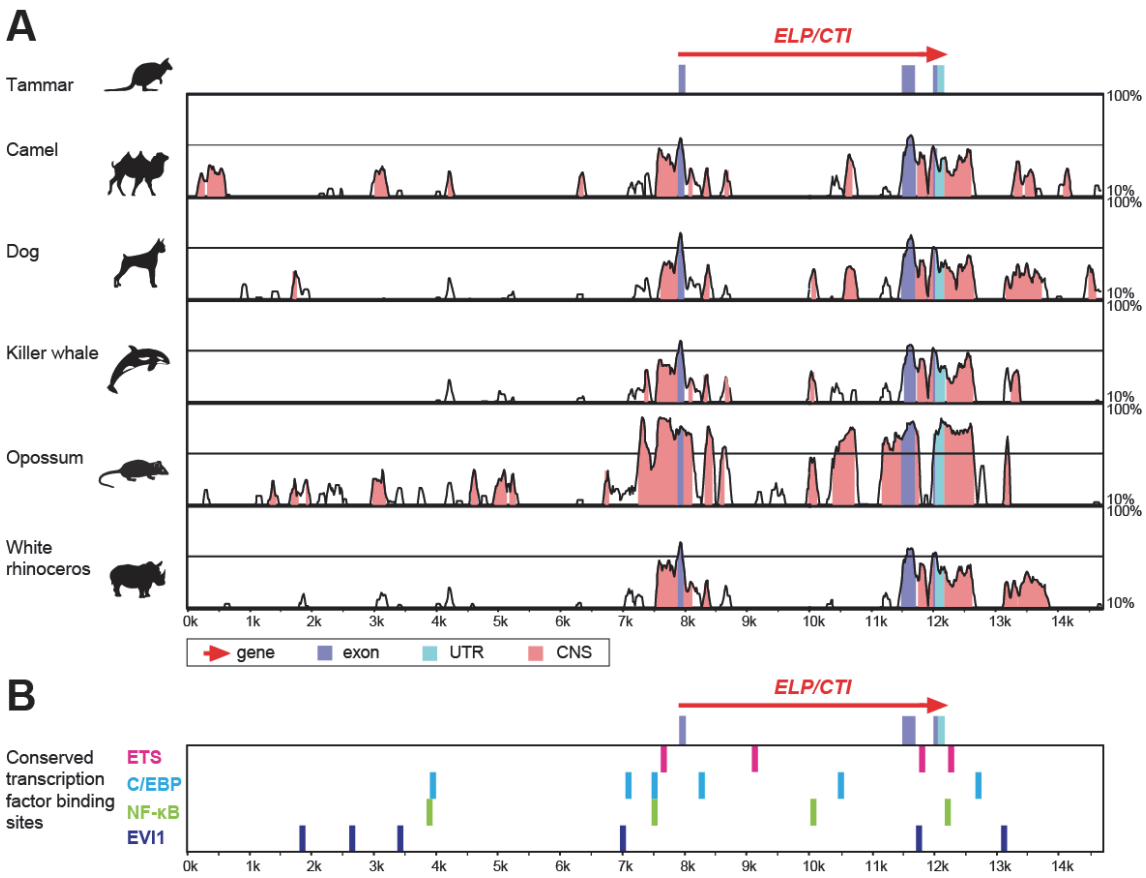


Fig. 7. VISTA plot of pairwise alignments for selected marsupial *ELP* and eutherian *CTI* genes and flanking regions

A. Sequence homology within the tammar, camel, dog, killer whale, opossum and white rhinoceros *ELP* or *CTI* genes and 5' and 3' flanking regions was determined with mVISTA. Tammar *ELP* was used as the reference sequence (horizontal axis, kb). Red horizontal arrows indicate gene location and direction of transcription. Purple rectangles indicate coding exons and blue-green rectangles, UTRs. The right vertical axis indicates % identity within a 100 bp window for each pairwise comparison, ranging from 10% to 100%. Regions sharing greater than 25% identity are shaded pink and the black horizontal line indicates 70% identity. **B.** The location of conserved putative ETS, C/EBP, NF-κB and EVI1 TFBS within and flanking the *ELP* and *CTI* genes in all species are shown.

Conserved TFBS in Phase 2A-specific tammar gene promoters

To test whether genes with similar expression profile share the same regulatory mechanisms (Eisen *et al.*, 1998), we used tammar mammary gland microarrays (Pharo *et al.*, 2012; Sharp *et al.*, 2009) to identify genes that cluster with *ELP* (Table 2). Conserved TFBS were identified within promoter sequences extracted from a new tammar genome assembly. All gene promoters except *CERCAM1* had at least one overlapping NF-κB and C/EBP site and all had multiple ETS and C/EBP sites (data not shown).

Discussion

While tammar *ELP* expression is upregulated at parturition *in vivo* (Pharo *et al.*, 2012) and responds to lactogenic hormones in 3D mammary gland explants *in vitro* (Pharo, 2014; Simpson, 1998), the lack of *ELP* expression in 2D cell culture models (this study) confirmed that transcription of the *ELP* gene is complex and strictly controlled. The identification of the minimal tammar *ELP* promoter necessary for hormone-independent gene activation and conserved TFBS within the proximal promoter provided a better understanding of factors that may regulate *ELP* including MAF, AP2γ, C/EBP, NF-κB, ERRα/β and EVI1.

Activation and repression of the tammar *ELP* gene during the reproductive cycle

Though *ELP* is responsive to IHCPRL in the 3D mammary explant culture system, the *in vitro* response is contingent upon *ELP* expression *in vivo* (Pharo, 2014; Simpson, 1998). Similarly, the lack of *ELP* induction in 2D cell culture models (this study) suggests that transcriptional and/or other mechanisms regulate tammar *ELP*. Promoter reporter assays and *in silico* comparative genomics analyses suggested several possible models for the regulation of tammar *ELP* in the mammary gland throughout the lactation cycle and in other tissues (Fig. 8). In particular, the Phase- and tissue-specific expression of tammar *ELP* suggests that local and/or global epigenetic mechanisms may be involved. In the mouse, developmental stage- and tissue-specific chromatin organisation is an important regulator of mammary gland genes and their regulatory elements (Lemay *et al.*, 2013; Rijnkels *et al.*, 2013; Rijnkels *et al.*, 2010).

The absence of *ELP* expression in the virgin mammary gland, during involution and in other tammar tissues suggests that during these periods, the tammar *ELP* gene is within a region of 'closed', condensed chromatin (heterochromatin), i.e. genomic DNA is wrapped around histones which are tightly-packed and therefore inaccessible to

transcription factors (Fig. 8B (i)). Hence, chromatin conformation-controlling promoter and distal regulatory elements (DREs) presumably silence the *ELP* gene. Gene silencing is characterised by methylated DNA within the gene promoter region, in particular, di-methylation of lysine 9 on histone H3 (H3K9me2) and tri-methylation of lysine 27 on histone H3 (H3K27me3) (Pekowska *et al.*, 2011; Rijnkels *et al.*, 2013; Wang *et al.*, 2014). The absence of DNase I hypersensitivity sites (DHSs), i.e. bound transcription factors, and/or the lack histone H3 acetylation (H3Ac) are also indicators of gene silencing (Rijnkels *et al.*, 2013; Shahbazian and Grunstein, 2007).

During pregnancy (Phase 1), *ELP* is expressed at low-levels (Fig. 8A). Therefore, chromatin in the promoter region is presumably open and the gene-activating H3K4me3 (Pekowska *et al.*, 2011) may be attached (Fig. 8B (ii)) Although accessible to transcription factors, additional levels of redundancy may prevent high-level *ELP* expression during this period. *ELP* may suppressed by the binding of NF-κB to the overlapping C/EBP/NF-κB site within the proximal promoter. A similar mechanism suppresses *CSN2* transcription during murine pregnancy and involves the binding of an NF-κB p50/p65 to an overlapping NF-κB/STAT5/C/EBP site in the *CSN2* promoter (Doppler *et al.*, 2000; Geymayer and Doppler, 2000; Zhang *et al.*, 2004). Alternatively, C/EBPβ may suppress *ELP* during pregnancy as a C/EBPβ LAP:LIP ratio less than five suppresses *CSN2* expression during pregnancy (Dearth *et al.*, 2001; Raught *et al.*, 1995; Rosen *et al.*, 1999).

ELP gene expression and indeed, the tammar mammary gland may also be controlled by circadian clock proteins. In the mouse mammary gland, the change in physiological state from late pregnancy to early lactation is correlated with distinct changes in molecular clock factors (Casey *et al.*, 2014; Metz *et al.*, 2006). This *in vivo* response can be mimicked *in vitro* by treating proliferating HC11 cells with lactogenic hormones to induce MEC differentiation (Casey *et al.*, 2014; Metz *et al.*, 2006). Hence, tammar *ELP* transcription may be suppressed by molecular clock factors during pregnancy. Furthermore, the ability of ETS factors to act as both transcriptional activators and suppressors and the considerable overlap in the DNA-binding specificity of ETS proteins to the GGA(A/T) core (Sementchenko and Watson, 2000; Sharrocks, 2001; Wei *et al.*, 2010) suggest than an ETS repressor may prevent MAF binding.

Tammar *ELP* is induced at parturition (Pharo *et al.*, 2012) as secretory activity begins (Findlay, 1982) and expressed for 100-125 days pp (Simpson *et al.*, 1998). This period coincides with the permanent attachment of the pouch young to its mother's teat,

its' inability to mount a specific-immune response (Fig. 8A) and key developmental milestones (Nicholas *et al.*, 1997; Pharo *et al.*, 2012; Simpson, 1998; Tyndale-Biscoe and Renfree, 1987). High expression levels of *ELP* are presumably facilitated by 'open' chromatin (euchromatin) which is loosely-packed, a high H3K4me3:H3K4me1 ratio and the binding of H3K4me3 and H3K4me2 to DRE(s) and/or the *ELP* promoter, plus H3Ac and an increase in DNase I hypersensitivity sites (DHSs), as for murine milk protein genes during lactation (Rijnkels *et al.*, 2013; Rijnkels *et al.*, 2010). Additionally, the recruitment of RNA polymerase II and basal transcription machinery, plus the binding of AP2γ and MAF to the TGCCTGGGG(+) and CTTCCT(-) sites respectively within the -161/-86 region of tammar *ELP* are likely to activate gene transcription. As for murine *CSN2* regulation, a postpartum reduction of the NF-κB p50/p65 repressor (Doppler *et al.*, 2000) and the binding of C/EBPβ LAP (Qian and Zhao, 2014) to the overlapping NF-κB/C/EBP site may also enhance Phase 2A *ELP* expression; a C/EBPβ LAP:LIP ratio>100 favours milk protein gene transcription (Dearth *et al.*, 2001; Raught *et al.*, 1995; Rosen *et al.*, 1999).

During early Phase 2B, around 114-127 days pp (Elizabeth Pharo, unpublished data), *ELP* expression ceases, with the protein barely detectable in milk by day 129 pp (Simpson, 1998). The immunocompetent pouch young becomes intermittently attached to its mother's teat (Nicholas *et al.*, 1997; Tyndale-Biscoe and Renfree, 1987), and there are large increases in mammary gland alveoli diameter (Findlay, 1982) as milk production is boosted for the developing young (Nicholas *et al.*, 1997). During mid- and late lactation, involution and the remodelling of the mammary gland, closed chromatin most probably suppresses *ELP* transcription, with NF-κB, C/EBPα and/or ETS suppressors providing redundancy to the system. Pregnancy then triggers chromatin to become open once again.

In contrast, chromatin is seemingly closed in the mammary gland of non-pregnant tammars. This is supported by the lack of response of milk protein genes to lactogenic hormones in explants harvested from this stage (Kevin Nicholas, personal communication). Murine milk protein genes expressed throughout lactation (*CSN1*, *CSN2*, *CSN3*, *LGB* and *LALBA*) are associated with open chromatin (Rijnkels *et al.*, 2013; Rijnkels *et al.*, 2010) and this presumably so for the tammar. However, the Phase 2B-early Phase 3-specific expression of *WAP* (Simpson *et al.*, 2000) and late 2B-Phase 3-specific transcription of *LLPA* and *LLPB* (Trott, 1999) suggests that Phase-specific chromatin structure is an important mechanism of marsupial milk protein gene regulation. For eutherians, different chromatin conformations during the colostrum and

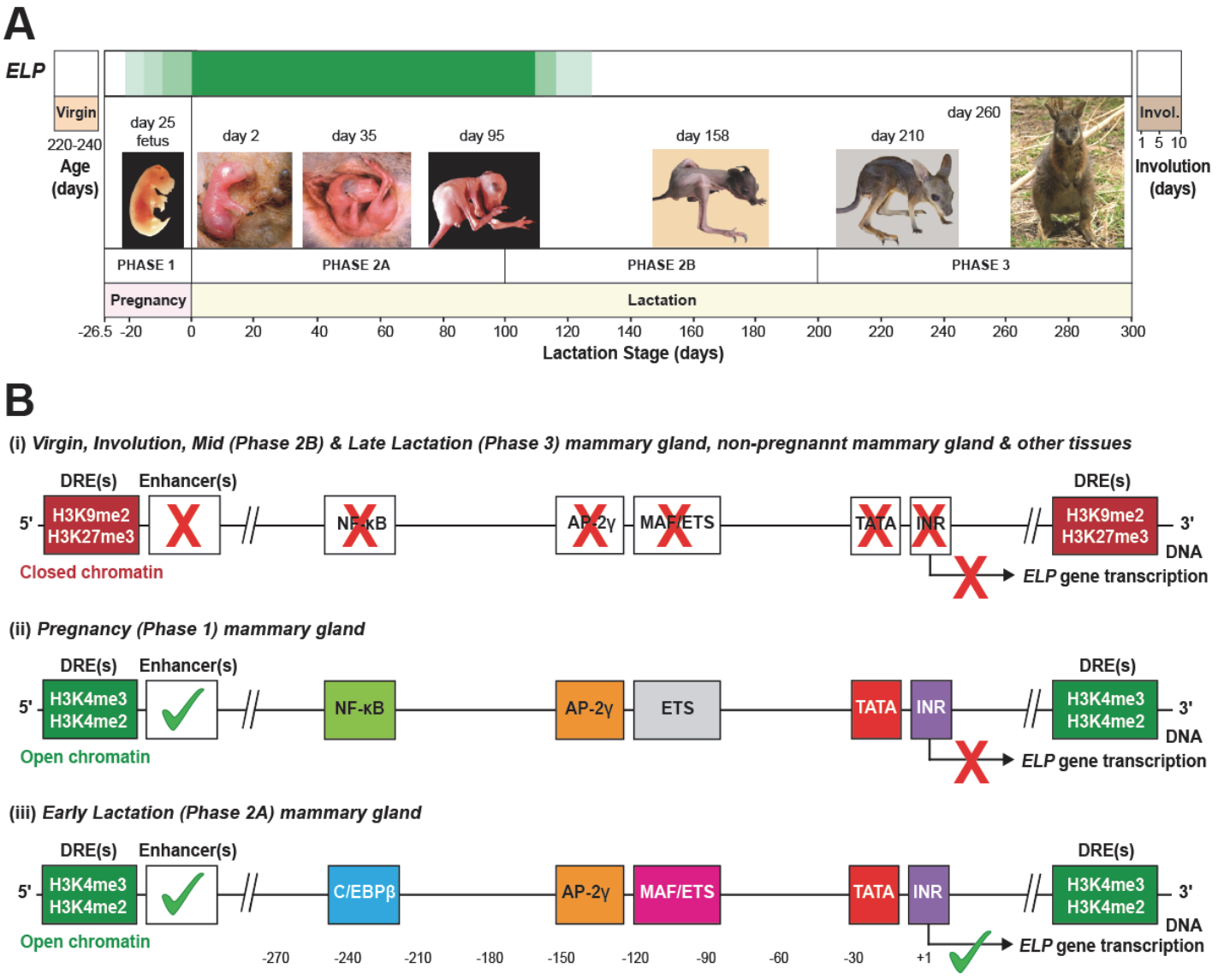

A
ELP
Virgin
220-240
Age
(days)
day 25
fetus
day 2
day 35
day 95
day 158
day 210
day 260
Invol.
1 5 10
Involution
(days)
PHASE 1
PHASE 2A
PHASE 2B
PHASE 3
Pregnancy
Lactation
-26.5 -20 0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300
Lactation Stage (days)
B
(i) Virgin, Involution, Mid (Phase 2B) & Late Lactation (Phase 3) mammary gland, non-pregnannt mammary gland & other tissues
DRE(s)
Enhancer(s)
5'
H3K9me2
H3K27me3
NF-κB
AP-2γ
MAF/ETS
TATA
INR
DRE(s)
H3K9me2
H3K27me3
3'
DNA
Closed chromatin
ELP gene transcription
(ii) Pregnancy (Phase 1) mammary gland
DRE(s)
Enhancer(s)
5'
H3K4me3
H3K4me2
NF-κB
AP-2γ
ETS
TATA
INR
DRE(s)
H3K4me3
H3K4me2
3'
DNA
Open chromatin
ELP gene transcription
(iii) Early Lactation (Phase 2A) mammary gland
DRE(s)
Enhancer(s)
5'
H3K4me3
H3K4me2
C/EBPβ
AP-2γ
MAF/ETS
TATA
INR
DRE(s)
H3K4me3
H3K4me2
3'
DNA
Open chromatin
ELP gene transcription
-270 -240 -210 -180 -150 -120 -90 -60 -30 +1

Fig. 8. Putative regulatory mechanisms involved in the regulation of the tammar *ELP* gene

A. Stylised diagram of *ELP* expression and pouch young development during the lactation cycle. *ELP* is expressed at high levels during Phase 2A-early Phase 2B only (green box), with low-level expression during pregnancy (Phase 1, light green boxes). *ELP* is undetected in the virgin gland and during involution (white boxes). **B.** Regulatory elements that may control the Phase 2A-early Phase 2B, mammary gland-specific expression of tammar *ELP*. (i) Closed chromatin and hence H3K9me2 and/or H3K27me3 binding to 5'- and/or 3'-DRE(s) may prevent *ELP* expression in all tissues except the Phase 1 and 2A mammary gland; (ii) Open chromatin may permit low-level *ELP* expression during Phase 2A, with NF-κB and an ETS-like element suppressing high-level gene expression; (iii) Open chromatin promoted by the binding of H3K4me3 and/or H3K4me2 to DRE(s) allows C/EBPβ, AP2γ, MAF/ETS and the basal transcriptional machinery to bind to and activate high-level *ELP* expression. Putative binding sites and genomic DNA are not drawn to scale.

mature milk production periods would explain the temporal expression of *CTI*.

The hormone-independent activation of tammar *ELP*

Though the precise identity of MAF is unknown (Mink *et al.*, 1992), a number of ETS family candidates exist, at least 24 of which are expressed in the normal mouse mammary gland (Coletta *et al.*, 2004; Galang *et al.*, 2004). Both ELF1 (E47-like factor 1) and the ETS1-like factor bind to the MAF consensus site (Welte *et al.*, 1994b), while ELF5 is a key regulator of alveolar cell fate, lactation phenotype and milk protein gene expression (Oakes *et al.*, 2008). Either an ETS1- or ELF1-like factor activates the mammary gland-specific, hormone-independent transcription of rodent WAP during lactation (Welte *et al.*, 1994b), while tammar WAP is transcribed during mid-lactation only (Simpson, 1998; Simpson *et al.*, 2000). It is interesting to speculate that the same ETS factor activates both tammar *ELP* and *WAP* with the Phase-specific expression of these genes regulated by different chromatin states. Furthermore, the identification of putative ETS TFBS sites in the promoters of the major milk protein genes (Malewski *et al.*, 2002) suggests that this family plays a vital role in the regulation of lactation. ETS factors can also act synergistically with other transcription factors, e.g. AP1, NF-κB, STAT5A, SP1, CBP/p300 and many others (Li *et al.*, 2000; Sharrocks, 2001). The interaction between ETS and AP2γ has not been documented in the mammary gland, but the close proximity of these sites in both the ELP and CTI proximal promoters and the boost in LGB synthesis (Kuss *et al.*, 2003) suggests that these factors cooperatively activate tammar ELP. Interestingly, EVI1 may also affect ELP transcription as it can bind to an ETS-like AGGAAG(+) motif (Glass *et al.*, 2013), identical to the MAF/ETS CTTCCT(-) site.

The retention of a highly-conserved canonical TATA box (TATAWAWR) (Juven-Gershon *et al.*, 2008) in all marsupial ELP promoters (TATAAAAC) suggests that this is the original TATA box, with a second upstream TATA box (TATAATTT) acquired by tammar ELP after the divergence of macropodids from other marsupials. Although tammar ELP has two TSS, only one mature protein is secreted into milk (Simpson *et al.*, 1998). The lack of a conserved CTI TATA motif was surprising but only ~10% of vertebrate promoters have a canonical TATA box, with at least 70% of promoters initiated by the recruitment of TATA binding protein (TBP) to the promoter by SP1 and/or SP3 (Sandelin *et al.*, 2007; Wobus *et al.*, 2008; Yang *et al.*, 2007). Though both ELP and CTI have low-affinity SP1/SP3 TFBS, the conservation of a marsupial ELP-specific, SP1/SP3-like, CAMP1a element (RGAGGTGGGGAACAT) suggests that this element may an important regulator of Phase 2A-specific gene expression. Additionally,

non-coding RNAs, e.g. miRNAs are differentially expressed during tammar lactation (Modepalli *et al.*, 2014) and may regulate the post-transcriptional production of tammar ELP.

Indirect activation of tammar *ELP*

The absence of a conserved STAT5A site (indicative of prolactin-responsiveness) in the proximal promoters of both *ELP* (and *CTI*) was surprising given the *in vitro* response of *ELP* to lactogenic hormones in mammary explants (Pharo, 2014). This apparent anomaly may be due to the indirect effects of lactogenic hormones on *ELP* transcription. For example, down-stream effectors of the prolactin and insulin-signalling pathways may activate *ELP* transcription (Pharo, 2014). PRL-binding to the short form of the PRLR activates both the mitogen-activated protein kinase (MAPK) and PI3K/AKT pathways (Binart *et al.*, 2010) while insulin also acts via the PI3K/AKT and insulin receptor signalling pathways (Lemay *et al.*, 2007; Oliver and Watson, 2013). Furthermore, the short-lived pulse of prolactin at parturition (Hinds, 1988) and the pulse of 17β-oestradiol ~30 hrs postpartum (Shaw and Renfree, 2006) may trigger the activation of estrogen-responsive genes, e.g. *prolactin*, *PRLR*, *ELF1*, *lactoferrin*, *ERRα* itself (Tang *et al.*, 2004), and thus tammar *ELP*.

The importance of extracellular matrix and a 3D tissue structure

The interaction between exogenous ECM and primary MECs is crucial for the formation of mammospheres and the expression of eutherian milk protein genes *in vitro* (Aggeler *et al.*, 1988; Li *et al.*, 1987; Streuli *et al.*, 1995). Unlike eutherian MECs, tammar PMGCs secrete their own matrix which has a Phase-specific composition and phenotype (Wanyonyi *et al.*, 2013a; Wanyonyi *et al.*, 2013b) and form mammospheres *in vitro* after 24-48 h in culture (personal communication, Sonia Mailer). Therefore, Phase 2A-specific ECM may be essential for *ELP* transcription, as demonstrated for the early lactation-specific endogenous *cathelicin 1a* (*MaeuCath1a, CAMP1a*) gene in tammar PMGCs (Wanyonyi *et al.*, 2013b). The precise mechanism(s) of milk protein gene activation and repression by ECM is unknown, but may involve the binding of ECM-response element(s) to the promoter (Qian and Zhao, 2014; Rijnkels *et al.*, 2013) and/or paracrine signalling via cell-cell contact.

Conclusions

This study has provided new insights into the regulation of the tammar *ELP* gene. *ELP* transcription is rigorously controlled and involves both hormone-dependent and hormone-independent factors. During pregnancy, NF-κB most-likely suppresses *ELP*

transcription. At parturition, the surge of lactogenic hormones, changes in steroids, open chromatin conformation and the binding of MAF/ETS, AP2ɣ and C/EBP to the proximal promoter provide a plausible model for tammar *ELP* gene induction. The correlation of ELP secretion with the absence of an acquired immune system in the PY, plus the sophisticated regulation of the gene suggests that ELP may have a crucial protective role as protease inhibitor during the early lactation period.

References

Aggeler, J., Park, C.S., and Bissell, M.J. (1988). Regulation of milk protein and basement membrane gene expression: the influence of the extracellular matrix. *J Dairy Sci* **71**: 2830-2842.

Ali, S., Edery, M., Pellegrini, I., Lesueur, L., Paly, J., Djiane, J., and Kelly, P.A. (1992). The Nb2 form of prolactin receptor is able to activate a milk protein gene promoter. *Mol Endocrinol* **6**: 1242-1248.

Binart, N., Bachelot, A., and Bouilly, J. (2010). Impact of prolactin receptor isoforms on reproduction. *Trends Endocrinol Metab* **21**: 362-368.

Brambell, F.W.R. (1970). The transmission of passive immunity from mother to young, Vol 18, North-Holland Publishing Company, Amsterdam.

Brennan, A.J., Sharp, J.A., Lefevre, C., Topcic, D., Auguste, A., Digby, M., and Nicholas, K.R. (2007). The tammar wallaby and fur seal: models to examine local control of lactation. *J Dairy Sci* **90 Suppl 1**: E66-75.

Brisken, C., and Ataca, D. (2015). Endocrine hormones and local signals during the development of the mouse mammary gland. *Wiley Interdisciplinary Reviews: Developmental Biology* **4**: 181-195.

Cane, K.N. (2005). The physiological and molecular regulation of lactation in fur seals. PhD Thesis, Department of Zoology (Melbourne: University of Melbourne).

Casey, T.M., Crodian, J., Erickson, E., Kuropatwinski, K.K., Gleiberman, A.S., and Antoch, M.P. (2014). Tissue-specific changes in molecular clocks during the transition from pregnancy to lactation in mice. *Biol Reprod* **90**: 127.

Cechova, D., Svestkova, V.V., Keil, B., and Sorm, F. (1969). Similarities in primary structures of cow colostrum trypsin inhibitor and bovine basic pancreatic trypsin inhibitor. *FEBS Lett* **4**: 155-156.

Coletta, R.D., Jedlicka, P., Gutierrez-Hartmann, A., and Ford, H.L. (2004). Transcriptional control of the cell cycle in mammary gland development and tumorigenesis. *J Mammary Gland Biol Neoplasia* **9**: 39-53.

Collet, C., Joseph, R., and Nicholas, K. (1991). A marsupial β-lactoglobulin gene: characterization and prolactin-dependent expression. *J Mol Endocrinol* **6**: 9-16.

Collet, C., Joseph, R., and Nicholas, K. (1992). Molecular characterization and *in-vitro* hormonal requirements for expression of two casein genes from a marsupial. *J Mol Endocrinol* **8**: 13-20.

De Leo, A.A., Lefevre, C., Topcic, D., Pharo, E., Cheng, J.F., Frappell, P., Westerman, M., Graves, J.A.M., and Nicholas, K.R. (2006). Characterization of two whey protein genes in the Australian dasyurid marsupial, the stripe-faced dunnart (*Sminthopsis macroura*). *Cytogenet Genome Res* **115**: 62-69.

Dearth, L.R., Hutt, J., Sattler, A., Gigliotti, A., and DeWille, J. (2001). Expression and function of CCAAT/enhancer binding protein β (C/EBPβ) LAP and LIP isoforms in mouse mammary gland, tumors and cultured mammary epithelial cells. *J Cell Biochem*

82: 357-370.

Demmer, J., Ross, I.K., Ginger, M.R., Piotte, C.K., and Grigor, M.R. (1998). Differential expression of milk protein genes during lactation in the common brushtail possum (*Trichosurus vulpecula*). *Journal of Molecular Endocrinology* **20**: 37-44.

Demmer, J., Stasiuk, S.J., Grigor, M.R., Simpson, K.J., and Nicholas, K.R. (2001). Differential expression of the whey acidic protein gene during lactation in the brushtail possum (*Trichosurus vulpecula*). *Biochim Biophys Acta* **1522**: 187-194.

Doppler, W., Geymayer, S., and Weirich, H.G. (2000). Synergistic and antagonistic interactions of transcription factors in the regulation of milk protein gene expression. Mechanisms of cross-talk between signalling pathways. *Adv Exp Med Biol* **480**: 139-146.

Doppler, W., Hock, W., Hofer, P., Groner, B., and Ball, R.K. (1990). Prolactin and glucocorticoid hormones control transcription of the b-casein gene by kinetically distinct mechanisms. *Mol Endocrinol* **4**: 912-919.

Doppler, W., Welte, T., and Philipp, S. (1995). CCAAT/enhancer-binding protein isoforms β and δ are expressed in mammary epithelial cells and bind to multiple sites in the β-casein gene promoter. *J Biol Chem* **270**: 17962-17969.

Edwards, M.J., Hinds, L.A., Deane, E.M., and Deakin, J.E. (2012). A review of complementary mechanisms which protect the developing marsupial pouch young. *Dev Comp Immunol* **37**: 213-220.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *PNAS* **95**: 14863-14868.

Findlay, L. (1982). The mammary glands of the tammar wallaby (*Macropus eugenii*) during pregnancy and lactation. *J Reprod Fertil* **65**: 59-66.

Frigo, L., and Woolley, P.A. (1997). Growth and Development of Pouch Young of the Stripe-faced Dunnart, *Sminthopsis macroura* (Marsupialia : Dasyuridae), in Captivity. *Australian Journal of Zoology* **45**: 157-170.

Galang, C.K., Muller, W.J., Foos, G., Oshima, R.G., and Hauser, C.A. (2004). Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors. *J Biol Chem* **279**: 11281-11292.

Geymayer, S., and Doppler, W. (2000). Activation of NF-κB p50/p65 is regulated in the developing mammary gland and inhibits STAT5-mediated β-casein gene expression. *Faseb Journal* **14**: 1159-1170.

Gigliotti, A.P., and DeWille, J.W. (1998). Lactation status influences expression of CCAAT/enhancer binding protein isoform mRNA in the mouse mammary gland. *J Cell Physiol* **174**: 232-239.

Glass, C., Wuertzer, C., Cui, X., Bi, Y., Davuluri, R., Xiao, Y.-Y., Wilson, M., Owens, K., Zhang, Y., and Perkins, A. (2013). Global identification of EVI1 target genes in Acute Myeloid Leukemia. *PLoS ONE* **8**: e67134.

Gouilleux, F., Wakao, H., Mundt, M., and Groner, B. (1994). Prolactin induces

phosphorylation of Tyr694 of Stat5 (MGF), a prerequisite for DNA binding and induction of transcription. *EMBO J* **13**: 4361-4369.

Green, B. (1984). Composition of milk and energetics of growth in marsupials. In Symposia of the Zoological Society of London 51 Physiological Strategies in Lactation, M. Peaker, R.G. Vernon, and C.H. Knight, eds. (London: Academic Press), pp. 369–387.

Green, B., Newgrain, K., and Merchant, J. (1980). Changes in milk composition during lactation in the tammar wallaby (*Macropus eugenii*). *Aust J Biol Sci* **33**: 35-42.

Hennighausen, L., and Robinson, G.W. (2005). Information networks in the mammary gland. *Nature Reviews Molecular Cell Biology* **6**: 715-725.

Hinds, L.A. (1988). Hormonal control of lactation. In The Developing Marsupial Models for Biomedical Research, C.H. Tyndale-Biscoe, and P.A. Janssens, eds. (Berlin, Germany: Springer-Verlag), pp. 55-67.

Horseman, N.D., and Gregerson, K.A. (2014). Prolactin actions. *Journal of Molecular Endocrinology* **52**: R95-R106.

Hughes, K., and Watson, C.J. (2012). The spectrum of STAT functions in mammary gland development. *JAK-STAT* **1**: 151-158.

Jenness, R. (1974). The composition of milk. In Lactation: A Comprehensive Treatise, B.L. Larson, and V.R. Smith, eds. (New York: Academic Press, Inc.), pp. 3-107.

Joss, J., Molloy, M., Hinds, L., and Deane, E. (2007). Proteomic analysis of early lactation milk of the tammar wallaby (*Macropus eugenii*). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **2**: 150-164.

Joss, J.L., Molloy, M.P., Hinds, L., and Deane, E. (2009). A longitudinal study of the protein components of marsupial milk from birth to weaning in the tammar wallaby (*Macropus eugenii*). *Dev Comp Immunol* **33**: 152-161.

Juven-Gershon, T., Hsu, J.-Y., Theisen, J.W.M., and Kadonaga, J.T. (2008). The RNA polymerase II core promoter - the gateway to transcription. *Current Opinion in Cell Biology* **20**: 253-259.

Kanai, A., Nonomura, N., Yoshimura, M., and Oka, T. (1993). DNA-binding proteins and their *cis*-acting sites controlling hormonal induction of a mouse β-casein: CAT fusion protein in mammary epithelial cells. *Gene* **126**: 195-201.

Kolb, A.F., Albang, R., Brem, G., Erfle, V., Gunzburg, W.H., and Salmons, B. (1995). Characterization of a protein that binds a negative regulatory element in the mammary-specific whey acidic protein promoter. *Biochem Biophys Res Commun* **217**: 1045-1052.

Kolb, A.F., Gunzburg, W.H., Albang, R., Brem, G., Erfle, V., and Salmons, B. (1994). Negative regulatory elements in the mammary specific whey acidic protein promoter. *Journal of Cellular Biochemistry* **56**: 245-261.

Kruse, P.E. (1983). The importance of colostral immunoglobulins and their absorption from the intestine of the newborn animals. *Ann Rech Vet* **14**: 349-353.

Kuss, A.W., Gogol, J., and Geidermann, H. (2003). Associations of a polymorphic AP-2 binding site in the 5'-flanking region of the bovine beta-lactoglobulin gene with milk proteins. *J Dairy Sci* **86**: 2213-2218.

Kwek, J.H., Iongh, R.D., Digby, M.R., Renfree, M.B., Nicholas, K.R., and Familari, M. (2009). Cross-fostering of the tammar wallaby (*Macropus eugenii*) pouch young accelerates fore-stomach maturation. *Mech Develop* **126**: 449-463.

Laskowski, M., Jr., and Laskowski, M. (1951). Crystalline trypsin inhibitor from colostrum. *J Biol Chem* **190**: 563-573.

Laskowski, M., and Qasim, M.A. (2000). What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim Biophys Acta* **1477**: 324-337.

Lee, C.S., and Oka, T. (1992a). A pregnancy-specific mammary nuclear factor involved in the repression of the mouse β-casein gene transcription by progesterone. *J Biol Chem* **267**: 5797-5801.

Lee, C.S., and Oka, T. (1992b). Progesterone regulation of a pregnancy-specific transcription repressor to β-casein gene promoter in mouse mammary gland. *Endocrinology* **131**: 2257-2262.

Lefèvre, C.M., Digby, M.R., Whitley, J.C., Strahm, Y., and Nicholas, K.R. (2007). Lactation transcriptomics in the Australian marsupial, *Macropus eugenii*: transcript sequencing and quantification. *BMC Genomics* **8**: 417.

Lemay, D.G., Neville, M.C., Rudolph, M.C., Pollard, K.S., and German, J.B. (2007). Gene regulatory networks in lactation: identification of global principles using bioinformatics. *BMC Syst Biol* **1**: 56.

Lemay, D.G., Pollard, K.S., Martin, W.F., Freeman Zadrowski, C., Hernandez, J., Korf, I., German, J.B., and Rijnkels, M. (2013). From genes to milk: genomic organization and epigenetic regulation of the mammary transcriptome. *PLoS One* **8**: e75030.

Li, M.L., Aggeler, J., Farson, D.A., Hatier, C., Hassell, J., and Bissell, M.J. (1987). Influence of a reconstituted basement membrane and its components on casein gene expression and secretion in mouse mammary epithelial cells. *Proc Natl Acad Sci U S A* **84**: 136-140.

Li, R., Pei, H., and Watson, D.K. (2000). Regulation of Ets function by protein - protein interactions. *Oncogene* **19**: 6514-6523.

Lubon, H., and Hennighausen, L. (1988). Conserved region of the rat α-lactalbumin promoter is a target site for protein binding *in vitro*. *Biochem J* **256**: 391-396.

Malewski, T. (1998). Computer analysis of distribution of putative *cis*- and *trans*-regulatory elements in milk protein gene promoters. *Biosystems* **45**: 29-44.

Malewski, T., Gajewska, M., Zebrowska, T., and Zwierzchowski, L. (2002). Differential induction of transcription factors and expression of milk protein genes by prolactin and growth hormone in the mammary gland of rabbits. *Growth Horm IGF Res* **12**: 41-53.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H., *et al.* (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* **42**: D142-147.

Menzies, K.K., Lefevre, C., Macmillan, K.L., and Nicholas, K.R. (2009). Insulin regulates milk protein synthesis at multiple levels in the bovine mammary gland. *Funct Integr Genomics* **9**: 197-217.

Mink, S., Hartig, E., Jennewein, P., Doppler, W., and Cato, A.C. (1992). A mammary cell-specific enhancer in mouse mammary tumor virus DNA is composed of multiple regulatory elements including binding sites for CTF/NFI and a novel transcription factor, mammary cell-activating factor. *Mol Cell Biol* **12**: 4906-4918.

Modepalli, V., Kumar, A., Hinds, L.A., Sharp, J.A., Nicholas, K.R., and Lefevre, C. (2014). Differential temporal expression of milk miRNA during the lactation cycle of the marsupial tammar wallaby (*Macropus eugenii*). *BMC Genomics* **15**: 1012.

NH&MRC (National Health & Medical Research Council, Australia) (2013) Australian code for the care and use of animals for scientific purposes, 8th edition. Canberra: National Health and Medical Research Council.

Nicholas, K.R., Simpson, K.J., Wilson, M., Trott, J., and Shaw, D. (1997). The tammar wallaby: a model to study putative autocrine-induced changes in milk composition. *J Mammary Gland Biol Neoplasia* **2**: 299-310.

Nicholas, K.R., and Tyndale-Biscoe, C.H. (1985). Prolactin-dependent accumulation of α-lactalbumin in mammary gland explants from the pregnant tammar wallaby (*Macropus eugenii*). *J Endocrinol* **106**: 337-342.

Oakes, S.R., Naylor, M.J., Asselin-Labat, M.L., Blazek, K.D., Gardiner-Garden, M., Hilton, H.N., Kazlauskas, M., Pritchard, M.A., Chodosh, L.A., Pfeffer, P.L., *et al.* (2008). The Ets transcription factor Elf5 specifies mammary alveolar cell fate. *Genes Dev* **22**: 581-586.

Oliver, C.H., and Watson, C.J. (2013). Making milk: A new link between STAT5 and Akt1. *JAKSTAT* **2**: e23228.

Osada, S., Yamamoto, H., Nishihara, T., and Imagawa, M. (1996). DNA binding specificity of the CCAAT/enhancer-binding protein transcription factor family. *J Biol Chem* **271**: 3891-3896.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J.-C., Ferrier, P., and Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO Journal* **30**: 4198-4210.

Pharo, E.A. (2014). Expression of the mammary gland-specific tammar wallaby *Early Lactation Protein* gene is maintained *in vitro* in the absence of prolactin. *Mol Cell Endocrinol* **382**: 871-880.

Pharo, E.A., Cane, K.N., McCoey, J., Buckle, A.M., Oosthuizen, W.H., Guinet, C., and Arnould, J.P.Y. (2016) A colostrum trypsin inhibitor gene expressed in the cape fur seal mammary gland during lactation. *Gene*. **578**: 7-16.

Pharo, E.A., De Leo, A.A., Renfree, M.B., Thomson, P.C., Lefevre, C.M., and Nicholas, K.R. (2012). The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene. *BMC Evol Biol* **12**: 8.

Piotte, C.P., and Grigor, M.R. (1996). A novel marsupial protein expressed by the mammary gland only during the early lactation and related to the Kunitz proteinase inhibitors. *Archives of Biochemistry and Biophysics* **330**: 59-64.

Qian, X., and Zhao, F.Q. (2014). Current major advances in the regulation of milk protein gene expression. *Crit Rev Eukaryot Gene Expr* **24**: 357-378.

Raught, B., Khursheed, B., Kazansky, A., and Rosen, J. (1994). YY1 represses β-casein gene expression by preventing the formation of a lactation-associated complex. *Mol Cell Biol* **14**: 1752-1763.

Raught, B., Liao, W.S., and Rosen, J.M. (1995). Developmentally and hormonally regulated CCAAT/enhancer-binding protein isoforms influence β-casein gene expression. *Mol Endocrinol* **9**: 1223-1232.

Rawlings, N.D., Waller, M., Barrett, A.J., and Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **42**: D503-509.

Rijnkels, M., Freeman-Zadrowski, C., Hernandez, J., Potluri, V., Wang, L., Li, W., and Lemay, D.G. (2013). Epigenetic modifications unlock the milk protein gene loci during mouse mammary gland development and differentiation. *PLoS One* **8**: e53270.

Rijnkels, M., Kabotyanski, E., Montazer-Torbati, M.B., Beauvais, C.H., Vassetzky, Y., Rosen, J.M., and Devinoy, E. (2010). The epigenetic landscape of mammary gland development and functional differentiation. *J Mammary Gland Biol Neoplasia* **15**: 85-100.

Riley, L.G., Gardiner-Garden, M., Thomson, P.C., Wynn, P.C., Williamson, P., Raadsma, H.W., and Sheehy, P.A. (2010). The influence of extracellular matrix and prolactin on global gene expression profiles of primary bovine mammary epithelial cells *in vitro*. *Anim Genet* **41**: 55-63.

Robinson, G.W., Johnson, P.F., Hennighausen, L., and Sterneck, E. (1998). The C/EBPβ transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland. *Genes Dev* **12**: 1907-1916.

Rosen, J.M., Wyszomierski, S.L., and Hadsell, D. (1999). Regulation of milk protein gene expression. *Annual Review of Nutrition* **19**: 407-436.

Rosen, J.M., Zahnow, C., Kazansky, A., and Raught, B. (1998). Composite response elements mediate hormonal and developmental regulation of milk protein gene expression. *Biochem Soc Symp* **63**: 101-113.

Roy, L., Cates, S., Schifferli, K., Pichet, J., Ciccarone, V., Bennett, S., and Hawley-Newson, P. (1999). High transfection efficiency of cloned cell lines. In Focus, pp. 62-62.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual, Third edn (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* **8**: 424-436.

Scott, C.J., and Taggart, C.C. (2010). Biologic protease inhibitors as novel therapeutic agents. *Biochimie*.

Sementchenko, V.I., and Watson, D.K. (2000). Ets target genes: past, present and future. *Oncogene* **19**: 6533-6548.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annual Review of Biochemistry* **76**: 75-100.

Sharp, J.A., Digby, M., Lefevre, C., Mailer, S., Khalil, E., Topcic, D., Auguste, A., Kwek, J.H., Brennan, A.J., Familari, M.*, et al.* (2009). The comparative genomics of tammar wallaby and Cape fur seal lacation models to examine functions of milk proteins. In Milk Proteins: from Expression to Food, A. Thompson, M. Boland, and H. Singh, eds. (San Diego, CA: Academic Press/Elsevier), pp. 55-79.

Sharp, J.A., Watt, A., Bisana, S., Modepalli, V., Wanyonyi, S., Kumar, A., Kwek, J., Collins, R., Lefèvre, C., and Nicholas, K.R. (2014). The comparative genomics of monotremes, marsupials, and pinnipeds: models to examine functions of milk proteins. In Milk Proteins : From Expression to Food (Second Edition), H. Singh, M. Boland, and A. Thompson, eds. (San Diego: Academic Press), pp. 75-112.

Sharrocks, A.D. (2001). The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol* **2**: 827-837.

Shaw, G., and Renfree, M.B. (2006). Parturition and perfect prematurity: birth in marsupials. *Australian Journal of Zoology* **54**: 139-149.

Simpson, K. (1998). The tammar wallaby, *Macropus eugenii*, a model to study autocrine and endocrine regulation of lactation. In Department of Agriculture (Melbourne: Latrobe University).

Simpson, K., Shaw, D., and Nicholas, K. (1998). Developmentally-regulated expression of a putative protease inhibitor gene in the lactating mammary gland of the tammar wallaby, *Macropus eugenii*. *Comp Biochem Phys B* **120**: 535-541.

Simpson, K.J., Ranganathan, S., Fisher, J.A., Janssens, P.A., Shaw, D.C., and Nicholas, K.R. (2000). The gene for a novel member of the whey acidic protein family encodes three four-disulfide core domains and is asynchronously expressed during lactation. *J Biol Chem* **275**: 23074-23081.

Streuli, C.H., Schmidhauser, C., Bailey, N., Yurchenco, P., Skubitz, A.P., Roskelley, C., and Bissell, M.J. (1995). Laminin mediates tissue-specific gene expression in mammary epithelia. *J Cell Biol* **129**: 591-603.

Tang, S., Han, H., and Bajic, V.B. (2004). ERGDB: Estrogen Responsive Genes Database. *Nucleic Acids Research* **32**: D533-D536.

Topcic, D., Auguste, A., De Leo, A.A., Lefevre, C., Digby, M.R., and Nicholas, K.R. (2009). Characterization of the tammar wallaby (*Macropus eugenii)* whey acidic protein gene: new insights into the function of the protein. *Evol Dev* **11**: 363-375.

Topper, Y.J., and Freeman, C.S. (1980). Multiple hormone interactions in the developmental biology of the mammary gland. *Physiol Rev* **60**: 1049-1106.

Trott, J.F. (1999). Endocrine and local control of milk protein gene expression in the tammar wallaby, *Macropus eugenii*. In Department of Molecular Biology, Genetics & Zoology (Melbourne: University of Melbourne).

Trott, J.F., Adams, T.E., Wilson, M., and Nicholas, K.R. (2005). Positive and negative regulatory elements in the late lactation protein-A gene promoter from the tammar wallaby (*Macropus eugenii*). *Bba-Gene Struct Expr* **1728**: 65-76.

Trott, J.F., Wilson, M.J., Hovey, R.C., Shaw, D.C., and Nicholas, K.R. (2002). Expression of novel lipocalin-like milk protein gene is developmentally-regulated during lactation in the tammar wallaby, *Macropus eugenii*. *Gene* **283**: 287-297.

Tyndale-Biscoe, H., and Renfree, M. (1987). Reproductive Physiology of Marsupials (Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo: Cambridge University Press).

Veselsky, L., Cechova, D., and Jonakova, V. (1978). Secretion and immunochemical properties of the trypsin inhibitor from bovine colostrum. *Hoppe Seylers Z Physiol Chem* **359**: 873-878.

Wakao, H., Gouilleux, F., and Groner, B. (1994). Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response. *EMBO J* **13**: 2182-2191.

Wang, Y., Li, X., and Hu, H. (2014). H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics* **103**: 222-228.

Wanyonyi, S.S., Lefevre, C., Sharp, J.A., and Nicholas, K.R. (2013a). The extracellular matrix locally regulates asynchronous concurrent lactation in tammar wallaby (Macropus eugenii). *Matrix Biology* **32**: 342-351.

Wanyonyi, S.S., Lefevre, C., Sharp, J.A., and Nicholas, K.R. (2013b). The extracellular matrix regulates *MaeuCath1a* gene expression. *Dev Comp Immunol* **40**: 289-299.

Watson, C., and Burdon, T. (1996). Prolactin signal transduction mechanisms in the mammary gland: the role of the Jak/Stat pathway. *Rev Reprod* **1**: 1-5.

Wei, G.H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R.*, et al.* (2010). Genome-wide analysis of ETS-family DNA-binding *in vitro* and *in vivo*. *EMBO J* **29**: 2147-2160.

Welte, T., Garimorth, K., Philipp, S., and Doppler, W. (1994a). Prolactin-dependent activation of a tyrosine phosphorylated DNA binding factor in mouse mammary epithelial cells. *Mol Endocrinol* **8**: 1091-1102.

Welte, T., Garimorth, K., Philipp, S., Jennewein, P., Huck, C., Cato, A.C., and Doppler, W. (1994b). Involvement of Ets-related proteins in hormone-independent mammary cell-specific gene expression. *Eur J Biochem* **223**: 997-1006.

Winklehner-Jennewein, P., Geymayer, S., Lechner, J., Welte, T., Hansson, L., Geley, S., and Doppler, W. (1998). A distal enhancer region in the human β-casein gene

mediates the response to prolactin and glucocorticoid hormones. *Gene* **217**: 127-139.

Wobus, M., Wandel, E., Prohaska, S., Findei, S., Tschöp, K., and Aust, G. (2008). Transcriptional regulation of the human CD97 promoter by Sp1/Sp3 in smooth muscle cells. *Gene* **413**: 67-75.

Yamashita, H., Nevalainen, M.T., Xu, J., LeBaron, M.J., Wagner, K.U., Erwin, R.A., Harmon, J.M., Hennighausen, L., Kirken, R.A., and Rui, H. (2001). Role of serine phosphorylation of Stat5a in prolactin-stimulated β-casein gene expression. *Molecular and Cellular Endocrinology* **183**: 151-163.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389**: 52-65.

Zahnow, C.A., Cardiff, R.D., Laucirica, R., Medina, D., and Rosen, J.M. (2001). A role for CCAAT/enhancer binding protein β-liver-enriched inhibitory protein in mammary epithelial cell proliferation. *Cancer Res* **61**: 261-269.

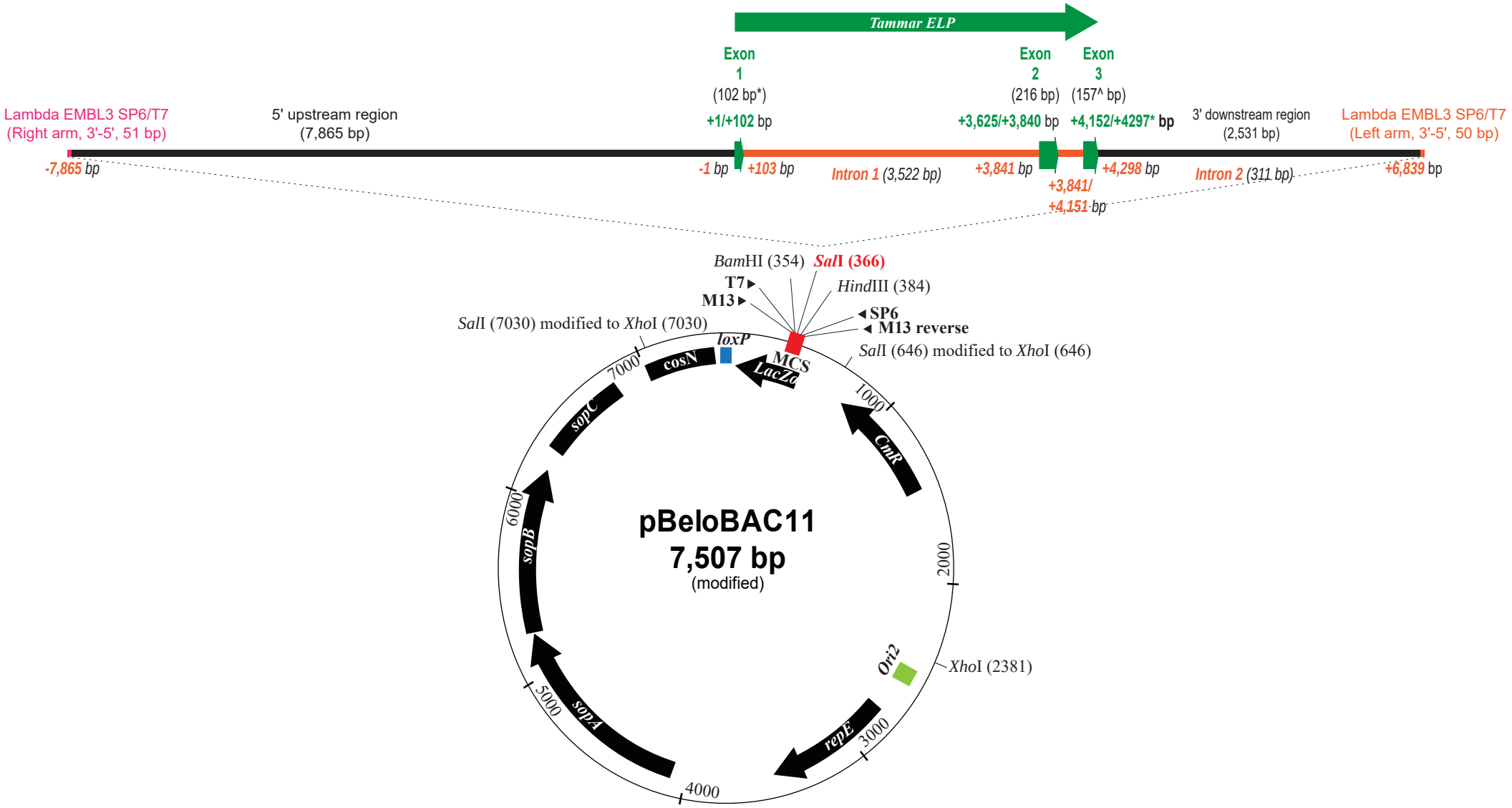
Zhang, H., Lee, L., and Ip, M.M. (2004). The liver-enriched inhibitory protein isoform of CCAAT/enhancer-binding protein β, but not nuclear factor-κB, mediates the transcriptional inhibition of beta-casein by tumor necrosis factor-alpha. *Endocrinology* **145**: 2833-2844.

Zhou, J., and Cidlowski, J.A. (2005). The human glucocorticoid receptor: one gene, multiple proteins and diverse responses. *Steroids* **70**: 407-417.

Acknowledgements

This research was funded by the Cooperative Research Centre for Innovative Dairy Products (CRC-IDP) and the Department of Zoology, The University of Melbourne, Australia. EAP was the recipient of an Australian Postgraduate Award (APA) and KNC, a Dairy Australia Postgraduate Student Award and both EAP and KNC received a scholarship top-up from the CRC-IDP. We thank Prof. Kevin Nicholas for his helpful advice, Dr. Hong-Jian Zhu for the use of his GloMax-96 Microplate Luminometer, Josie Iaria for her assistance with dual luciferase assays and Sonia Mailer for assistance with the preparation of tammar PMGCs.

Appendix A Supplementary material



Supplementary file 2. Map of pBeloBACII-14.7kbtELP

A tammar genomic library (liver) in the *E. coli* phage vector lambda EMBL3 T7/SP6 was screened with tammar *ELP* cDNA and a positive clone isolated (Simpson, 1998). The clone was *Sal*I digested and the ~14.7 kb genomic DNA fragment (-7,865/+6839, +1 denotes the TSS of the short transcript) cloned into the *Sal*I (366) site of a modified pBeloBACII vector in which the *Sal*I (646 and 7030) had been modified to *Xho*I. For numbering simplicity, *the TSS of the short transcript is used as the start of exon 1 and, ^up to and including the polyadenylation signal is used as the end of exon 3.

Simpson, K. (1998). The tammar wallaby, *Macropus eugenii*, a model to study autocrine and endocrine regulation of lactation. PhD Thesis, Department of Agriculture (Melbourne: Latrobe University).

Supplementary file 2. Cell Culture Methods: HC11, KIM-2 and MDA-MB-231 immortalised cell lines and Tammar mammary gland primary cells

Cell Culture

1. HC11 Cell Culture

HC11 cells (Ball et al., 1988) were cultured in maintenance media consisting of Roswell Park Memorial Institute (RPMI) 1640 supplemented with 10% heat inactivated FBS, 2 mM L-glutamine, 5 μg/mL bovine insulin (I-5500, Sigma), 10 ng/mL human EGF (354001, BD Biosciences) and Penicillin-Streptomycin (100 IU/mL, 100 μg/mL). Cells were incubated at 37°C in 5% CO_2 as described (Hynes et al., 1990) and the culture media changed every two days. HC11s are a derivative of murine COMMA-1D MECs, which were originally isolated from the mammary gland of BALB/c mice during mid-pregnancy (Danielson et al., 1984).

1.1 Stable co-transfection of the 14.7 kb tammar ELP *gene and neomycin-resistance gene into HC11 cells*

At least fifteen stable lines were created by co-transfection of the linearised 14.7 kb tammar *ELP* gene (-7,865/+6839) and a plasmid expressing the neomycin-resistance gene (pSV2-Neo, (Southern and Berg, 1982)) into HC11 cells. The complete 14.7 kb tammar *ELP* gene was released from the pBeloBACII-14.7kbtELP plasmid by *Sal*I digestion, gel extracted and purified. Negative control lines were also created: the *neomycin-resistance* gene alone, the 14.7 kb tammar *ELP* gene alone and cells alone (no DNA), with the latter two controls used to establish a cell death curve in response to antibiotic selection. HC11 cells were seeded in 6-well plates at 4 x 10^5 cells per well in HC11 proliferation (maintenance) media without antibiotics. The following day, cells were transfected with 4 μg of DNA and 8 μL Lipofectamine 2000 (LF 2000, Invitrogen) in a total volume of 500 μL in OptiMEM I Reduced Serum Medium (Invitrogen), as per the manufacturer's instructions. The 14.7 kb tammar *ELP* linearised genomic DNA and the *neomycin-resistance* gene were co-transfected into HC11 cells at two different ratios: 14:1 and 29: 1 (14.7 kb tELP: pSV2-Neo). All cells were trypsinised the following day at a 1 in 10 dilution and grown in maintenance media. Forty-eight hours post-transfection, selection for neomycin-resistant cells was commenced with the addition of 400

μg/mL Geneticin Selective Antibiotic. Cells were trypsinised and split to establish clonal lines. Culture media was changed every second day and cell lines expanded over an eight-ten week period.

1.2 Screening of the HC11 stable transfectants for the 14.7 kb tammar ELP *gene*

Stable lines were analysed for the incorporation of the 14.7 kb tammar *ELP* gene using a PCR-based screening approach. Genomic DNA was isolated from stable lines [putative positive lines (+14.7kb tELP + pSV2-Neo) and negative lines (+pSV2-Neo)] with TriPure Isolation Reagent (Roche) according to the manufacturer's instructions. PCR-based screening with *Taq* DNA Polymerase in Storage Buffer B (M1661, Promega Corporation) was performed using three sets of primers (Supplementary file 5); two sets of tammar *ELP* gene-specific primers and one set of endogenous mouse *CSN2* promoter primers (-259/+7, numbering relative to the transcription start site, Table 1), the latter as a positive control for genomic DNA integrity. The tammar *ELP* primers amplified two regions of the gene, the promoter (-542/+20) and the region spanning exons 2 and 3 (+3625/+4176), so as to distinguish between genomic DNA (552 bp product) and mRNA/cDNA contamination (241 bp product). Tammar *ELP* numbering is given relative to the short transcription start site.

1.3 Hormonal induction of milk protein gene expression in HC11 cell lines with the stably-transfected 14.7 kb tammar ELP *gene*

The lactogenic hormone response of six positive HC11 lines containing the stably-integrated 14.7 kb tammar *ELP* gene and three negative lines containing the neomycin-resistance gene alone was tested using the HC11 differentiation protocol (Supplementary file 8A) kindly provided by Michael Kazlauskas (Garvan Institute, Sydney, Australia). Briefly, cells were plated in 2 mL proliferation media supplemented with bovine insulin (5 μg/mL), human EGF (10 ng/mL) and FBS (10%) at 1.0 x 10^5 cells per well in a 6-well plate and 48 h later, the media changed. The following day, the confluent cells were washed with 1X PBS and fresh proliferation media minus EGF added. From days 4 to 7, inclusive, media was changed daily and milk protein gene expression induced with differentiation media (proliferation media minus EGF but supplemented with 1 μM dexamethasone (D, 08893, Sigma) and 5 μg/mL ovine prolactin: IDPRL (insulin, cortisol and prolactin) treatment.

As a negative control, cells were cultured in proliferation media minus EGF: I (insulin) treatment. On day 8, cells were harvested. Total RNA was extracted with TriPure Isolation Reagent and expression of tammar *ELP* and endogenous mouse *CSN2* assayed by Northern analysis, as described (Pharo et al., 2012). Gene expression levels were quantified by PhosphorImager analysis.

1.4 Hormonal induction of HC11 cells transiently transfected with tammar ELP promoter constructs

The lactogenic hormone-responsiveness of HC11 cells transiently transfected with tammar *ELP*, or mouse *CSN2* promoter constructs and the pSV-β-galactosidase (pSV-β-gal) plasmid as an internal transfection control, was based upon the HC11 differentiation protocol for transiently transfected cells (Supplementary file 9A) from Michael Kazlauskas, Garvan Institute. Briefly, cells were seeded in 2 mL proliferation media at 4.0 x 10^5 cells/well in 35 mm wells and transfected the following day with 4 μg DNA (3.2 μg promoter construct and 0.8 μg pSV-β-gal) and LF 2000 (8 μL). Transfection complexes were removed 4-6 h post-transfection and replaced with either maintenance media (I), or differentiation media (IDPRL). Transgene expression was assayed 48 h post-transfection, either by fluorescence microscopy for d2EGFP, or cell lysates collected for luciferase and β-galactosidase assays.

2. KIM-2 Cell Culture

KIM-2 conditional immortalised cells isolated from the mouse mammary (Gordon et al., 2000) were cultured in 50% Dulbecco's Modified Eagle Medium (DMEM) and 50% Ham's F-12 Nutrient Mixture, supplemented with 10% heat-inactivated FBS, Penicillin-Streptomycin (100 U/mL/100 μg/mL), bovine insulin (5 μg/mL), human EGF (10 ng/mL) and linoleic acid (5 μg/mL, L8384, Sigma) at 37°C in 5% CO_2. Every 1 in 5 passages, cells were cultured on collagen-coated flasks, as described (Gordon et al., 2000). Differentiation and transient transfection of KIM-2 cells were carried out as per Supplementary files 8A and 9A respectively.

3. MDA-MB-231 Cell Culture

MDA-MB-231 cells originally isolated from a patient with a breast adrenocarcioma (Cailleau et al.,

1974) were cultured in DMEM supplemented with 10% heat-inactivated FBS, 4 mM L-glutamine, Penicillin-Streptomycin (100 U/mL/100 µg/mL), at 37°C in 5% CO_2. Transient transfections of MDA-MB-231 cells were carried out by seeding culture dishes (96-well) 24 h prior to transfection at 1.8 x 10^4 cells per well in media (100 µL) without antibiotics. Cells were transfected with 50 ng plasmid DNA (*ELP* promoter construct) and ViaFect (0.2 µL, Promega) in Opti-MEM I (Invitrogen). The pRL-CMV *Renilla* luciferase plasmid (0.5 ng) was co-transfected as an internal transfection control and transgene expression assayed 24 h post-transfection using dual luciferase assays.

4. Culture of tammar wallaby primary mammary gland cells (PMGCs)

4.1 Preparation of PMGCs from tammar mammary glands

Mammary glands from mid-pregnant (d16P), or late-pregnant tammars (d24P) were extracted, weighed, cut into pieces and finely chopped with scissors. Tissue was placed in Digestion media [1X Hanks' Balanced Salt Solution (HBSS), Fungizone (5 µg/mL), Penicillin-Streptomycin (100 U/mL/100 µg/mL), Bovine Serum Albumin (0.35%), 5 µg/mL Collagenase, Type 2 (Worthington Biochemical Corporation, UK), glucose (0.5 mM)], 1 g tissue/10 mL media in a sterile flask and incubated at 37°C in a shaking incubator (200 rpm) for 2 h (Sharp et al., 2008). The mixture was pipetted at least once during the incubation to break up tissue clumps. The digested tissue was filtered under vacuum through 150 µm mesh using a Nalgene filter unit, collected and centrifuged at 500 *g* for 5 min at RT (Thermo IEC Centra-CL2, Thermo Electron Corporation). The supernatant was removed and the pellet of cells re-suspended and washed three times with Wash media [1X HBSS, DNase I (20 µg/mL), Trypsin Inhibitor from soyabean (1 mg/mL, Sigma)]. Briefly, the cell pellet was re-suspended with Wash media (2 mL) and filtered under vacuum through 50 µm mesh. The filtrate was collected and centrifuged at 500 *g* for 5 min, the supernatant removed and the pellet re-suspended in 2 mL was media. This step was repeated a further two times, but on the last wash, the pellet was re-suspended with 40 mL Wash media. Cells were pelleted, re-suspended in PMGC media (see below) and counted (Sambrook and Russell, 2001). The cell suspension was centrifuged again at 500 *g* for 5 min, the supernatant removed and equal volumes of FBS and 20% FBS/80% DMSO (Sigma)

added. Cells were frozen at ~2 x 10^7 cells/mL at -80°C in a Cryo 1°C Freezing Container (5100-0001, Nalgene) and then stored long-term in liquid N_2.

4.2 Hormonal induction and transient transfection of tammar PMGCs

Tammar PMGCs consisted of a mixed population of cells, including mammary epithelial cells, fibroblasts and myoepithelial cells (Personal communication, Sonia Mailer). PMGCs were maintained in media optimized by Sonia Mailer, a member of the research group; who also performed much of the original characterisation of the PMGCs. PMGC media consisted of 50% DMEM and 50% F-12 Nutrient Mixture supplemented with 10% heat-inactivated FBS, Penicillin-Streptomycin (100 U/mL/100 µg/mL), 2 mM L-glutamine, porcine insulin (1 µg/mL), cortisol (1 µg/mL) and high quality, tissue-culture grade EGF (10 ng/mL, E-4127, Sigma). Differentiation of PMGCs with lactogenic hormones and transient transfections were carried out as for HC11 and KIM-2 cells (Supplementary files 8A and 9A) respectively.

References

Ball, R.K., Friis, R.R., Schoenenberger, C.A., Doppler, W., and Groner, B. (1988). Prolactin regulation of □-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *EMBO Journal.* **7**: 2089-2095.

Cailleau, R., Young, R., Olivé, M., and Reeves, W.J. (1974). Breast Tumor Cell Lines From Pleural Effusions. *Journal of the National Cancer Institute* **53**: 661-674.

Danielson, K.G., Oborn, C.J., Durban, E.M., Butel, J.S., and Medina, D. (1984). Epithelial mouse mammary cell line exhibiting normal morphogenesis *in vivo* and functional differentiation *in vitro*. *Proceedings of the National Academy of Sciences of the U S A* **81**: 3756-3760.

Gordon, K.E., Binas, B., Chapman, R.S., Kurian, K.M., Clarkson, R.W., Clark, A.J., Lane, E.B., and Watson, C.J. (2000). A novel cell culture model for studying differentiation and apoptosis in the mouse mammary gland. *Breast Cancer Research* **2**: 222-235.

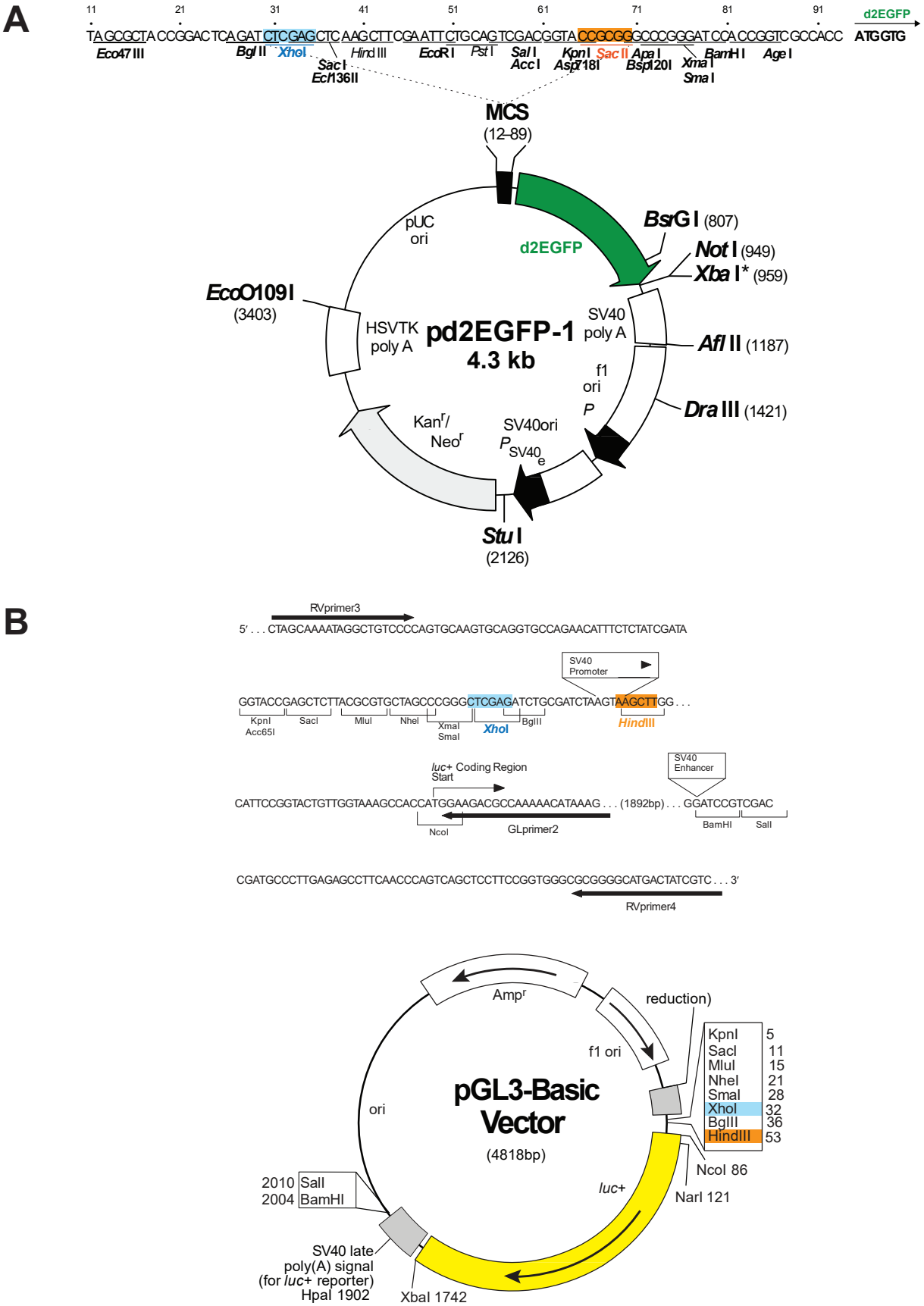
Hynes, N.E., Taverna, D., Harwerth, I.M., Ciardiello, F., Salomon, D.S., Yamamoto, T., and Groner, B. (1990). Epidermal growth factor receptor, but not c-erbB-2, activation prevents lactogenic hormone induction of the b-casein gene in mouse mammary epithelial cells. *Molecular and Cellular Biology* **10**: 4027-4034.

Pharo, E.A., De Leo, A.A., Renfree, M.B., Thomson, P.C., Lefevre, C.M., and Nicholas, K.R. (2012). The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene. *BMC Evolutionary Biology* **12**: 8.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual, Third edn (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Sharp, J.A., Mailer, S.L., Thomson, P.C., Lefevre, C., and Nicholas, K.R. (2008). Identification and transcript analysis of a novel wallaby (*Macropus eugenii*) basal-like breast cancer cell line. *Molecular Cancer* **7**: 1.

Southern, P.J., and Berg, P. (1982). Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *Journal of Molecular and Applied Genetics* **1**: 327-341.



Supplementary file 3. Plasmid maps of pd2EGFP-1 and pGL3-Basic used for the tammar *ELP* promoter reporter assays

A. Tammar *ELP* promoter fragments were originally cloned into the *Xho*I/*Sac*II digested pd2EGFP-1 promoterless vector. **B.** They were then transferred into the *Hind*III cut and blunted-ended and *Xho*I cut pGL3-Basic vector (refer to section 2.3).

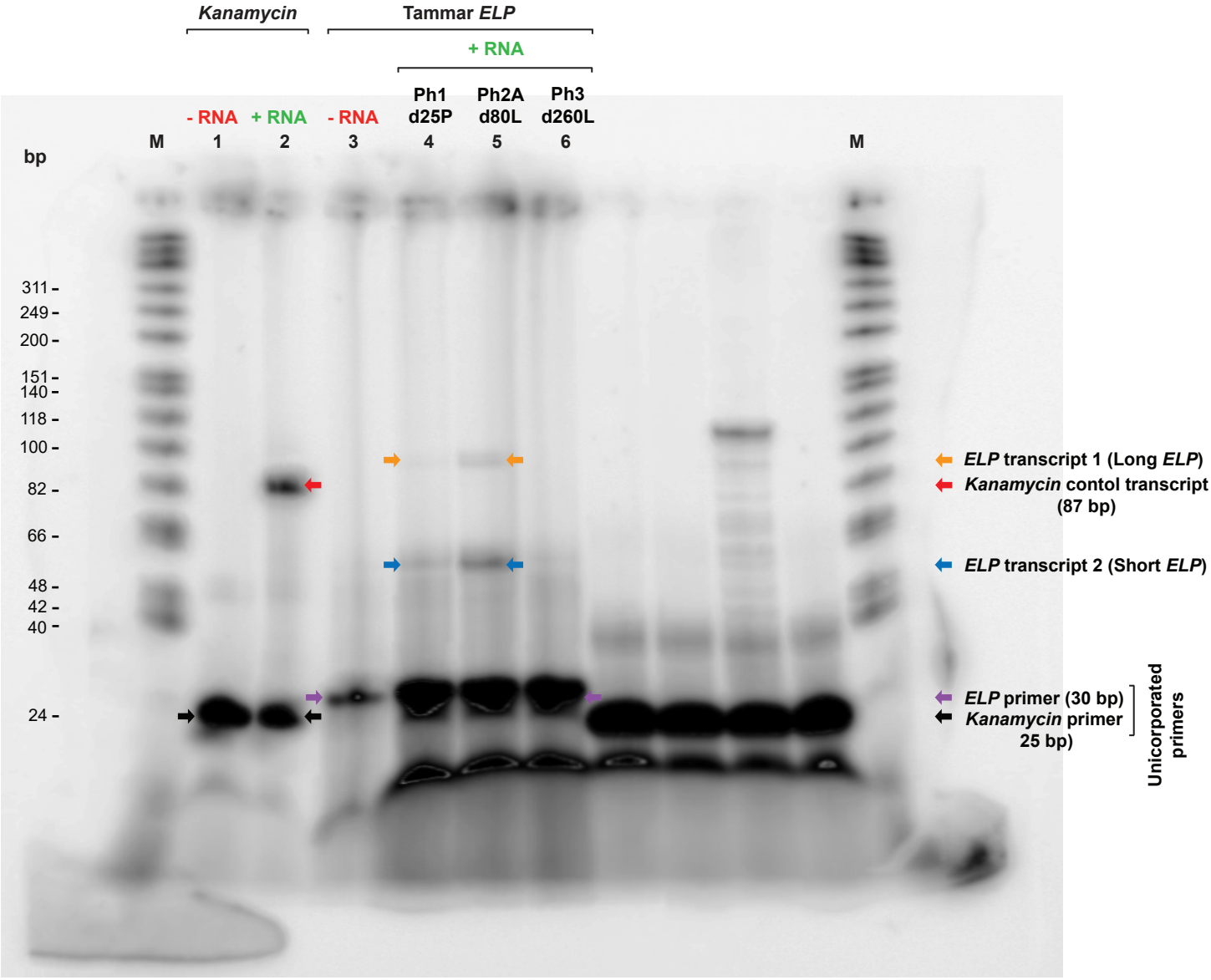
Supplementary file 4 – Accession numbers for sequences used in this study

GENE	COMMON NAME	ACCESSION NUMBER(S) Gene ID/locus
ELP	Tammar wallaby	JN191335
	Koala	JN191337
	Tasmanian devil	BK008719
	Stripe-faced dunnart	AC186006
	Fat-tailed dunnart	JN191339
	South American (Gray short-tailed) opossum	BK008086
CTI	Cape fur seal	KC152480
	Pacific walrus	101362989
	Weddell seal	LOC102745256
	Dog (Labrador, Boxer breeds)	BK008082
	Giant panda	BK008084
	Polar bear	103669871
	Cat (Abyssinian breed)	BK008083
	Cow (Hereford breed)	ENSBTAG00000016127, LOC104973896
	Goat (Yunnan black breed)	LOC102174856
	Sheep (Texel breed)	LOC105613397
	Wild yak	LOC102286180
	Tibetan antelope (Chiru)	LOC102333010
	Water buffalo	LOC102405068
	Pig (Duroc breed)	LOC100513767
	Alpaca	LOC102539473
	Wild Bactrian camel	LOC105087824
	Killer whale	LOC101273120
	Bottle-nosed dolphin	BK008085, LOC101340157
	Scammon’s minke whale	LOC103003268
	Southern white rhinoceros	LOC101393639

Supplementary file 5. Primers and conditions used to amplify selected marsupial *ELP* and eutherian *CTI* promoter regions and to screen HC11 cells for the stable integration of the 14.7kb tammar *ELP* gene

Promoter/Genomic region*	Primer F (5' … 3')	Primer F (5' … 3')	PCR Product Size (bp)	PCR Conditions
Fat-tailed dunnart *ELP*	**Scrass_-594bp_F** (28 bp) CCAAGTATTGCTCTCTCATTTGAACCTC	**Scrass_Ex1_R** (28 bp) GTTTTTCTGAGCTGGTCATGCCAGCCAG	662	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 68°C for 1 min]; 68°C for 10 min.
Koala *ELP*	**ScPF_-414F** (27 bp) CTCACCTCCCAAGTCAAGGCTACTCAG	**ELP_KR_1027R** (28 bp) GGTGATGATGACAGCCATTGTGAGTTTC	1470	94°C for 2 min; 35 cycles of [94°C for 30 s; 54°C for 30 s; 68°C for 4 min]; 68°C for 10 min.
Cape fur seal *CTI*	**Cf_-3301bp_F** (26 bp) GCTCCCTTGTATGCAACTCTCTGGTC	**Cf_Ex_1R** (26 bp) AGAGGACAAGGAACGGGCTGAACTTC	3096	94°C for 2 min; 35 cycles of [94°C for 30 s; 53°C for 30 s; 68°C for 4 min]; 68°C for 10 min.
Tammar *ELP* promoter# (-542/+20)*	**553^_*Xho*I_tELP_F** (30bp) CTG*CTCGAG*ACTTAGGCATCAGAGTGTCTT	**+23^_*Sac*II_tELP_R** (30 bp) GTG*CCGCGG*GTAGACTACACTTGATTAGTG	579	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 72°C for 30 s]; 72°C for 10 min.
Exon 2 to Exon 3 tammar *ELP*# (+3,625/+4,176)*	**ELP_Ex_2F** (24 bp) ACCAAATCCAAGTGAACTCCCTGG	**ELP_Ex_3R** (25 bp) CTTCAGTTATTATTGTTCCTGTTCT	552	94°C for 2 min; 35 cycles of [94°C for 30 s; 55°C for 30 s; 72°C for 30 s]; 72°C for 10 min.

** Gene/promoter region of the construct (relative to the transcription start site (short TSS), +1)*
 # Primer pairs used to screen HC11 cells for the stable integration of the 14.7kb tammar ELP gene
^ Position of the restriction enzyme site incorporated into the promoter construct



Supplementary file 6. Identification of the transcription start sites of the tammar *ELP* gene

Primer extension analysis of tammar *ELP* – original membrane. A 30 bp [γ-^{32}P]-ATP end-labeled *ELP* reverse primer was used to prime first strand *ELP* cDNA synthesis from tammar mammary gland total RNA (10 µg) during pregnancy and lactation. Kanamycin experimental controls were used, consisting of a [γ-^{32}P]-ATP end-labelled Kanamycin control primer (25 bp) with or without 1.2kb Kanamycin Control RNA (10 µg) and formed the positive and negative controls respectively. Products were electrophesed through a 6% polyacrylamide/7 M urea gel, exposed to a PhosphorImager screen and analysed with ImageQuant software. 1-2. Kanamycin controls, 1. -RNA, 2. +RNA, 3-6. tammar *ELP*, 3. –RNA, 4. + Phase 1 (P1) d25P total RNA, 5. +P2A d80L total RNA, 6. +P3 d260L total RNA. Primer extension product sizes were determined with φX174 HinfI DNA Markers (M). Two different *tELP* primer extension products were detected for Phase 1 and 2A (lanes 4 and 5): transcript 1 (long, 96 bp, orange arrows) and transcript 2 (short, 61 bp, blue arrows). The kanamycin control product (87 bp, red arrow), and the unincorporated kanamycin and *ELP* control primers (black and purple arrows respectively) are also shown.

Supplementary file 7. The location of putative transcription factor consensus binding-sites within the tammar wallaby *ELP* gene

High affinity DNA-binding sites are listed, with less-stringent sites also noted (*). Sites are numbered relative to the start of transcription (short transcript), with those within the promoter in black text, exon 1 (purple), intron 1 (blue), exon 2 (red), intron 2 (green) and exon 3 (orange). Publications from which sites were retrieved are also listed.

Transcription Factor DNA-binding site consensus sequence (5' → 3')[1,2]	Position of DNA-binding site	References
TATA box (TBP, TATA binding protein) TATAWAWR, TATA(A/T)A(A/T)A	-1189/-1182, -66/-59, -30/-23, 2489/2496, 2501/2508, 3403/3410	(Ghosh et al., 1998; Wobbe and Struhl, 1990; Yang et al., 2007)
STAT/GAS[3] (Signal transducers and activators of transcription/IFN-γ (interferon gamma activation site) TTNCNNNAA NTT(C/A)(C/T)N(T/G)AAA	-1310/-1302, -1244/-1236, -560/-552, -121/-113, 91/106, 323/331, 499/507, 572/580, 1203/1211, 1244/1252, 2703/2711, 3320/3328, 3728/3736,	(Darnell et al., 1994; Decker et al., 1997)
STAT5 TTCNNNGAA	-560/-552, 1811/1819, 3027/3035	(Burdon et al., 1994; Groner and Gouilleux, 1995; Wakao et al., 1994; Wakao et al., 1992; Watson et al., 1991)
C/EBP (CCAAT/enhancer binding protein T(T/G)NNGNAA(T/G) RTTGCGYAAY	-788/-780, -609/-601, -122/-115, -78/-70, 99/107, 351/359, 500/508, 534/542, 571/579, 1020/1028, 1142/1150, 1226/1234, 1593/1601, 1792/1800, 2002/2008, 2167/2175, 2190/2204, 2702/2710, 3727/3735, 3859/3867	(Doppler et al., 1995; Osada et al., 1996; Raught et al., 1995; Rosen et al., 1998)
MAF (Mammary cell-activating factor) GRRGSAAGK, AGGAAG, CTTCCT(-)	-1518/-1510, -1294/-1289, -859/-851, -409/-404, -123/-115, -79/-71, 479/484, 810/818, 913/921, 1020/1029, 1227/1235, 1248/1253, 1481/1486, 1566/1574, 1634/1639, 1674/1679, 1831/1839, 2168/2176, 2191/2199, 2753/2758, 3070/3075, 3234/3239, 3859/3864, 4019/4027, 4210/4215	(Lenasi et al., 2005; Malewski et al., 2002; Mink et al., 1992; Welte et al., 1994)
ETS (E26 transformation-specific sequence site)GGA(A/T) core	There are numerous sites (too many to list) with a GGA(A/T) core within the tammar *ELP* gene	(Alberstein et al., 2007; Sementchenko and Watson, 2000; Wasylyk et al., 1993; Welte et al., 1994)
GRE/PRE (Glucocorticoid/ Progesterone response element) GGTACA(N)$_3$TGTTCT CT(a/T)CANNNTGT(C/T)CT **GRE half-site** TGTTCT, TGTNCY* **PRE** TGTTCACT	-1601/-1596*, -1476/-1471*, -1456/-1449, -1151/-1146*, -776/-771*, -761/-756*, -131/-126*, 213/218*, 939/944, 952/957*, 1119/1124*, 1147/1152*, 1438/1443*, 1654/1659*, 2121/2126*, 2220/2225, 2255/2269, 2564/2569*, 3495/3600*	(Garlatti et al., 1994; Truss et al., 1991)
YY1 (Yin and Yang-1) (C/G/A)(G/T)(C/T/A)CATN(T/A) (T/G/C)CCATCT, ACATNTT*	-1288/-1283, -129/-123*, 92/97, 363/368, 495/501, 581/587, 662/667, 1866/1871, 1902/1907, 1939/1944, 2463/2468, 2514/2519, 2960/2609, 3080/3085, 3214/3219, 4014/4063	(Hyde-DeRuyscher et al., 1995; Kim et al., 2006; Meier and Groner, 1994; Raught et al., 1994; Yant et al., 1995)
WPPCS (Whey protein promoter core sequence) TGGCARNNWCKKC	-1182/-1170, -1063/-1051, -1028/-1016, -317/-309, -201/-189, 140/152, 1421/1433, 1691/1703, 2950/2962, 4221/4233	(Kolb et al., 1994; Lubon and Hennighausen, 1988)

Transcription Factor DNA-binding site consensus sequence (5' → 3')[1]	Position of DNA-binding site	References
½CTF/NFI (CCAAT-box-transcription factor/Nuclear factor I) TTGGC(N_{2-5})GCCAA **NFI half site:** TTGGC	-1028/-1024, -987/-983, -759/-755, -511/-507, -359/-355, 852/857, 1420/1424, 1573/1577, 1693/1697, 3361/3365, 3470/3475, 4102/4107	(Belikov et al., 2004; Meisterernst et al., 1988; Zorbas et al., 1992)
NF-κB (Nuclear factor κB) GGGRHTYYCC GGG(A/G)NN(A/G)(A/G)CC	-1548/-1539, -1547/-1538, -252/-243, -102/-95, 3315/3326	(Kunsch et al., 1992; Sen and Baltimore, 1986)
PMF (Pregnancy-specific mammary nuclear factor) TGAT(N_{5-8})ATCA, ATCA(N_{5-8})TGAT, TGAAT(N_{4-7})ATTCA	-1637/-1628, -974/-955, 703/737, 1334/1356, 2521/2546, 2530/2543, 4178/4209	(Lee and Oka, 1992a, b; Malewski, 1998)
SP1/SP3 (Specificity protein) GGGCGG **GC Box** GGGCGGG **GT Box** GGTGGGG **GA Box** GGGAGG	-1333/-1328, -228/-223, -87/-82, -40/-35, -36/-31, 128/133, 912/917, 1455/1460, 3608/3613	(Gidoni et al., 1984; Suske, 1999; Yu et al., 1991)
IRF (Interferon regulatory factor) GAAANNGAAANN G(A)AAASYGAAASY **IRF3** GAAA(C/G)(C/G)GAAAN(T/C) **IRF7** GAA(A/T)N(C/T)GAAAN(T/C)	-1321/-1308, -1337/-1326, -63/-51, 1230/1241	(Lin et al., 2000; Tanaka et al., 1993)
E-Box CACGTG, CANNTG	-1659/-1654, -1635/-1630, -985/-980, -966/-980, -943/-938, -653/-648, -110/-105	(Blackwell et al., 1990; Carr and Sharp, 1990; Kako and Ishida, 1998)
IRE/S (Insulin response element/ sequence) **Positive Effect** (no consensus) **Negative Effect** T(G/A)TTT(T/G)(G/T)	-1417/-1411, -518/-512, 560/566, 1317/1325, 1365/1374, 2535/2543, 2580/2588, 2595/2603, 3449/3457, 4055/4063	(O'Brien et al., 2001; Pierreux et al., 1999)
IRE-ABP (IRE-A binding protein) PyC(A/T)TTG(A/T)(A/T)	-1104/-1097	(Buggs et al., 1998)
ACAAAG enhancer (F11/contactin) ACAAAG	-1103/-1098, -779/-774, 949/954, 1122/1127, 2790/2795, 3098/3103	(Mink et al., 1992; Ouatas et al., 2002)
ERE (Estrogen response element) GGTCANNNTG(A/T)CC **ERE half site** TGACC	-1385/-1381, -97/-93, 270/274, 401/405, 632/641, 1522/1526, 1607/1611, 3067/3071, 3182/3186, 3501/3506, 3522/3526	(Duan et al., 2008; Naar et al., 1991)
AP1 (Activating protein-1) TGA(C/G)TCA TGAATCA GGAATCA	1396/1404	(Duan et al., 2008; Lee et al., 1987)
AP2 (Activating protein-2) GCCNNNGGC	-572/-564	(Duan and Clemmons, 1995; Imagawa et al., 1987)
NFY (Nuclear factor Y) (T/C)(A/G)(A/G)CCAATC(A/G)	-683/-672, -197/-185, 1857/1868, 3303/3314	(Bi et al., 1997; Li et al., 1992; Mantovani, 1998; Mitchell and Tjian, 1989)

Transcription Factor DNA-binding site consensus sequence (5' → 3')[1]	Position of DNA-binding site	References
NFAT (Nuclear activator of T-cells) (T/A)GGAAA(A/N)(ATC)N GGAAAN **core**	-522/-513, 346/355, 569/607, 1099/1107, 1465/1473, 4020/4028, 2155/2163, 3151/3159, 2121/2126, 2220/2225, 2255/2269, 3120/3128	(Latinis et al., 1997; Rao et al., 1997)
OCT1 (Octamer protein 1) ATGCAAAT	-973/-961, -966/-953, 203/215, 1499/1511, 2038/2050	(Dong and Zhao, 2007; Kemler et al., 1991; Zhao et al., 2002)
CREB (cAMP response element binding protein, CBP) TGACGTCA	-1628/-1624, -1388/-1381, -1351/-1344, 722/729, 3075/3082	(Daniel et al., 1998; Montminy et al., 1986)
PIT1 (Pituitary factor-1) (A/T)TTATNCAT	2421/2428, 2441/2448	(Ingraham et al., 1990; Ingraham et al., 1988)
GATA1[4] (GATA protein-1) (A/G)GATA(A/G)	-277/-272, 882/887, 1035/1040, 1086/1091, 1344/1349, 2356/2361, 2825/2829, 2826/2831, 3353/3358	(Aird et al., 1994; Ko and Engel, 1993)
GATA2 (GATA protein-2) GATCT, GATTT, GATAA, (A/G)GATA(A/G)	-1638/-1634, -297/-293	(Ko and Engel, 1993)
MZF1, MZF2 (Myeloid zinc finger) AAGAAGGGGA, GGTGGGGA, AGTGGGGA	-880/-871, -136/-129, -49/-42, 116/123, 381/388, 1477/1484, 1469/1478, 4120/4127	(Murai et al., 1998; Zelko and Folz, 2003)
HSF (Heat shock factor) nGAAnnTTCn nTTCnnGAAn	-1652/-1643, -1614/-1605, -49/-42, 1200/1209	(Bonner et al., 1994)

[1]Nucleotide codes: A = adenine; C = cytosine; G = guanine; T = thymine; K = G/T; M = A/C; R= A/G; S = G/C; W = A/T; Y = C/T; B = C/G/T; H = A/C/T; V = A/C/G; N = A/C/G/T.
[2]TFBS are shown for the (+) strand unless otherwise indicated (-)
[3]STAT2 cannot bind to STAT/GAS sites (Wesoly et al., 2007).
[4]There were an additional ten GATA2 sites within the first intron of t*ELP* (data not shown).

References

Aird, W.C., Parvin, J.D., Sharp, P.A., and Rosenberg, R.D. (1994). The interaction of GATA-binding proteins and basal transcription factors with GATA box-containing core promoters. A model of tissue-specific gene expression. J Biol Chem **269**: 883-889.

Alberstein, M., Amit, M., Vaknin, K., O'Donnell, A., Farhy, C., Lerenthal, Y., Shomron, N., Shaham, O., Sharrocks, A.D., Ashery-Padan, R., *et al.* (2007). Regulation of transcription of the RNA splicing factor hSlu7 by Elk-1 and Sp1 affects alternative splicing. RNA **13**: 1988-1999.

Belikov, S., Holmqvist, P.H., Astrand, C., and Wrange, O. (2004). Nuclear factor 1 and octamer transcription factor 1 binding preset the chromatin structure of the mouse mammary tumor virus promoter for hormone induction. J Biol Chem **279**: 49857-49867.

Bi, W., Wu, L., Coustry, F., de Crombrugghe, B., and Maity, S.N. (1997). DNA Binding Specificity of the CCAAT-binding Factor CBF/NF-Y. J Biol Chem **272**: 26562-26572.

Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N., and Weintraub, H. (1990). Sequence-specific DNA binding by the c-Myc protein. Science **250**: 1149-1151.

Bonner, J.J., Ballou, C., and Fackenthal, D.L. (1994). Interactions between DNA-bound trimers of the yeast heat shock factor. Mol Cell Biol **14**: 501-508.

Buggs, C., Nasrin, N., Mode, A., Tollet, P., Zhao, H.F., Gustafsson, J.A., and Alexander-Bridges, M. (1998). IRE-ABP (insulin response element-A binding protein), an SRY-like protein, inhibits C/EBPa (CCAAT/enhancer-binding protein a)-stimulated expression of the sex-specific cytochrome P450 2C12 gene. Mol Endocrinol **12**: 1294-1309.

Burdon, T.G., Demmer, J., Clark, A.J., and Watson, C.J. (1994). The mammary factor MPBF is a prolactin-induced transcriptional regulator which binds to STAT factor recognition sites. FEBS Lett **350**: 177-182.

Carr, C.S., and Sharp, P.A. (1990). A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. Mol Cell Biol **10**: 4384-4388.

Daniel, P.B., Walker, W.H., and Habener, J.F. (1998). Cyclic AMP signaling and gene regulation. Annual Review of Nutrition **18**: 353-383.

Darnell, J.E., Jr., Kerr, I.M., and Stark, G.R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. Science **264**: 1415-1421.

Decker, T., Kovarik, P., and Meinke, A. (1997). GAS elements: a few nucleotides with a major impact on cytokine-induced gene expression. J Interferon Cytokine Res **17**: 121-134.

Dong, B., and Zhao, F.Q. (2007). Involvement of the ubiquitous Oct-1 transcription factor in hormonal induction of b-casein gene expression. Biochem J **401**: 57-64.

Doppler, W., Welte, T., and Philipp, S. (1995). CCAAT/enhancer-binding protein isoforms β and d are expressed in mammary epithelial cells and bind to multiple sites in the b-casein gene promoter. J Biol Chem **270**: 17962-17969.

Duan, C., and Clemmons, D.R. (1995). Transcription factor AP-2 regulates human insulin-like growth factor binding protein-5 gene expression. J Biol Chem **270**: 24844-24851.

Duan, R., Ginsburg, E., and Vonderhaar, B.K. (2008). Estrogen stimulates transcription from the human prolactin distal promoter through AP1 and estrogen responsive elements in T47D human breast cancer cells. Molecular and Cellular Endocrinology **281**: 9-18.

Garlatti, M., Daheshia, M., Slater, E., Bouguet, J., Hanoune, J., Beato, M., and Barouki, R. (1994). A functional glucocorticoid-responsive unit composed of two overlapping inactive receptor-binding sites: evidence for formation of a receptor tetramer. Mol Cell Biol **14**: 8007-8017.

Ghosh, S., May, M.J., and Kopp, E.B. (1998). NF-kB and Rel proteins: evolutionarily conserved mediators of immune responses. Annu Rev Immunol **16**: 225-260.

Gidoni, D., Dynan, W.S., and Tjian, R. (1984). Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature **312**: 409-413.

Groner, B., and Gouilleux, F. (1995). Prolactin-mediated gene activation in mammary epithelial cells. Curr Opin Genet Dev **5**: 587-594.

Hyde-DeRuyscher, R.P., Jennings, E., and Shenk, T. (1995). DNA binding sites for the transcriptional activator/repressor YY1. Nucleic Acids Res **23**: 4457-4465.

Imagawa, M., Chiu, R., and Karin, M. (1987). Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. Cell **51**: 251-260.

Ingraham, H.A., Albert, V.R., Chen, R.P., Crenshaw 3d, E.B., Elsholtz, H.P., He, X., Kapiloff, M.S., Mangalam, H.J., Swanson, L.W., Treacy, M.N., *et al.* (1990). A family of POU-domain and Pit-1 tissue-specific transcription factors in pituitary and neuroendocrine development. Annu Rev Physiol **52**: 773-791.

Ingraham, H.A., Chen, R.P., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L., and Rosenfeld, M.G. (1988). A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell **55**: 519-529.

Kako, K., and Ishida, N. (1998). The role of transcription factors in circadian gene expression. Neurosci Res **31**: 257-264.

Kemler, I., Bucher, E., Seipel, K., Muller-Immergluck, M.M., and Schaffner, W. (1991). Promoters with the octamer DNA motif (ATGCAAAT) can be ubiquitous or cell type-specific depending on binding affinity of the octamer site and Oct-factor concentration. Nucleic Acids Res **19**: 237-242.

Kim, J.D., Hinz, A.K., Bergmann, A., Huang, J.M., Ovcharenko, I., Stubbs, L., and Kim, J. (2006). Identification of clustered YY1 binding sites in imprinting control regions. Genome Res **16**: 901-911.

Ko, L.J., and Engel, J.D. (1993). DNA-binding specificities of the GATA transcription factor family. Mol Cell Biol **13**: 4011-4022.

Kolb, A.F., Gunzburg, W.H., Albang, R., Brem, G., Erfle, V., and Salmons, B. (1994). Negative regulatory elements in the mammary specific whey acidic protein promoter. Journal of Cellular Biochemistry **56**: 245-261.

Kunsch, C., Ruben, S.M., and Rosen, C.A. (1992). Selection of optimal kappa B/Rel DNA-binding motifs: interaction of both subunits of NF-kappa B with DNA is required for transcriptional activation. Mol Cell Biol **12**: 4412-4421.

Latinis, K.M., Norian, L.A., Eliason, S.L., and Koretzky, G.A. (1997). Two NFAT transcription factor binding sites participate in the regulation of CD95 (Fas) ligand expression in activated human T cells. J Biol Chem **272**: 31427-31434.

Lee, C.S., and Oka, T. (1992a). A pregnancy-specific mammary nuclear factor involved in the repression of the mouse b-casein gene transcription by progesterone. J Biol Chem **267**: 5797-5801.

Lee, C.S., and Oka, T. (1992b). Progesterone regulation of a pregnancy-specific transcription repressor to b-casein gene promoter in mouse mammary gland. Endocrinology **131**: 2257-2262.

Lee, W., Haslinger, A., Karin, M., and Tjian, R. (1987). Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature **325**: 368-372.

Lenasi, T., Kokalj-Vokac, N., Narat, M., Baldi, A., and Dovc, P. (2005). Functional study of the equine b-casein and k-casein gene promoters. J Dairy Res **72 Spec No**: 34-43.

Li, X.Y., Mantovani, R., Hooft van Huijsduijnen, R., Andre, I., Benoist, C., and Mathis, D. (1992). Evolutionary variation of the CCAAT-binding transcription factor NF-Y. Nucleic Acids Res **20**: 1087-1091.

Lin, R., Genin, P., Mamane, Y., and Hiscott, J. (2000). Selective DNA binding and association with the CREB binding protein coactivator contribute to differential activation of a/b interferon genes by interferon regulatory factors 3 and 7. Mol Cell Biol **20**: 6342-6353.

Lubon, H., and Hennighausen, L. (1988). Conserved region of the rat a-lactalbumin promoter is a target site for protein binding *in vitro*. Biochem J **256**: 391-396.

Malewski, T. (1998). Computer analysis of distribution of putative *cis*- and *trans*-regulatory elements in milk protein gene promoters. Biosystems **45**: 29-44.

Malewski, T., Gajewska, M., Zebrowska, T., and Zwierzchowski, L. (2002). Differential induction of transcription factors and expression of milk protein genes by prolactin and growth hormone in the mammary gland of rabbits. Growth Horm IGF Res **12**: 41-53.

Mantovani, R. (1998). A survey of 178 NF-Y binding CCAAT boxes. Nucleic Acids Res **26**: 1135-1143.

Meier, V.S., and Groner, B. (1994). The nuclear factor YY1 participates in repression of the beta-casein gene promoter in mammary epithelial cells and is counteracted by mammary gland factor during lactogenic hormone induction. Mol Cell Biol **14**: 128-137.

Meisterernst, M., Gander, I., Rogge, L., and Winnacker, E.L. (1988). A quantitative analysis of nuclear factor I/DNA interactions. Nucleic Acids Res **16**: 4419-4435.

Mink, S., Hartig, E., Jennewein, P., Doppler, W., and Cato, A.C. (1992). A mammary cell-specific enhancer in mouse mammary tumor virus DNA is composed of multiple regulatory elements including binding sites for CTF/NFI and a novel transcription factor, mammary cell-activating factor. Mol Cell Biol **12**: 4906-4918.

Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**: 371-378.

Montminy, M.R., Sevarino, K.A., Wagner, J.A., Mandel, G., and Goodman, R.H. (1986). Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. Proc Natl Acad Sci U S A **83**: 6682-6686.

Murai, K., Murakami, H., and Nagata, S. (1998). Myeloid-specific transcriptional activation by murine myeloid zinc-finger protein 2. Proc Natl Acad Sci U S A **95**: 3461-3466.

Naar, A.M., Boutin, J.M., Lipkin, S.M., Yu, V.C., Holloway, J.M., Glass, C.K., and Rosenfeld, M.G. (1991). The orientation and spacing of core DNA-binding motifs dictate selective transcriptional responses to three nuclear receptors. Cell **65**: 1267-1279.

O'Brien, R.M., Streeper, R.S., Ayala, J.E., Stadelmaier, B.T., and Hornbuckle, L.A. (2001). Insulin-regulated gene expression. Biochem Soc Trans **29**: 552-558.

Osada, S., Yamamoto, H., Nishihara, T., and Imagawa, M. (1996). DNA binding specificity of the CCAAT/enhancer-binding protein transcription factor family. J Biol Chem **271**: 3891-3896.

Ouatas, T., Clare, S.E., Hartsough, M.T., De La Rosa, A., and Steeg, P.S. (2002). MMTV-associated transcription factor binding sites increase nm23-H1 metastasis suppressor gene expression in human breast carcinoma cell lines. Clin Exp Metastasis **19**: 35-42.

Pierreux, C.E., Rousseau, G.G., and Lemaigre, F.P. (1999). Insulin inhibition of glucocorticoid-stimulated gene transcription: requirement for an insulin response element? Molecular and Cellular Endocrinology **147**: 1-5.

Rao, A., Luo, C., and Hogan, P.G. (1997). Transcription factors of the NFAT family: regulation and function. Annual Review of Immunology **15**: 707-747.

Raught, B., Khursheed, B., Kazansky, A., and Rosen, J. (1994). YY1 represses b-casein gene expression by preventing the formation of a lactation-associated complex. Mol Cell Biol **14**: 1752-1763.

Raught, B., Liao, W.S., and Rosen, J.M. (1995). Developmentally and hormonally regulated CCAAT/enhancer-binding protein isoforms influence b-casein gene expression. Mol Endocrinol **9**: 1223-1232.

Rosen, J.M., Zahnow, C., Kazansky, A., and Raught, B. (1998). Composite response elements mediate hormonal and developmental regulation of milk protein gene expression. Biochem Soc Symp **63**: 101-113.

Sementchenko, V.I., and Watson, D.K. (2000). Ets target genes: past, present and future. Oncogene **19**: 6533-6548.

Sen, R., and Baltimore, D. (1986). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell **46**: 705-716.

Suske, G. (1999). The Sp-family of transcription factors. Gene **238**: 291-300.

Tanaka, N., Kawakami, T., and Taniguchi, T. (1993). Recognition DNA sequences of interferon regulatory factor 1 (IRF-1) and IRF-2, regulators of cell growth and the interferon system. Mol Cell Biol **13**: 4531-4538.

Truss, M., Chalepakis, G., Slater, E.P., Mader, S., and Beato, M. (1991). Functional interaction of hybrid response elements with wild-type and mutant steroid hormone receptors. Mol Cell Biol **11**: 3247-3258.

Wakao, H., Gouilleux, F., and Groner, B. (1994). Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response. EMBO J **13**: 2182-2191.

Wakao, H., Schmitt-Ney, M., and Groner, B. (1992). Mammary gland-specific nuclear factor is present in lactating rodent and bovine mammary tissue and composed of a single polypeptide of 89 kDa. J Biol Chem **267**: 16365-16370.

Wasylyk, B., Hahn, S.L., and Giovane, A. (1993). The Ets family of transcription factors. Eur J Biochem **211**: 7-18.

Watson, C.J., Gordon, K.E., Robertson, M., and Clark, A.J. (1991). Interaction of DNA-binding proteins with a milk protein gene promoter *in vitro*: identification of a mammary gland-specific factor. Nucl Acids Res **19**: 6603-6610.

Welte, T., Garimorth, K., Philipp, S., Jennewein, P., Huck, C., Cato, A.C., and Doppler, W. (1994). Involvement of Ets-related proteins in hormone-independent mammary cell-specific gene expression. Eur J Biochem **223**: 997-1006.

Wesoly, J., Szweykowska-Kulinska, Z., and Bluyssen, H.A. (2007). STAT activation and differential complex formation dictate selectivity of interferon responses. Acta Biochim Pol **54**: 27-38.

Wobbe, C.R., and Struhl, K. (1990). Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. Mol Cell Biol **10**: 3859-3867.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene **389**: 52-65.

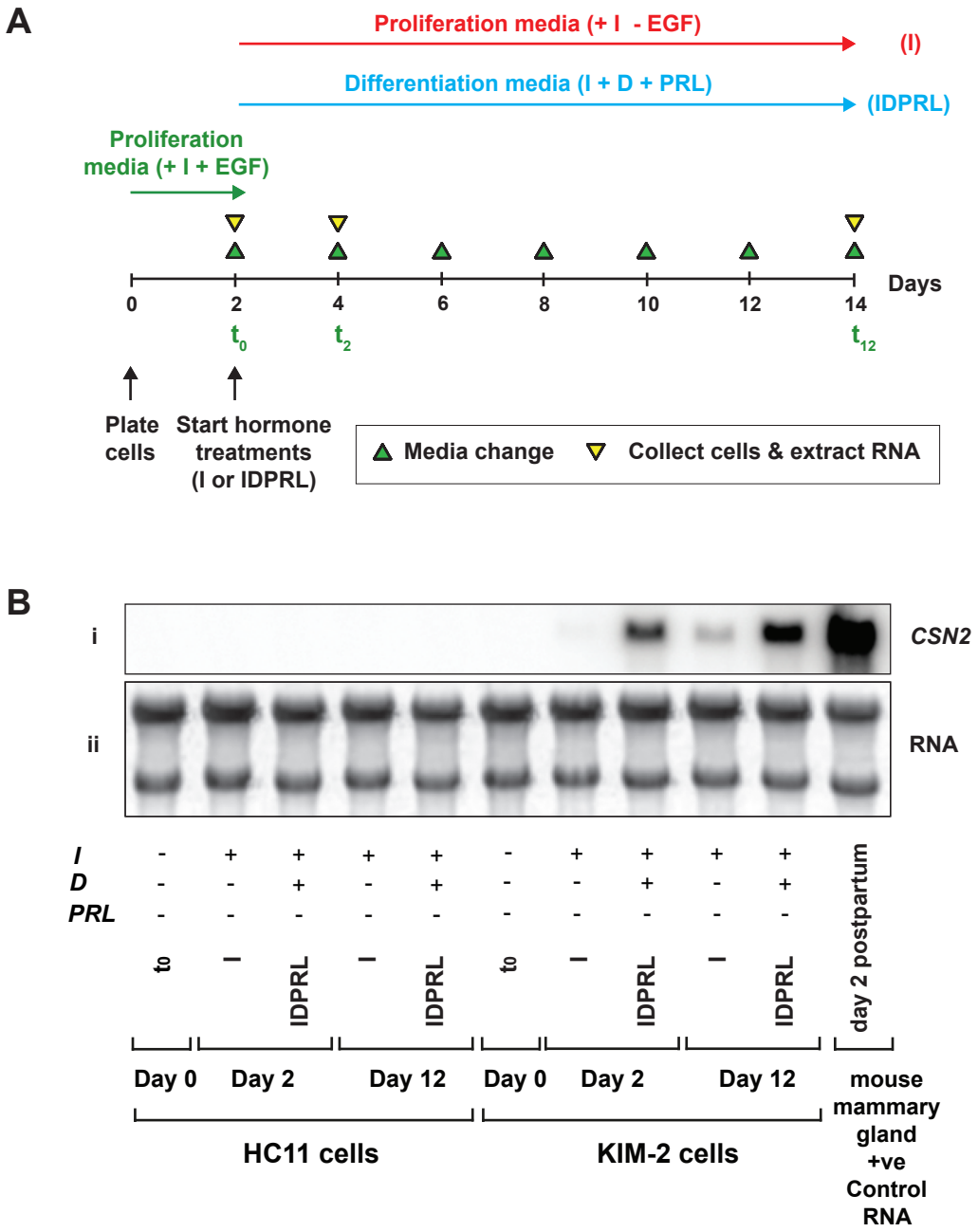
Yant, S.R., Zhu, W., Millinoff, D., Slightom, J.L., Goodman, M., and Gumucio, D.L. (1995). High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster. Nucleic Acids Res **23**: 4353-4362.

Yu, C.Y., Motamed, K., Chen, J., Bailey, A.D., and Shen, C.K. (1991). The CACC box upstream of human embryonic epsilon globin gene binds Sp1 and is a functional promoter element in vitro and in vivo. J Biol Chem **266**: 8907-8915.

Zelko, I.N., and Folz, R.J. (2003). Myeloid zinc finger (MZF)-like, Kruppel-like and Ets families of transcription factors determine the cell-specific expression of mouse extracellular superoxide dismutase. Biochem J **369**: 375-386.

Zhao, F.Q., Adachi, K., and Oka, T. (2002). Involvement of Oct-1 in transcriptional regulation of b-casein gene expression in mouse mammary gland. Biochim Biophys Acta **1577**: 27-37.

Zorbas, H., Rein, T., Krause, A., Hoffmann, K., and Winnacker, E.L. (1992). Nuclear factor I (NF I) binds to an NF I-type site but not to the CCAAT site in the human alpha-globin gene promoter. J Biol Chem **267**: 8478-8484.

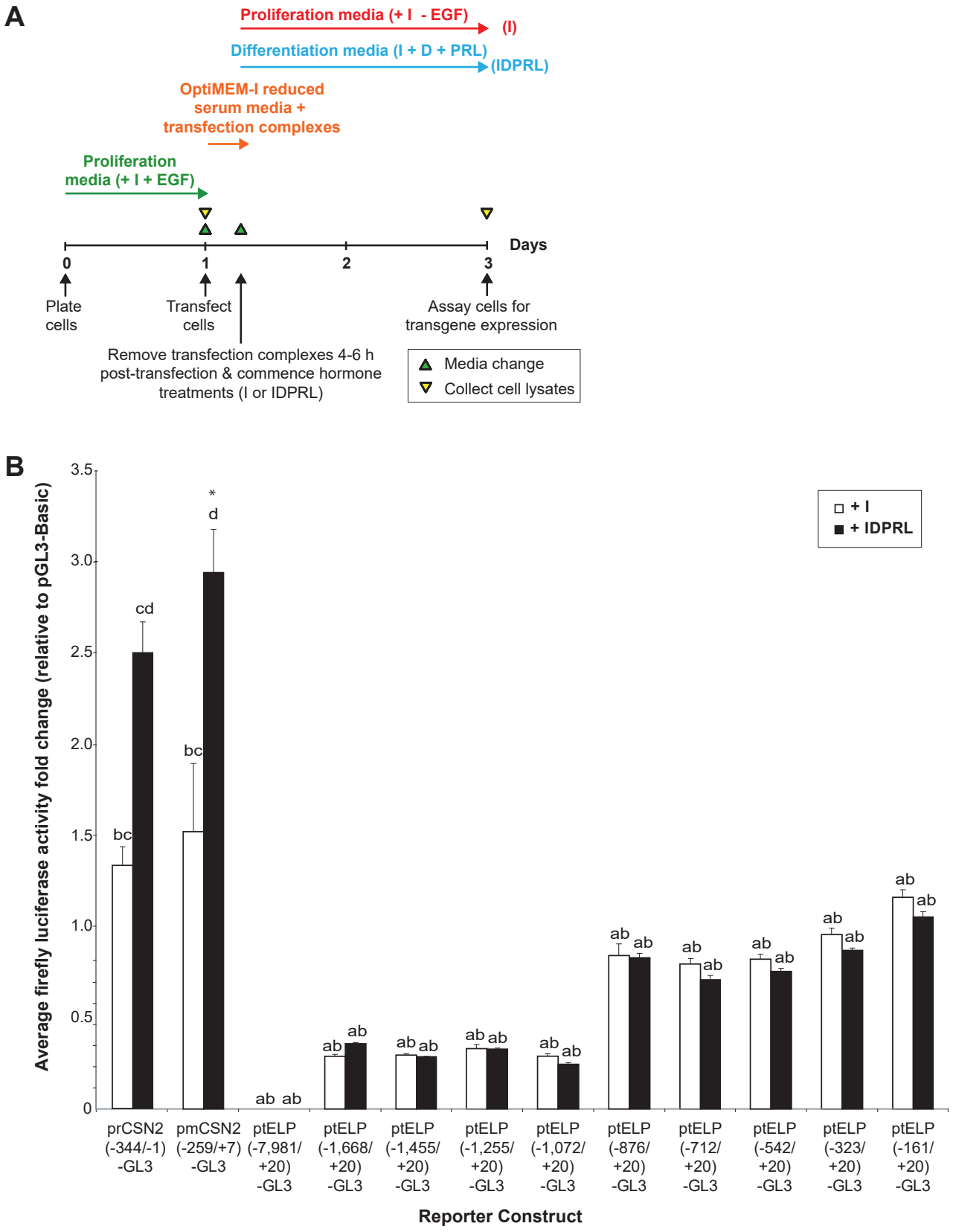


Supplementary file 8. Prolactin-responsiveness of HC11 and KIM-2 murine mammary epithelial cells in an *in vitro* culture system

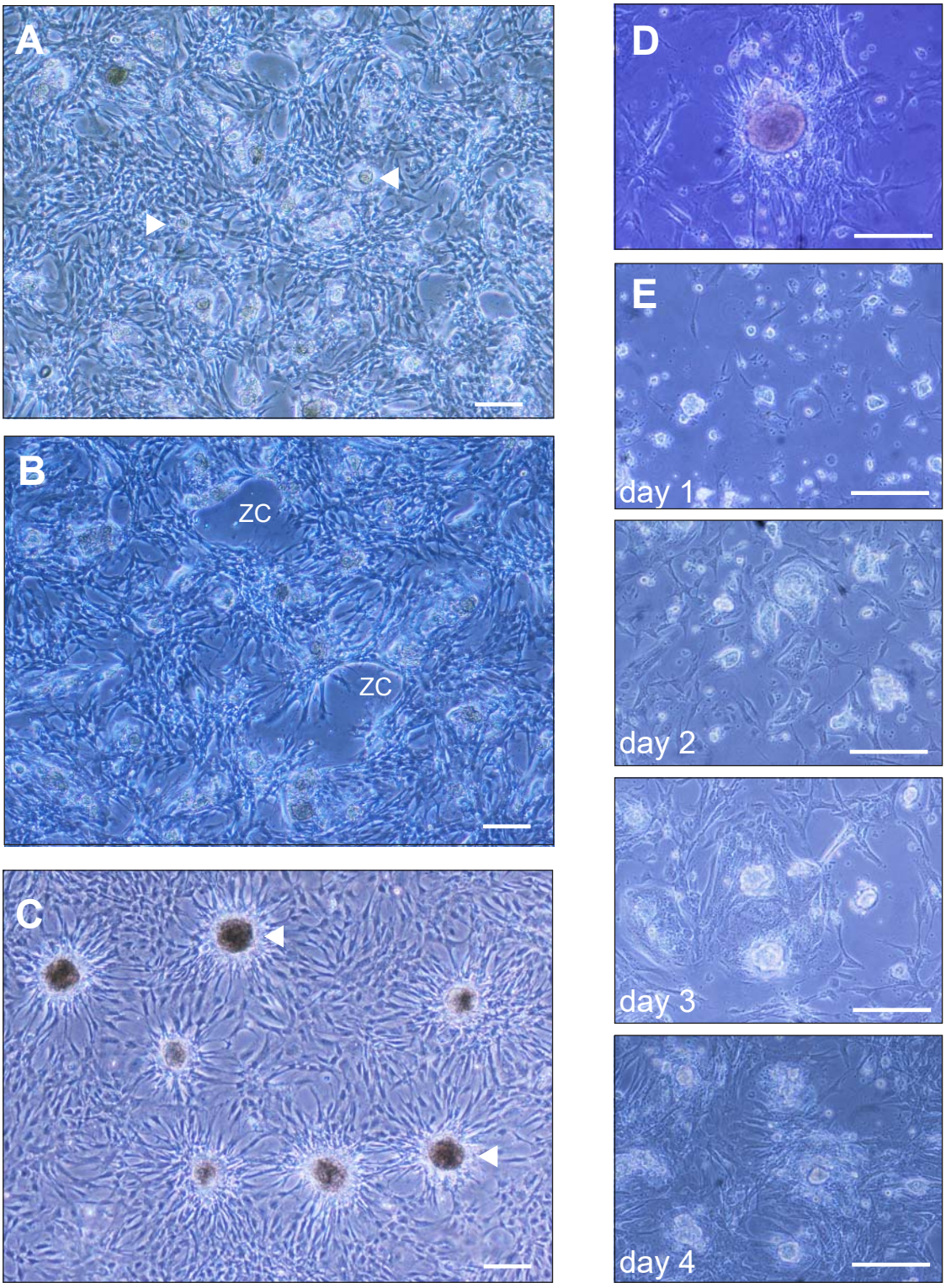
A. Experiment design. The protocol for the differentiation of HC11 and KIM-2 cells, i.e., the induction of milk protein gene expression, was based upon the procedure for KIM-2 cells (Gordon, *et al*. 2000). Cells were grown to confluence in 6-well plates (35 mm wells) in proliferation media (for ~2 days, t_0). Milk protein gene expression induced with differentiation media [5 µg/mL insulin, 1 µM dexamethasone and 5 µg/mL prolactin (IDPRL treatment)]. The negative control for prolactin-responsiveness was Proliferation media -EGF (I treatment). All treatments were performed in duplicate. **B.** Induction of murine *CSN2* expression in HC11 and KIM-2 cells was determined by Northern analysis of total RNA (10 µg). *CSN2* expression was not detected in HC11 cells (i). In contrast, endogenous *CSN2* was induced in KIM-2 cells treated with IDPRL. *CSN2* expression was highest in positive control RNA isolated from the mouse mammary gland at day 2 of lactation (d2L), as expected. Ethidium bromide staining of ribosomal bands indicates equivalence of sample loading and RNA integrity (ii).

Reference

Gordon, K.E., Binas, B., Chapman, R.S., Kurian, K.M., Clarkson, R.W., Clark, A.J., Lane, E.B., and Watson, C.J. (2000). A novel cell culture model for studying differentiation and apoptosis in the mouse mammary gland. Breast Cancer Res 2, 222-235.

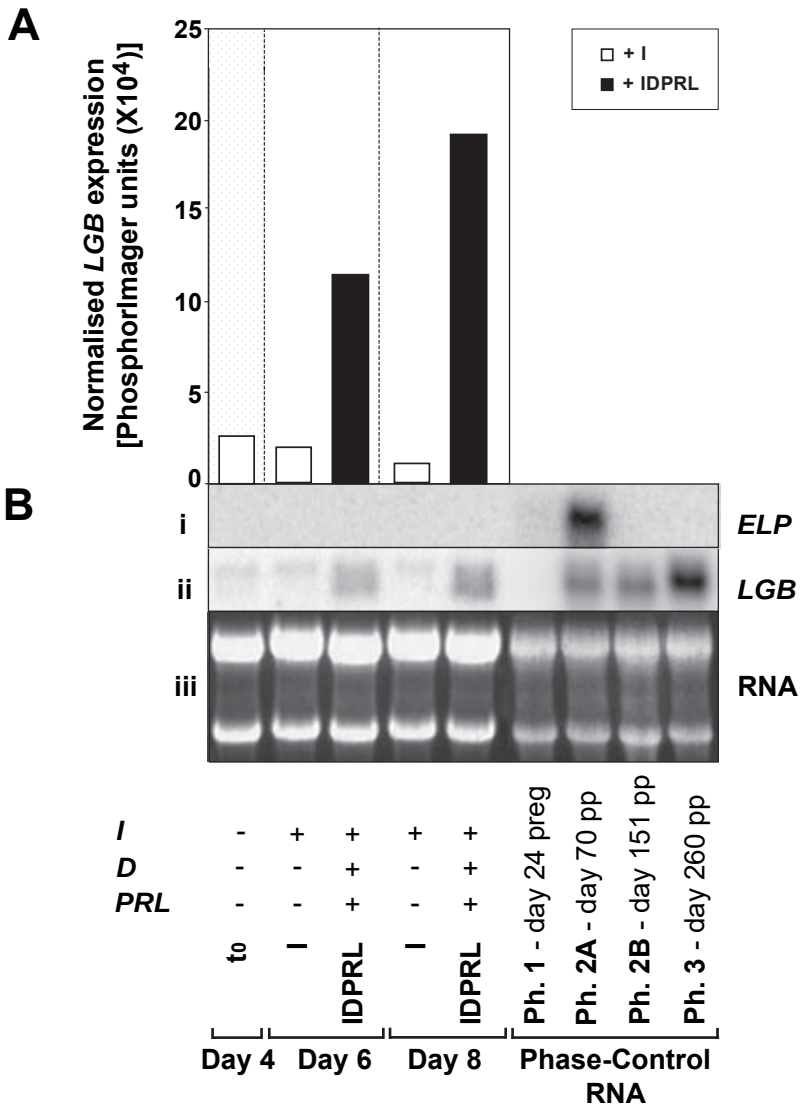


Supplementary file 9. Activity of tammar *ELP* promoter constructs in KIM-2 cells
KIM-2 cells were transiently co-transfected with tammar *ELP*, or rat or mouse *CSN* promoter constructs and pSV-β-Galactosidase as an internal transfection control. A. Experiment design. B. Results represent average firefly luciferase activity normalised to β-Gal activity and expressed as fold change relative to pGL3-Basic activity ± SEM (n=2) for each of two experiments. The PRL-responsiveness of KIM-2 cells was confirmed by the increased luciferase activity of the IDRPRL-treated murine *CSN2* promoter construct compared to the insulin alone treatment ($p<0.05$). In contrast, there was no significant difference in luciferase activity for any of the tammar *ELP* promoter constructs between the I and IDPRL treatments ($p>0.05$).



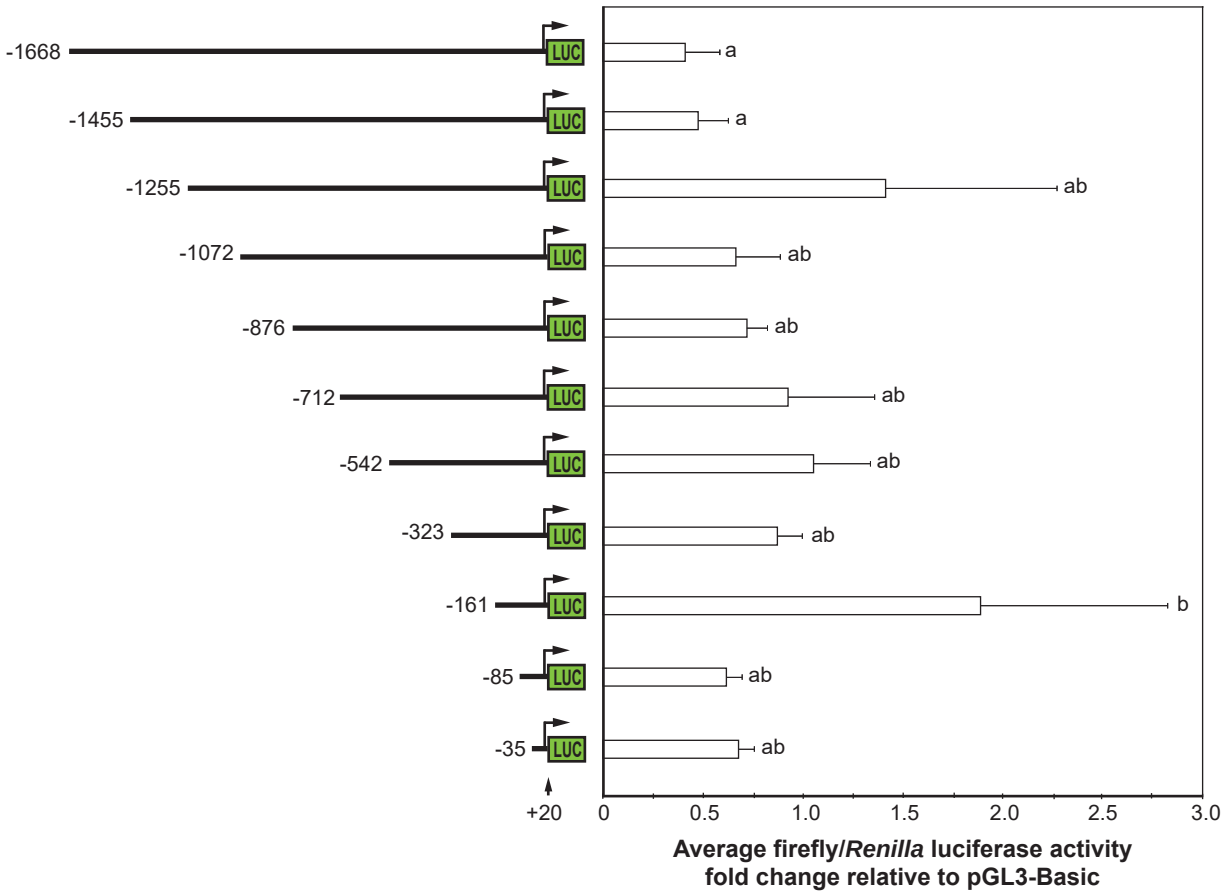
Supplementary file 10. Growth and differentiation of tammar primary mammary gland cells (PMGCs) on tissue culture plastic

A. When cultured in proliferation media, PMGCs harvested from tammar mammary tissue (day 16 of pregnancy, d16P) spontaneously formed aggregates of cells (immature mammospheres, white arrowheads) to which cells were recruited. **B**. This phenomenon produced zones of clearing (ZC) on the TCP. **C.** Larger, but fewer mature mammospheres (white arrowheads) and an increase in fibroblast numbers are favoured by continuous passaging of PMGCs. **D.** Mammosphere formation. **E.** Proliferation of d16P PMGCs over a 4-day period (day 1, day 2, day 3, day 4). Magnification: **A**, **B**, **C** 40X, **D**, **E** (days 1 to 4, inclusive) 200X. Scale bars: **A**, **B**, **C** = 200 µm, **D**, **E** (days 1 to day 4, inclusive) = 100 µm.



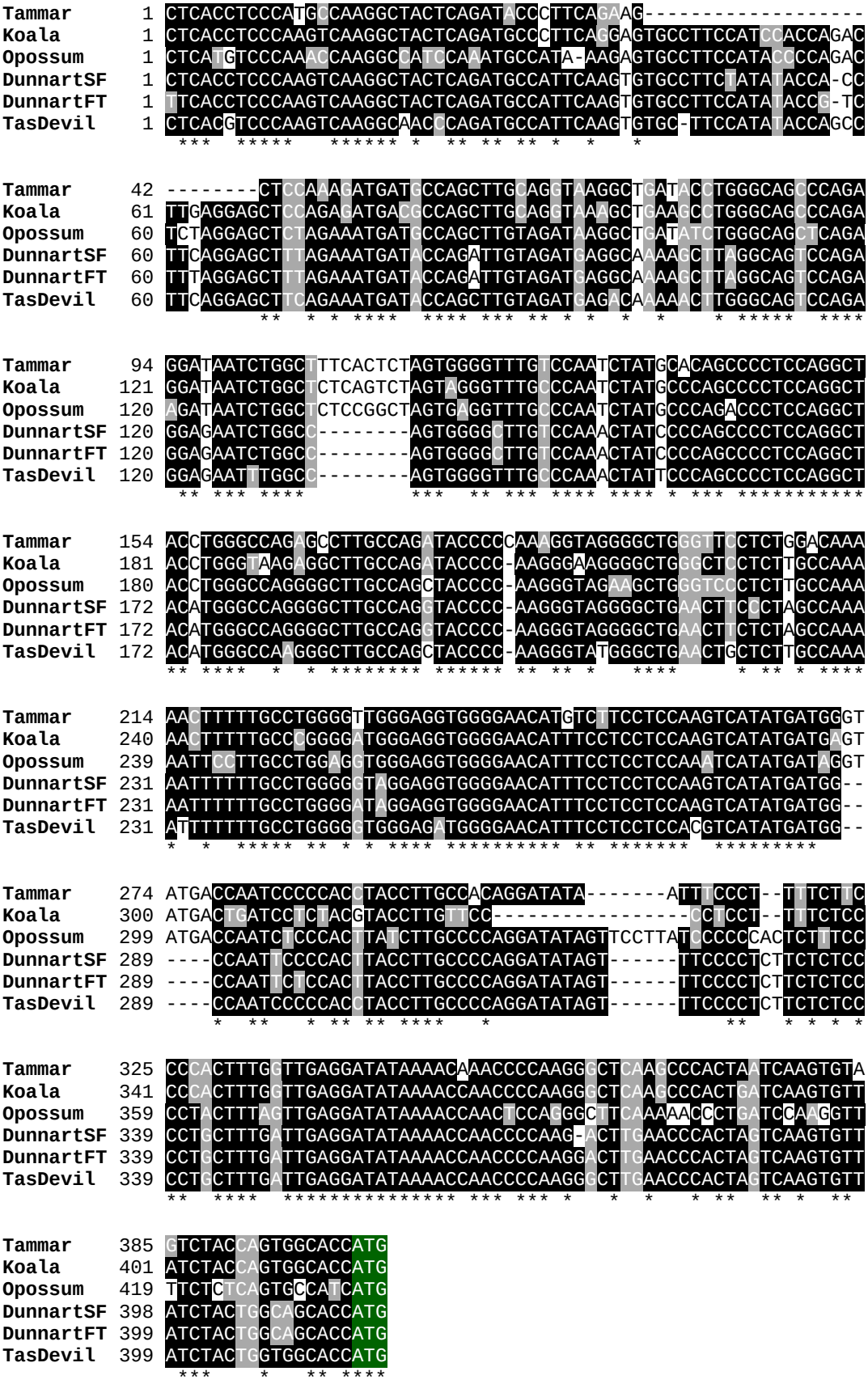
Supplementary file 11. Lactogenic hormone-responsiveness of tammar primary mammary gland cells (PMGCs)

Tammar PMGCs from d16P were grown to confluence in culture dishes (35 mm wells) in proliferation media for 3 days and cultured for an additional day in proliferation media minus EGF (day 4). On day 4, milk protein gene expression was induced over a 4-day period with differentiation media supplemented with insulin (5 μg/mL), dexamethasone (1 μM) and ovine prolactin (5 μg/mL) (IDPRL treatment). Proliferation media minus EGF (I treatment) formed the negative control. Media was changed on day 6 and cells collected on both day 6 and day 8. All treatments were performed in duplicate (n=2). Total RNA (10 μg) was extracted from cells and expression of *ELP* and *LGB* assayed by Northern Analysis. **A.** *LGB* gene expression (PhosphorImager units normalised to RNA loading) was much higher in IDPRL-treated PMGCs compared to those treated with insulin alone. *LGB* expression increased over the 4-day period of treatment with lactogenic hormones. **B.** Tammar (i) *ELP* expression was not detected in PMGCs by Northern Analysis, unlike (ii) *LGB*. The two tammar *LGB* transcripts were visible in the IDPRL-treated cells only (iii) Ribosomal bands indicate RNA integrity and loading.

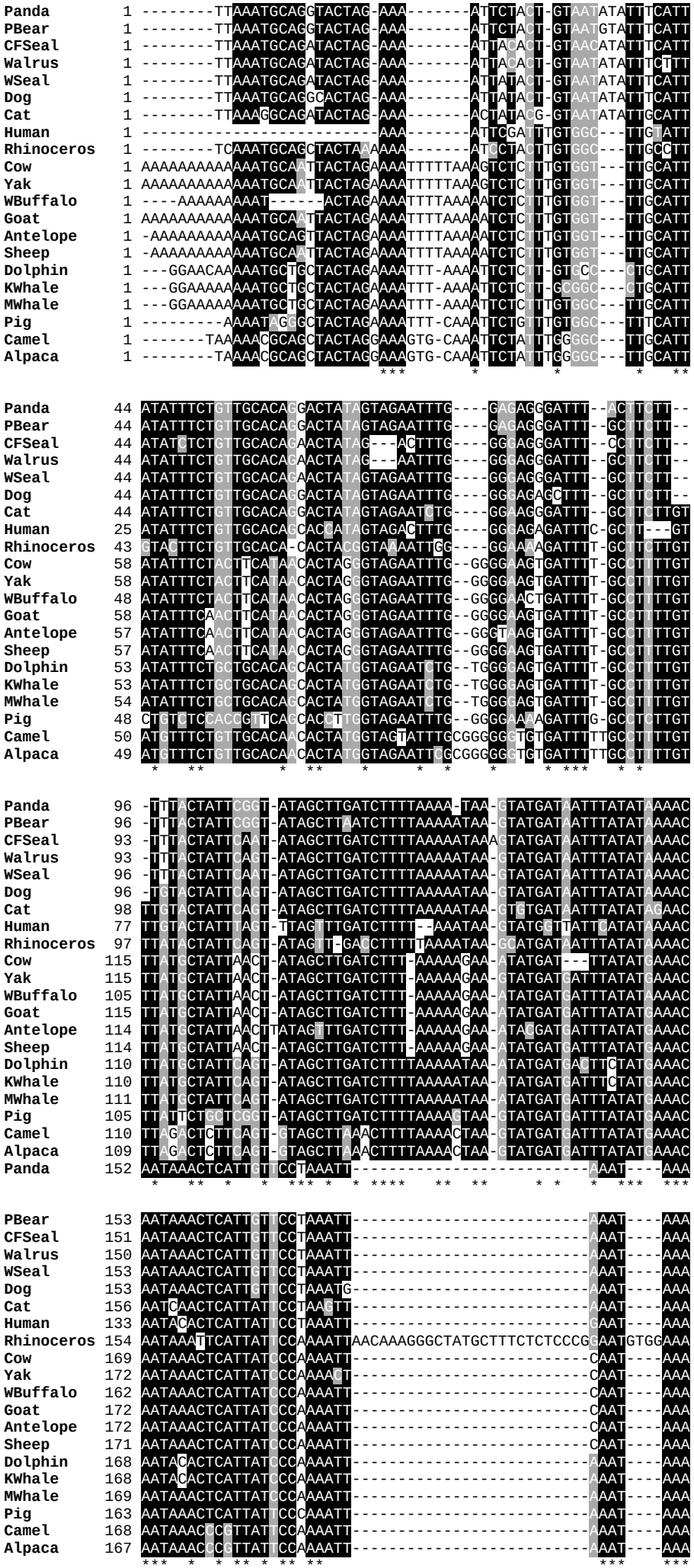


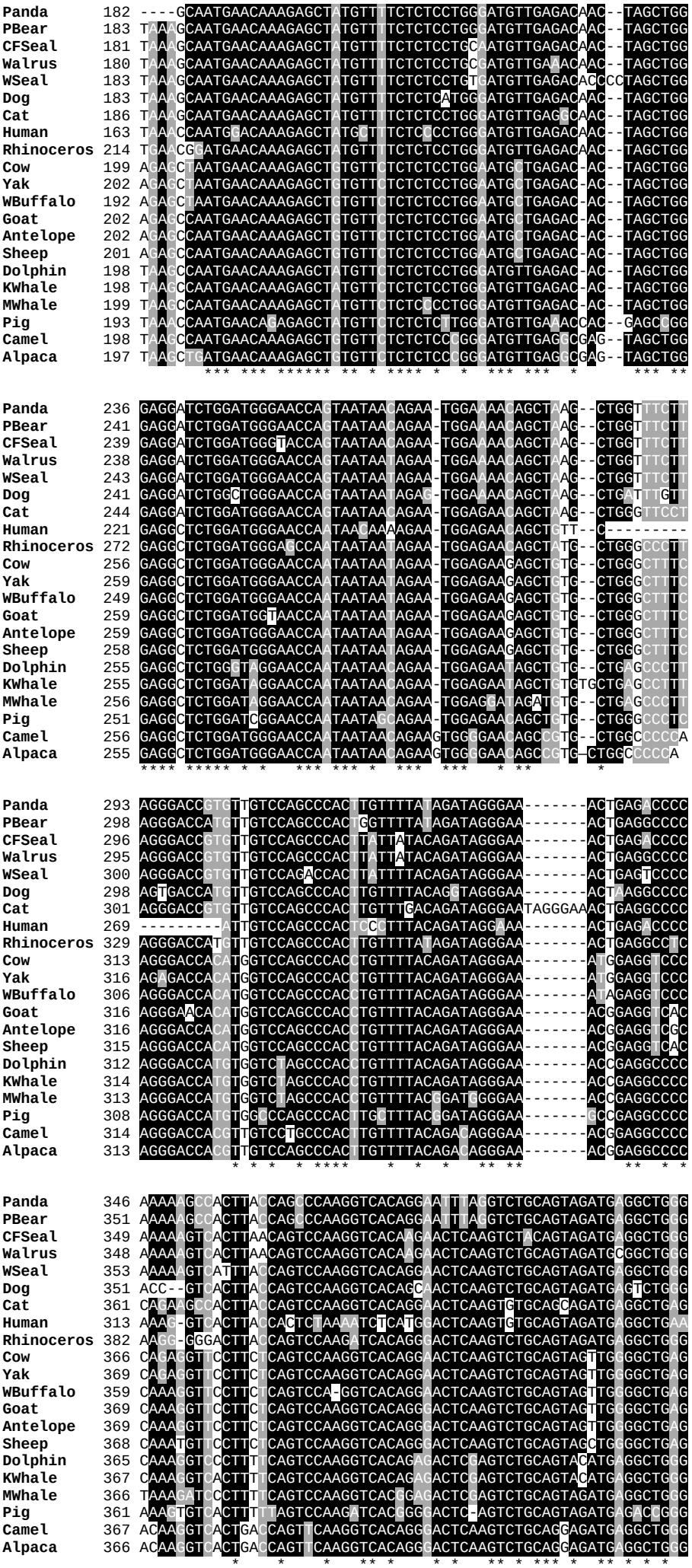
Supplementary file 12. Hormone-independent activity of tammar *ELP* promoter constructs in MDA MB 231 cells

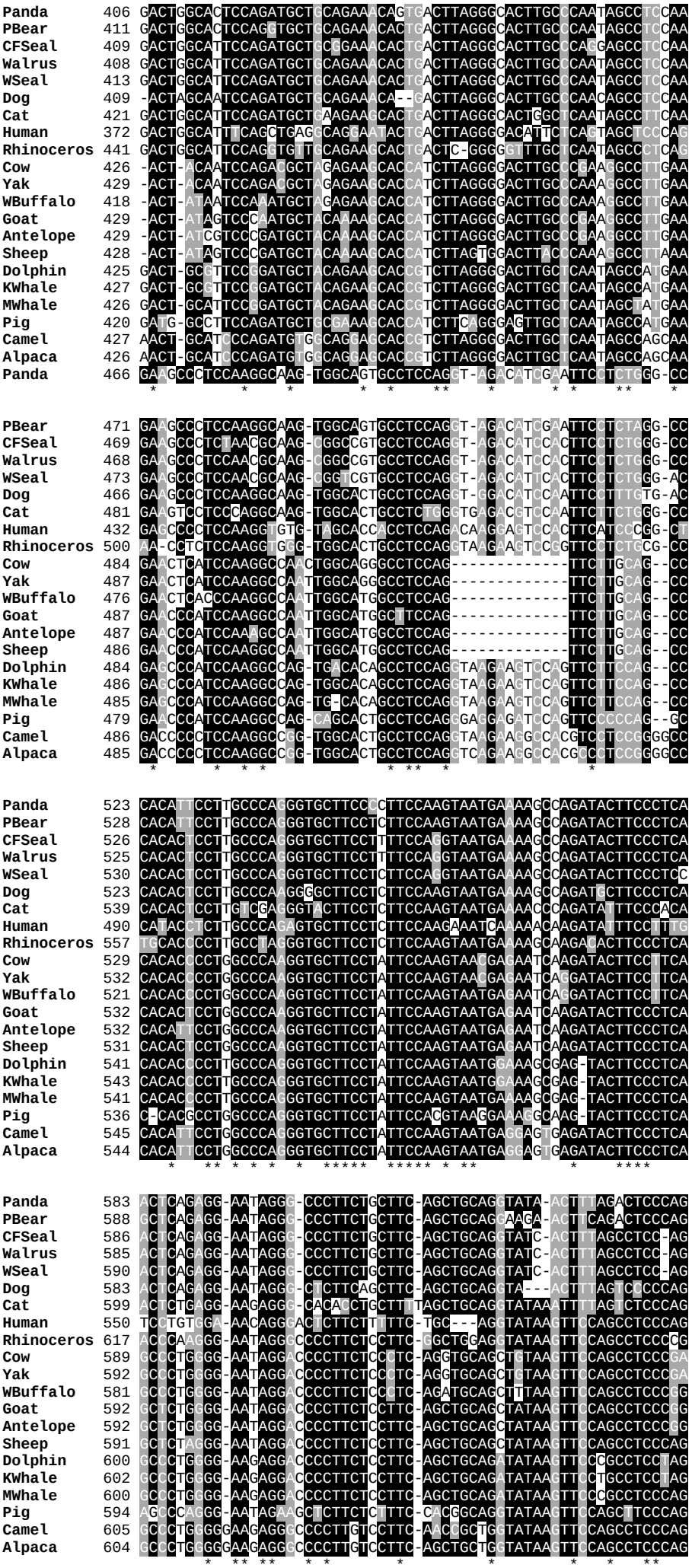
MDA MB 231 cells were transiently co-transfected with tammar *ELP* promoter constructs in the firefly luciferase plasmid pGL3-Basic and the *Renilla* luciferase pRL-CMV as an internal transfection control. Results represent average firefly/*Renilla* luciferase activity expressed as fold change relative to pGL3-Basic activity ± SEM (n=3) for two duplicate experiments. Although average firefly/*Renilla* activity was highest for the -161/+20 tammar *ELP* promoter construct, activity was only significantly higher ($p<0.05$) than for the -118/+20 and -1455/+20 *ELP* promoter constructs.

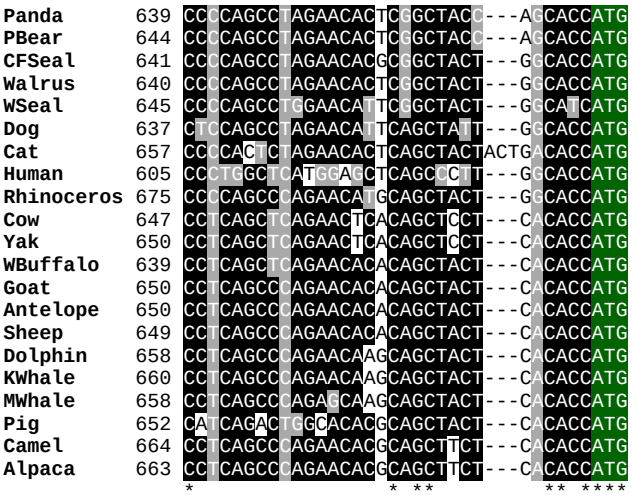


Supplementary file 13. Alignment of marsupial *ELP* proximal promoters
ClustalW2 alignment of the marsupial *ELP* proximal promoters. Nucleotides identical in all 6 sequences are indicated by an asterisk (*), those shared by at least 70% of sequences (5 out of 6) are shaded black and conserved nucleotide types (either purines or pyrimidines) shared by more than 70% of sequences are shaded grey. The putative translation start site is shaded green.

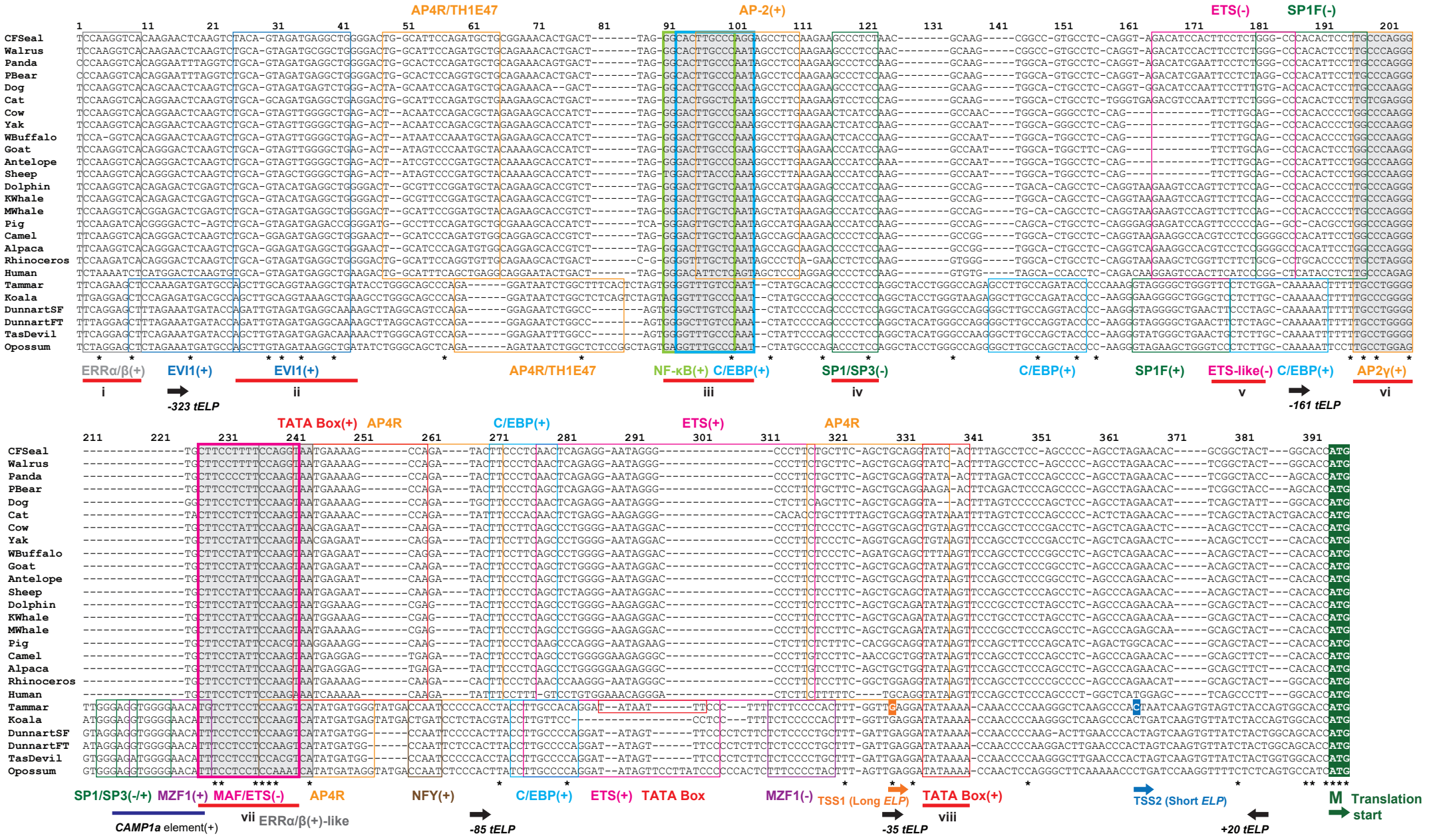

Panda 1 --------TTAAATGCAGGTACTAG-AAA-------ATTCTACT-GTAATATATTTCATT
PBear 1 --------TTAAATGCAGGTACTAG-AAA-------ATTCTACT-GTAATGTATTTCATT
CFSeal 1 --------TTAAATGCAGATACTAG-AAA-------ATTACACT-GTAACATATTTCATT
Walrus 1 --------TTAAATGCAGATACTAG-AAA-------ATTACACT-GTAATATATTTCTTT
WSeal 1 --------TTAAATGCAGATACTAG-AAA-------ATTATACT-GTAATATATTTCATT
Dog 1 --------TTAAATGCAGGCACTAG-AAA-------ATTATACT-GTAATATATTTCATT
Cat 1 --------TTAAAGGCAGATACTAG-AAA-------ACTATACG-GTAATATATTGCATT
Human 1 ------------------------AAA-------ATTCGATTTGTGGC---TTGTATT
Rhinoceros 1 --------TCAAATGCAGCTACTAAAAA-------ATGCTACTTGTGGC---TTGCCTT
Cow 1 AAAAAAAAAAAAATGCAATTACTAGAAAATTTTTAAAGTCTCTTTGTGGT---TTGCATT
Yak 1 AAAAAAAAAAAAATGCAATTACTAGAAAATTTTTAAAGTCTCTTTGTGGT---TTGCATT
WBuffalo 1 ----AAAAAAAAAT------ACTAGAAAATTTTAAAAATCTCTTTGTGGT---TTGCATT
Goat 1 AAAAAAAAAAAAATGCAATTACTAGAAAATTTTAAAAATCTCTTTGTGGT---TTGCATT
Antelope 1 -AAAAAAAAAAAATGCAGTTACTAGAAAATTTTAAAAATCTCTTTGTGGT---TTGCATT
Sheep 1 -AAAAAAAAAAAATGCAATTACTAGAAAATTTTAAAAATCTCTTTGTGGT---TTGCATT
Dolphin 1 ---GGAACAAAAATGCTGCTACTAGAAAATTT-AAAATTCTCTT-GTCCC---CTGCATT
KWhale 1 ---GGAAAAAAAATGCTGCTACTAGAAAATTT-AAAATTCTCTT-GCGGC---CTGCATT
MWhale 1 ---GGAAAAAAAATGCTGCTACTAGAAAATTT-AAAATTCTCTTTGTGGC---TTGCATT
Pig 1 ---------AAAATAGGGCTACTAGAAAATTT-CAAATTCTGTTTGTGGC---TTTCATT
Camel 1 -------TAAAAACGCAGCTACTAGGAAAGTG-CAAATTCTATTTGGGGC---TTGCATT
Alpaca 1 --------TAAAACGCAGCTACTAGGAAAGTG-CAAATTCTATTTGGGGC---TTGCATT
Panda 44 ATATTTCTGTTGCACAGGACTATAGTAGAATTTG----GAGAGGGATTT--ACTTCTT--
PBear 44 ATATTTCTGTTGCACAGGACTATAGTAGAATTTG----GAGAGGGATTT--GCTTCTT--
CFSeal 44 ATATCTCTGTTGCACAGAACTATAG---ACTTTG----GGGAGGGATTT--CCTTCTT--
Walrus 44 ATATTTCTGTTGCACAGAACTATAG---AATTTG----GGGAGGGATTT--GCTTCTT--
WSeal 44 ATATTTCTGTTGCACAGAACTATAGTAGAATTTG----GGGAGGGATTT--GCTTCTT--
Dog 44 ATATTTCTGTTGCACAGGACTATAGTAGAATTTG----GGGAGAGCTTT--GCTTCTT--
Cat 44 ATATTTCTGTTGCACAGGACTATAGTAGAATCTG----GGAAGGGATTT--GCTTCTTGT
Human 25 ATATTTCTGTTGCACAGCACCATAGTAGACTTTG----GGGAGAGATTTC-GCTT---GT
Rhinoceros 43 GTACTTCTGTTGCACA-CACTACGGTAAAATTGC----GGAAAAGATTTT-GCTTCTTGT
Cow 58 ATATTTCTACTTCATAACACTAGGGTAGAATTTG--GGGGAAGTGATTTT-GCCTTTTGT
Yak 58 ATATTTCTACTTCATAACACTAGGGTAGAATTTG--GGGGAAGTGATTTT-GCCTTTTGT
WBuffalo 48 ATATTTCTACTTCATAACACTAGGGTAGAATTTG--GGGGAACTGATTTT-GCCTTTTGT
Goat 58 ATATTTCAACTTCATAACACTAGGGTAGAATTTG--GGGGAAGTGATTTT-GCCTTTTGT
Antelope 57 ATATTTCAACTTCATAACACTAGGGTAGAATTTG--GGGTAAGTGATTTT-GCCTTTTGT
Sheep 57 ATATTTCAACTTCATAACACTAGGGTAGAATTTG--GGGGAAGTGATTTT-GCCTTTTGT
Dolphin 53 ATATTTCTGCTGCACAGCACTATGGTAGAATCTG--TGGGGAGTGATTTT-GCCTTTTGT
KWhale 53 ATATTTCTGCTGCACAGCACTATGGTAGAATCTG--TGGGGAGTGATTTT-GCCTTTTGT
MWhale 54 ATATTTCTGCTGCACAGCACTATGGTAGAATCTG--TGGGGAGTGATTTT-GCCTTTTGT
Pig 48 CTGTGTCCACCGTTCAGCACCTTGGTAGAATTTG--GGGGAAAAGATTTG-GCCTCTTGT
Camel 50 ATGTTTCTGTTGCACAACACTATGGTAGTATTTGCGGGGGCTGTGATTTTTGCCTTTTGT
Alpaca 49 ATGTTTCTGTTGCACAACACTATGGTAGAATTGCGGGGGCTGTGATTTTTGCCTTTTGT
Panda 96 -TTTACTATTCGGT-ATAGCTTGATCTTTTAAAA-TAA-GTATGATAATTTATATAAAAC
PBear 96 -TTTACTATTCGGT-ATAGCTTAATCTTTTAAAAATAA-GTATGATAATTTATATAAAAC
CFSeal 93 -TTTACTATTCAAT-ATAGCTTGATCTTTTAAAAATAAAGTATGATAATTTATATAAAAC
Walrus 93 -TTTACTATTCAGT-ATAGCTTGATCTTTTAAAAATAA-GTATGATAATTTATATAAAAC
WSeal 96 -TTTACTATTCAAT-ATAGCTTGATCTTTTAAAAATAA-GTATGATAATTTATATAAAAC
Dog 96 -TGTACTATTCAGT-ATAGCTTGATCTTTTAAAAATAA-GTATGATAATTTATATAAAAC
Cat 98 TTGTACTATTCAGT-ATAGCTTGATCTTTTAAAAATAA-GTGTGATAATTTATATAGAAC
Human 77 TTGTACTATTTAGT-TTAGTTTGATCTTTT--AAATAA-GTATGGTTATTCATATAAAAC
Rhinoceros 97 TTATACTATTCAGT-ATAGTT-GACCTTTTTAAAATAA-GCATGATAATTTATATAAAAC
Cow 115 TTATGCTATTAACT-ATAGCTTGATCTTT-AAAAAGAAA-ATATGAT---TTATATGAAAC
Yak 115 TTATGCTATTAACT-ATAGCTTGATCTTT-AAAAAGAAA-GTATGATGATTTATATGAAAC
WBuffalo 105 TTATGCTATTAACT-ATAGCTTGATCTTT-AAAAAGAAA-ATATGATGATTTATATAAAAC
Goat 115 TTATGCTATTAACT-ATAGCTTGATCTTT-AAAAAGAAA-ATATGATGATTTATATGAAAC
Antelope 114 TTATGCTATTAACTTATAGTTGATCTTT-AAAAAGAAA-ATACGATGATTTATATGAAAC
Sheep 114 TTATGCTATTAACT-ATAGCTTGATCTTT-AAAAAGAAA-ATATGATGATTTATATGAAAC
Dolphin 110 TTATGCTATTCAGT-ATAGCTTGATCTTTTAAAAATAA-ATATGATGACTTCTATGAAAC
KWhale 110 TTATGCTATTCAGT-ATAGCTTGATCTTTTAAAAATAA-ATATGATGATTTCTATGAAAC
MWhale 111 TTATGCTATTCAGT-ATAGCTTGATCTTTTAAAAATAA-ATATGATGATTTATATGAAAC
Pig 105 TTATTCTGCTCGGT-ATAGCTTGATCTTTTAAAAGTAA-GTATGATGATTTATATGAAAC
Camel 110 TTAGACTCTTCAGT-GTAGCTTAAACTTTTAAAACTAA-GTATGATGATTTATATGAAAC
Alpaca 109 TTAGACTCTTCAGT-GTAGCTTAAACTTTTAAAACTAA-GTATGATGATTTATATGAAAC
Panda 152 AATAAACTCATTGTTCCTAAATT----------AAAT----AAA
PBear 153 AATAAACTCATTGTTCCTAAATT----------AAAT----AAA
CFSeal 151 AATAAACTCATTGTTCCTAAATT----------AAAT----AAA
Walrus 150 AATAAACTCATTGTTCCTAAATT----------AAAT----AAA
WSeal 153 AATAAACTCATTGTTCCTAAATT----------AAAT----AAA
Dog 153 AATAAACTCATTGTTCCTAAATG----------AAAT----AAA
Cat 156 AATCAACTCATTATTCCTAAGTT----------AAAT----AAA
Human 133 AATACACTCATTATTCCTAAATT----------GAAT----AAA
Rhinoceros 154 AATAAATTCATTATTCCAAAATTAACAAAGGGCTATGCTTTCTCTCCCGGAATGTGGAAA
Cow 169 AATAAACTCATTATCCCAAAATT----------CAAT----AAA
Yak 172 AATAAACTCATTATCCCAAAACT----------CAAT----AAA
WBuffalo 162 AATAAACTCATTATCCCAAAATT----------CAAT----AAA
Goat 172 AATAAACTCATTATCCCAAAATT----------CAAT----AAA
Antelope 172 AATAAACTCATTATCCCAAAATT----------CAAT----AAA
Sheep 171 AATAAACTCATTATCCCAAAATT----------CAAT----AAA
Dolphin 168 AATACACTCATTATCCCAAAATT----------AAAT----AAA
KWhale 168 AATACACTCATTATCCCAAAATT----------AAAT----AAA
MWhale 169 AATAAACTCATTATCCCAAAATT----------AAAT----AAA
Pig 163 AATAAACTCATTATTCCCAAATT----------AAAT----AAA
Camel 168 AATAAACCCGTTATTCCAAAATT----------AAAT----AAA
Alpaca 167 AATAAACCCGTTATTCCAAAATT----------AAAT----AAA


Panda 182 ----GCAATGAACAAAGAGCTATGTTTTCTCTCCTGGGATGTTGAGACAAC--TAGCTGG
PBear 183 TAAAGCAATGAACAAAGAGCTATGTTTTCTCTCCTGGGATGTTGAGACAAC--TAGCTGG
CFSeal 181 TAAAGCAATGAACAAAGAGCTATGTTTTCTCTCCTGCAATGTTGAGACAAC--TAGCTGG
Walrus 180 TAAAGCAATGAACAAAGAGCTATGTTTTCTCTCCTGCGATGTTGAAACAAC--TAGCTGG
WSeal 183 TAAAGCAATGAACAAAGAGCTATGTTTTCTCTCCTGTGATGTTGAGACACCCCTAGCTGG
Dog 183 TAAAGCAATGAACAAAGAGCTATGTTTTCTCTCATGGGATGTTGAGACAAC--TAGCTGG
Cat 186 TAAAGCAATGAACAAAGAGCTATGTTTTCTCTCCTGGGATGTTGAGGCAAC--TAGCTGG
Human 163 TAAACCAATGGACAAAGAGCTATGCTTTCTCCCCTGGGATGTTGAGACAAC--TAGCTGG
Rhinoceros 214 TGAACGGATGAACAAAGAGCTATGTTTTCTCTCCTGGGATGTTGAGACAAC--TAGCTGG
Cow 199 AGAGCTAATGAACAAAGAGCTGTGTTGTCTCTCCTGGAATGCTGAGAC-AC--TAGCTGG
Yak 202 AGAGCTAATGAACAAAGAGCTGTGTTGTCTCTCCTGGAATGCTGAGAC-AC--TAGCTGG
WBuffalo 192 AGAGCTAATGAACAAAGAGCTGTGTTGTCTCTCCTGGAATGCTGAGAC-AC--TAGCTGG
Goat 202 AGAGCCAATGAACAAAGAGCTGTGTTGTCTCTCCTGGAATGCTGAGAC-AC--TAGCTGG
Antelope 202 AGAGCCAATGAACAAAGAGCTGTGTTGTCTCTCCTGGAATGCTGAGAC-AC--TAGCTGG
Sheep 201 AGAGCCAATGAACAAAGAGCTGTGTTGTCTCTCCTGGAATGCTGAGAC-AC--TAGCTGG
Dolphin 198 TAAGCCAATGAACAAAGAGCTATGTTGTCTCTCCTGGGATGTTGAGAC-AC--TAGCTGG
KWhale 198 TAAGCCAATGAACAAAGAGCTATGTTGTCTCTCCTGGGATGTTGAGAC-AC--TAGCTGG
MWhale 199 TAAGCCAATGAACAAAGAGCTATGTTGTCTCCCCTGGGATGTTGAGAC-AC--TAGCTGG
Pig 193 TAAACCAATGAACAGAGAGCTATGTTGTCTCTCTTGGGATGTTGAAACCAC--GAGCGGG
Camel 198 TAAGCCAATGAACAAAGAGCTGTGTTGTCTCTCCCGGGATGTTGAGGCGAG--TAGCTGG
Alpaca 197 TAAGCTGATGAACAAAGAGCTGTGTTGTCTCTCCCGGGATGTTGAGGCGAG--TAGCTGG
*** *** ****** ** * **** * * *** *** * *** **
Panda 236 GAGGATCTGGATGGGAACCAGTAATAACAGAA-TGGAAAAGAGCTAAG--CTGGTTTCTT
PBear 241 GAGGATCTGGATGGGAACCAGTAATAACAGAA-TGGAAAAGAGCTAAG--CTGGTTTCTT
CFSeal 239 GAGGATCTGGATGGGTACCAGTAATAACAGAA-TGGAAAAGAGCTAAG--CTGGTTTCTT
Walrus 238 GAGGATCTGGATGGGAACCAGTAATAATAGAA-TGGAAAACAGCTAAG--CTGGTTTCTT
WSeal 243 GAGGATCTGGATGGGAACCAGTAATAATAGAA-TGGAAAACAGCTAAG--CTGGTTTCTT
Dog 241 GAGGATCTGGCTGGGAACCAGTAATAATAGAG-TGGAAAAGAGCTAAG--CTGATTTGTT
Cat 244 GAGGATCTGGATGGGAACCAGTAATAACAGAA-TGGAGAAGAGCTAAG--CTGGGTTCCT
Human 221 GAGGCTCTGGATGGGAACCAATAACAAAGAAA-TGGAGAAGAGCTGTT--C---------
Rhinoceros 272 GAGGCTCTGGATGGGAGCCAATAATAATAGAA-TGGAGAAGAGCTATG--CTGGGCCCTT
Cow 256 GAGGCTCTGGATGGGAACCAATAATAATAGAA-TGGAGAAGAGCTGTG--CTGGGCTTTC
Yak 259 GAGGCTCTGGATGGGAACCAATAATAATAGAA-TGGAGAAGAGCTGTG--CTGGGCTTTC
WBuffalo 249 GAGGCTCTGGATGGGAACCAATAATAATAGAA-TGGAGAAGAGCTGTG--CTGGGCTTTC
Goat 259 GAGGCTCTGGATGGTAACCAATAATAATAGAA-TGGAGAAGAGCTGTG--CTGGGCTTTC
Antelope 259 GAGGCTCTGGATGGGAACCAATAATAATAGAA-TGGAGAAGAGCTGTG--CTGGGCTTTC
Sheep 258 GAGGCTCTGGATGGGAACCAATAATAATAGAA-TGGAGAAGAGCTGTG--CTGGGCTTTC
Dolphin 255 GAGGCTCTGGGTAGGAACCAATAATAACAGAA-TGGAGAATAGCTGTG--CTGAGCCCTT
KWhale 255 GAGGCTCTGGATAGGAACCAATAATAACAGAA-TGGAGAATAGCTGTGTGCTGAGCCTTT
MWhale 256 GAGGCTCTGGATAGGAACCAATAATAACAGAA-TGGAGGATAGATGTG--CTGAGCCCTT
Pig 251 GAGGCTCTGGATCGGAACCAATAATAGCAGAA-TGGAGAAGAGCTGTG--CTGGGCCCTC
Camel 256 GAGGCTCTGGATGGGAACCAATAATAACAGAAGTGGGGAACAGCCGTG--CTGGCCCCCA
Alpaca 255 GAGGCTCTGGATGGGAACCAATAATAACAGAAGTGGGGAACAGCCGTG—CTGGCCCCCA
**** ***** * * *** *** * *** *** * ** * *
Panda 293 AGGGACCGTGTTGTCCAGCCCACTTGTTTTATAGATAGGGAA-------ACTGAGACCCC
PBear 298 AGGGACCATGTTGTCCAGCCCACTGGTTTTATAGATAGGGAA-------ACTGAGGCCCC
CFSeal 296 AGGGACCGTGTTGTCCAGCCCACTTATTATACAGATAGGGAA-------ACTGAGACCCC
Walrus 295 AGGGACCGTGTTGTCCAGCCCACTTATTATACAGATAGGGAA-------ACTGAGGCCCC
WSeal 300 AGGGACCGTGTTGTCCAGACCACTTATTTTACAGATAGGGAA-------ACTGAGTCCCC
Dog 298 AGTGACCATGTTGTCCAGCCCACTTGTTTTACAGGTAGGGAA-------ACTAAGGCCCC
Cat 301 AGGGACCGTGTTGTCCAGCCCACTTGTTTGACAGATAGGGAATAGGGAAACTGAGGCCCC
Human 269 ----------ATTGTCCAGCCCACTCCCTTTACAGATAGGAAA-------ACTGAGACCCC
Rhinoceros 329 AGGGACCATGTTGTCCAGCCCACTTGTTTTATAGATAGGGAA-------ACTGAGGCCTC
Cow 313 AGGGACCACATGGTCCAGCCCACCTGTTTTACAGATAGGGAA-------ATGGAGGTCCC
Yak 316 AGAGACCACATGGTCCAGCCCACCTGTTTTACAGATAGGGAA-------ATGGAGGTCCC
WBuffalo 306 AGGGACCACATGGTCCAGCCCACCTGTTTTACAGATAGGGAA-------ATAGAGGTCCC
Goat 316 AGGGAACACATGGTCCAGCCCACCTGTTTTACAGATAGGGAA-------ACGGAGGTCAC
Antelope 316 AGGGACCACATGGTCCAGCCCACCTGTTTTACAGATAGGGAA-------ACGGAGGTCGC
Sheep 315 AGGGACCACATGGTCCAGCCCACCTGTTTTACAGATAGGGAA-------ACGGAGGTCAC
Dolphin 312 AGGGACCATGTGGTCTAGCCCACCTGTTTTACAGATAGGGAA-------ACCGAGGCCCC
KWhale 314 AGGGACCATGTGGTCTAGCCCACCTGTTTTACAGATAGGGAA-------ACCGAGGCCCC
MWhale 313 AGGGACCATGTGGTCTAGCCCACCTGTTTTACGGATGGGGAA-------ACCGAGGCCCC
Pig 308 AGGGACCATGTGGCCCAGCCCACTTGGTTTACGGATAGGGAA-------GCCGAGGCCCC
Camel 314 AGGGACCACGTTGTCCTGCCCACTTGTTTTACAGACAGGGAA-------ACGGAGGCCCC
Alpaca 313 AGGGACCACGTTGTCCAGCCCACTTGTTTTACAGACAGGGAA-------ACGGAGGCCCC
* * * * **** * * * ** ** ** * *
Panda 346 AAAAAGCCACTTACCAGCCCAAGGTCACAGGAATTTAGGTCTGCAGTAGATGAGGCTGGG
PBear 351 AAAAAGCCACTTACCAGCCCAAGGTCACAGGAATTTAGGTCTGCAGTAGATGAGGCTGGG
CFSeal 349 AAAAAGTCACTTAACAGTCCAAGGTCACAAGAACTCAAGTCTACAGTAGATGAGGCTGGG
Walrus 348 AAAAAGTCACTTAACAGTCCAAGGTCACAAGAACTCAAGTCTGCAGTAGATGCGGCTGGG
WSeal 353 AAAAAGTCATTTACCAGTCCAAGGTCACAGGAACTCAAGTCTGCAGTAGATGAGGCTGGG
Dog 351 ACC--GTCACTTACCAGTCCAAGGTCACAGCAACTCAAGTCTGCAGTAGATGAGTCTGGG
Cat 361 CAGAAGCCACTTACCAGTCCAAGGTCACAGGAACTCAAGTGTGCAGCAGATGAGGCTGAG
Human 313 AAAG-GTCACTTACCACTCTAAAATCTCATGGACTCAAGTCTGCAGTAGATGAGGCTGAA
Rhinoceros 382 AAGG-GGGACTTACCAGTCCAAGATCACAGGGACTCAAGTCTGCAGTAGATGAGGCTGGG
Cow 366 CAGAGGTTCCTTCTCAGTCCAAGGTCACAGGAACTCAAGTCTGCAGTAGTTGGGGCTGAG
Yak 369 CAGAGGTTCCTTCTCAGTCCAAGGTCACAGGAACTCAAGTCTGCAGTAGTTGGGGCTGAG
WBuffalo 359 CAAAGGTTCCTTCTCAGTCCA-GGTCACAGGAACTCAAGTCTGCAGTAGTTGGGGCTGAG
Goat 369 CAAAGGTTCCTTCTCAGTCCAAGGTCACAGGGACTCAAGTCTGCAGTAGTTGGGGCTGAG
Antelope 369 CAAAGGTTCCTTCTCAGTCCAAGGTCACAGGGACTCAAGTCTGCAGTAGTTGGGGCTGAG
Sheep 368 CAAATGTTCCTTCTCAGTCCAAGGTCACAGGGACTCAAGTCTGCAGTAGCTGGGGCTGAG
Dolphin 365 CAAAGGTCCCTTTTCAGTCCAAGGTCACAGAGACTCGAGTCTGCAGTACATGAGGCTGGG
KWhale 367 CAAAGGTCACTTTTCAGTCCAAGGTCACAGAGACTCGAGTCTGCAGTACATGAGGCTGGG
MWhale 366 TAAAGATCCCTTTTCAGTCCAAGGTCACGGAGACTCGAGTCTGCAGTAGATGAGGCTGGG
Pig 361 AAAGTGTCACTTTTTAGTCCAAGATCACGGGGACTC-AGTCTGCAGTAGATGAGACCGGG
Camel 367 ACAAGGTCACTGACCAGTTCAAGGTCACAGGGACTCAAGTCTGCAGGAGATGAGGCTGGG
Alpaca 366 ACAAGGTCACTGACCAGTTCAAGGTCACAGGGACTCAAGTCTGCAGGAGATGAGGCTGGG
* * * ** * * * ** * *** * ** * * *


Panda 406 GACTGGCACTCCAGATGCTGCAGAAACAGTGACTTAGGGCACTTGCCCAATAGCCTCCAA
PBear 411 GACTGGCACTCCAGGTGCTGCAGAAACACTGACTTAGGGCACTTGCCCAATAGCCTCCAA
CFSeal 409 GACTGGCATTCCAGATGCTGCGGAAACACTGACTTAGGGCACTTGCCCAGGAGCCTCCAA
Walrus 408 GACTGGCATTCCAGATGCTGCAGAAACACTGACTTAGGGCACTTGCCCAATAGCCTCCAA
WSeal 413 GACTGGCATTCCAGATGCTGCAGAAACACTGACTTAGGGCACTTGCCCAATAGCCTCCAA
Dog 409 -ACTAGCAATCCAGATGCTGCAGAAACA--GACTTAGGGCACTTGCCCAACAGCCTCCAA
Cat 421 GACTGGCATTCCAGATGCTGCAAGAAGCACTGACTTAGGGCACTGGCCCAATAGCCTTCAA
Human 372 GACTGGCATTTCAGCTGAGGCAGGAATACTGACTTAGGGGACATTCTCAGTAGCTCCCAG
Rhinoceros 441 GACTGGCATTCCAGGTGTTGCAGAAGCACTGACTC-GGGGGTTTGCTCAATAGCCCTCAG
Cow 426 -ACT-ACAATCCAGACGCTAGAGAAGCACCATCTTAGGGGACTTGCCCGAAGGCCTTGAA
Yak 429 -ACT-ACAATCCAGACGCTAGAGAAGCACCATCTTAGGGGACTTGCCCAAAGGCCTTGAA
WBuffalo 418 -ACT-ATAATCCAAATGCTAGAGAAGCACCATCTTAGGGGACTTGCCCAAAGGCCTTGAA
Goat 429 -ACT-ATAGTCCCAATGCTACAAAAGCACCATCTTAGGGGACTTGCCCGAAGGCCTTGAA
Antelope 429 -ACT-ATCGTCCCGATGCTACAAAAGCACCATCTTAGGGGACTTGCCCGAAGGCCTTGAA
Sheep 428 -ACT-ATAGTCCCGATGCTACAAAAGCACCATCTTAGTGGACTTACCCAAAGGCCTTAAA
Dolphin 425 GACT-GCGTTCCGGATGCTACAGAAGCACCGTCTTAGGGGACTTGCTCAATAGCCATGAA
KWhale 427 GACT-GCGTTCCGGATGCTACAGAAGCACCGTCTTAGGGGACTTGCTCAATAGCCATGAA
MWhale 426 GACT-GCATTCCGGATGCTACAGAAGCACCGTCTTAGGGGACTTGCTCAATAGCTATGAA
Pig 420 GATG-GCCTTCCAGATGCTGCGAAAGCACCATCTTCAGGGAGTTGCTCAATAGCCATGAA
Camel 427 AACT-GCATCCCAGATGTGGCAGGAGCACCGTCTTAGGGGACTTGCTCAATAGCCAGCAA
Alpaca 426 AACT-GCATCCCAGATGTGGCAGGAGCACCGTCTTAGGGGACTTGCTCAATAGCCAGCAA
Panda 466 GAAGCCCTCCAAGGCAAG-TGGCAGTGCCTCCAGGT-AGACATCGAATTCCTCTGGG-CC
PBear 471 GAAGCCCTCCAAGGCAAG-TGGCAGTGCCTCCAGGT-AGACATCGAATTCCTCTAGG-CC
CFSeal 469 GAAGCCCTCTAACGCAAG-CGGCCGTGCCTCCAGGT-AGACATCCACTTCCTCTGGG-CC
Walrus 468 GAAGCCCTCCAACGCAAG-CGGCCGTGCCTCCAGGT-AGACATCCACTTCCTCTGGG-CC
WSeal 473 GAAGCCCTCCAACGCAAG-CGGTCGTGCCTCCAGGT-AGACATTCACTTCCTCTGGG-AC
Dog 466 GAAGCCCTCCAAGGCAAG-TGGCACTGCCTCCAGGT-GGACATCCAATTCCTTTGTG-AC
Cat 481 GAAGTCCTCCCAGGCAAG-TGGCACTGCCTCTGGGTGAGACCTCCAATTCTTCTGGG-CC
Human 432 GAGCCCCTCCAAGGTGTG-TAGCACCACCTCCAGACAAGCACTCCACTTCATCCCGG-CT
Rhinoceros 500 AA-CCTCTCCAAGGTGGG-TGGCACTGCCTCCAGGTAAGAACTCCGGTTCCTCTGCG-CC
Cow 484 GAACTGATCCAAGGCCAACTGGCAGGGCCTCCAG----------TTCTTGCAG--CC
Yak 487 GAACTCATCCAAGGCCAATTGGCAGGGCCTCCAG----------TTCTTGCAG--CC
WBuffalo 476 GAACTCACCCAAGGCCAATTGGCATGGCCTCCAG----------TTCTTGCAG--CC
Goat 487 GAACCCATCCAAGGCCAATTGGCATGGCTTCCAG----------TTCTTGCAG--CC
Antelope 487 GAACCCATCCAAAGCCAATTGGCATGGCCTCCAG----------TTCTTGCAG--CC
Sheep 486 GAACCCATCCAAGGCCAATTGGCATGGCCTCCAG----------TTCTTGCAG--CC
Dolphin 484 GAGCCCATCCAAGGCCAG-TGACACAGCCTCCAGGTAAGAAGCTCCAGTTCTTCCAG--CC
KWhale 486 GAGCCCATCCAAGGCCAG-TGGCACAGCCTCCAGGTAAGAAGCTCCAGTTCTTCCAG--CC
MWhale 485 GAGCCCATCCAAGGCCAG-TG-CACAGCCTCCAGGTAAGAAGCTCCAGTTCTTCCAG--CC
Pig 479 GAACCCATCCAAGGCCAG-CAGCACTGCCTCCAGCGAGGAGATCCAGTTCCCCCAG--GC
Camel 486 GACCCCCTCCAAGGCCGG-TGGCACTGCCTCCAGGTAAGAACGCCACCGTCCTCCGGGGCC
Alpaca 485 GACCCCCTCCAAGGCCGG-TGGCACTGCCTCCAGGTCAGAACGCCACGCCCTCCGGGGCC
Panda 523 CACATTCCTTGCCCAGGGTGCTTCCGCTTCCAAGTAATGAAAAGCCAGATACTTCCCTCA
PBear 528 CACATTCCTTGCCCAGGGTGCTTCCTCTTCCAAGTAATGAAAAGCCAGATACTTCCCTCA
CFSeal 526 CACACTCCTTGCCCAGGGTGCTTCCTTTCCAGGTAATGAAAAGCCAGATACTTCCCTCA
Walrus 525 CACACTCCTTGCCCAGGGTGCTTCCTTTCCAGGTAATGAAAAGCCAGATACTTCCCTCA
WSeal 530 CACACTCCTTGCCCAGGGTGCTTCCTCTTCCAGGTAATGAAAAGCCAGATACTTCCCTCC
Dog 523 CACACTCCTTGCCCAAGGGGCTTCCTCTTCCAAGTAATGAAAAGCCAGATGCTTCCCTCA
Cat 539 CACACTCCTTGTCGAGGGTACTTCCTCTTCCAAGTAATGAAAACCCAGATATTTCCCACA
Human 490 CATACCTCTTGCCCAGAGTGCTTCCTCTTCCAAGAAATCAAAAACAAGATATTTCCTTTG
Rhinoceros 557 TGCACCCCTTGCCTAGGGTGCTTCCTCTTCCAAGTAATGAAAAGCAAGACACTTCCCTCA
Cow 529 CACACCCCTGGCCCAAGGTGCTTCCTATTCCAAGTAACGAGAATCAAGATACTTCCTTCA
Yak 532 CACACCCCTGGCCCAAGGTGCTTCCTATTCCAAGTAACGAGAATCAGGATACTTCCTTCA
WBuffalo 521 CACACCCCTGGCCCAAGGTGCTTCCTATTCCAAGTAATGAGAATCAGGATACTTCCTTCA
Goat 532 CACACTCCTGGCCCAAGGTGCTTCCTATTCCAAGTAATGAGAATCAAGATACTTCCCTCA
Antelope 532 CACATTCCTGGCCCAAGGTGCTTCCTATTCCAAGTAATGAGAATCAAGATACTTCCCTCA
Sheep 531 CACACTCCTGGCCCAAGGTGCTTCCTATTCCAAGTAATGAGAATCAAGATACTTCCCTCA
Dolphin 541 CACACCCCTTGCCCAGGGTGCTTCCTATTCCAAGTAATGGAAAGCGAG-TACTTCCCTCA
KWhale 543 CACACCCCTTGCCCAGGGTGCTTCCTATTCCAAGTAATGGAAAGCGAG-TACTTCCCTCA
MWhale 541 CACACCCCTTGCCCAGGGTGCTTCCTATTCCAAGTAATGAAAAGCGAG-TACTTCCCTCA
Pig 536 C-CACGCCTGGCCCAGGGTGCTTCCTATTCCACGTAAGGAAAGGCAAG-TACTTCCCTCA
Camel 545 CACATTCCTGGCCCAGGGTGCTTCCTATTCCAAGTAATGAGGAGTGAGATACTTCCCTCA
Alpaca 544 CACATTCCTGGCCCAGGGTGCTTCCTATTCCAAGTAATGAGGAGTGAGATACTTCCCTCA
Panda 583 ACTCAGAGG-AATAGGG-CCCTTCTGCTTC-AGCTGCAGGTATA-ACTTTAGACTCCCAG
PBear 588 GCTCAGAGG-AATAGGG-CCCTTCTGCTTC-AGCTGCAGGAAGA-ACTTCAGACTCCCAG
CFSeal 586 ACTCAGAGG-AATAGGG-CCCTTCTGCTTC-AGCTGCAGGTATC-ACTTTAGCCTCC-AG
Walrus 585 ACTCAGAGG-AATAGGG-CCCTTCTGCTTC-AGCTGCAGGTATC-ACTTTAGCCTCC-AG
WSeal 590 ACTCAGAGG-AATAGGG-CCCTTCTGCTTC-AGCTGCAGGTATC-ACTTTAGCCTCC-AG
Dog 583 ACTCAGAGG-AATAGGG-CTCTTCAGCTTC-AGCTGCAGGTA---ACTTTAGTCCCCCAG
Cat 599 ACTCTGAGG-AAGAGGG-CACACCTGCTTTTAGCTGCAGGTATAAATTTTAGTCTCCCAG
Human 550 TCCTGTGGA-AACAGGGACTCTTCTTTTTC-TGC---AGGTATAAGTTCCAGCCTCCCAG
Rhinoceros 617 ACCCAAGGG-AATAGGGCCCCTTCTCCTTC-GGCTGGAGGTATAAGTTCCAGCCTCCCCG
Cow 589 GCCCTGGGG-AATAGGACCCCTTCTCCCTC-AGGTGCAGCTGTAAGTTCCAGCCTCCCGA
Yak 592 GCCCTGGGG-AATAGGACCCCTTCTCCCTC-AGGTGCAGCTGTAAGTTCCAGCCTCCCGA
WBuffalo 581 GCCCTGGGG-AATAGGACCCCTTCTCCCTC-AGATGCAGCTTTAAGTTCCAGCCTCCCGG
Goat 592 GCTCTGGGG-AATAGGACCCCTTCTCCTTC-AGCTGCAGCTATAAGTTCCAGCCTCCCGG
Antelope 592 GCTCTGGGG-AATAGGACCCCTTCTCCTTC-AGCTGCAGCTATAAGTTCCAGCCTCCCGG
Sheep 591 GCTCTAGGG-AATAGGACCCCTTCTCCTTC-AGCTGCAGCTATAAGTTCCAGCCTCCCAG
Dolphin 600 GCCCTGGGG-AAGAGGACCCCTTCTCCTTC-AGCTGCAGATATAAGTTCCGCCTCCTAG
KWhale 602 GCCCTGGGG-AAGAGGACCCCTTCTCCTTC-AGCTGCAGATATAAGTTCCTGCCTCCTAG
MWhale 600 GCCCTGGGG-AAGAGGACCCCTTCTCCTTC-AGCTGCAGATATAAGTTCCCGCCTCCCAG
Pig 594 AGCCCAGGG-AATAGAAGCTCTTCTCTTTC-CACGGCAGGTATAAGTTCCAGCTTCCCAG
Camel 605 GCCCTGGGGGAAGAGGGCCCCTTGTCCTTC-AACCGCTGGTATAAGTTCCAGCCTCCCAG
Alpaca 604 GCCCTGGGGGAAGAGGGCCCCTTGTCCTTC-AGCTGCTGGTATAAGTTCCAGCCTCCCAG



Supplementary file 14. Alignment of eutherian *CTI* proximal promoters
ClustalW2 alignment of the eutherian *CTI* proximal promoters. Nucleotides identical in all 21 sequences are indicated by an asterisk (*), those shared by at least 70% of sequences (15 out of 21) are shaded black and conserved nucleotide types (either purines or pyrimidines) shared by more than 70% of sequences are shaded grey. The putative translation start site is shaded green. The human and Weddell seal *CTI* promoter regions from the respective pseudogenes are included for comparison.



Supplementary file 15. Conserved transcription factor binding sites within 6 marsupial *ELP* and 20 eutherian *CTI* proximal promoters

Eight conserved regions (i-viii, red bars) within the *ELP* and *CTI* proximal promoter were identified (see also Fig. 5). Three highly-conserved regions (shaded grey) included: (iii) overlapping NF-κB and C/EBP sites, (vi) AP2γ and (vii) a MAF/ETS site which is partially overlapped by an ERRα/β-like motif. Conserved putative TFBS are boxed. Asterisks indicate nucleotides common to all sequences and alignment numbering is shown above sequences.

Supplementary file 16. Conserved *ELP* and *CTI*-specific putative TFBS identified within the proximal promoter region (Fig. 5).

Conserved TFBS motifs from 26 *ELP* and *CTI* sequences were identified by a Tomtom (Motif Comparison Tool Search) of public databases, with the identity and probabilities associated with selected database hits listed below.

Putative TFBS	Alignment numbering (Fig. 5)	Sequence (5'-3')	Database Hits	p-value
NF-κB(+) Consensus: GGGRCTTGCYC	85-95	GGCACTTGCCC GGCACTTGCCC GGCACTTGCCC GGCACTTGCCC GGCACTTGCCC GGCACTGGCTC GGGACTTGCCC GGGACTTGCCC GGGACTTGCCC GGGACTTGCCC GGGACTTGCCC TGGACTTACCC GGGACTTGCTC GGGACTTGCTC GGGACTTGCTC GGGAGTTGCTC GGGACTTGCTC GGGACTTGCTC GGGGTTTGCTC GGGACATTCTC GGGGTTTGTCC AGGGTTTGCCC GGGGCTTGTCC GGGGCTTGTCC GGGGTTTGCCC GAGGTTTGCCC	**MA0105.3 (NFKB1)** **MA0107.1 (RELA)** Rel homology region (RHR) factors	8.31e-05 5.00e-04
C/EBP(+) Consensus: CTTGCYCAAW	87-98	CTTGCCCAGG CTTGCCCAAT CTTGCCCAAT CTTGCCCAAC CTGGCTCAAT CTTGCCCGAA CTTGCCCAAA CTTGCCCAAA CTTGCCCGAA CTTGCCCGAA CTTACCCAAA CTTGCTCAAT CTTGCTCAAT CTTGCTCAAT GTTGCTCAAT CTTGCTCAAT CTTGCTCAAT TTTGCTCAAT CATTCTCAGT TTTGTCCAAT TTTGCCCAAT CTTGTCCAAA CTTGTCCAAA TTTGCCCAAA TTTGCCCAAT	**GBX2_DBD** Gastrulation and brain-specific homeobox protein 2 **Cebpb_DBD** CCAAT/enhancer-binding protein beta, Cebpb, CEBPB_MOUSE, IL-6DBP, Interleukin-6-dependent-binding protein, LAP, Liver-enriched transcriptional activatorF 1.0)	1.50e-03 3.65e-03

Putative TFBS	Alignment numbering (Fig. 5)	Sequence (5'-3')	Top hits	p-value
AP2γ(+) Consensus: KGCCYARGG	191-199	GGCACTTGCCC GGCACTTGCCC GGCACTTGCCC GGCACTTGCCC GGCACTTGCCC GGCACTGGCTC GGGACTTGCCC GGGACTTGCCC GGGACTTGCCC GGGACTTGCCC GGGACTTGCCC TGGACTTACCC GGGACTTGCTC GGGACTTGCTC GGGACTTGCTC GGGAGTTGCTC GGGACTTGCTC GGGACTTGCTC GGGGTTTGCTC GGGACATTCTC GGGGTTTGTCC AGGGTTTGCCC GGGGCTTGTCC GGGGCTTGTCC GGGGTTTGCCC GAGGTTTGCCC	**MA0154.2 (EBF1)** Early B-Cell Factor 1 **TFAP2C_full** Activating enhancer binding protein 2 gamma, B2RAN6_HUMAN, cDNA, FLJ95025, highly similar to Homo sapiens transcription factor AP-2 gamma, TFAP2C	2.70e-04 4.88e-04
MAF/ETS(-) Consensus: CTTCCT	217-222	CTTGCCCAGG CTTGCCCAAT CTTGCCCAAT CTTGCCCAAT CTTGCCCAAC CTGGCTCAAT CTTGCCCGAA CTTGCCCAAA CTTGCCCAAA CTTGCCCGAA CTTGCCCGAA CTTACCCAAA CTTGCTCAAT CTTGCTCAAT CTTGCTCAAT GTTGCTCAAT CTTGCTCAAT CTTGCTCAAT TTTGCTCAAT CATTCTCAGT TTTGTCCAAT TTTGCCCAAT CTTGTCCAAA CTTGTCCAAA TTTGCCCAAA TTTGCCCAAT	**MA0474.1 (Erg) (ETS)** **MA0098.2 (Ets1)**	8.24e-06 5.22e-05

Chapter Seven:

Developing recombinant protein expression of tammar ELP

Introduction

Marsupial *ELP* and eutherian *CTI* encode 8.47-9.68 kDa (Pharo *et al.*, 2016) milk proteins with a single Kunitz protease inhibitor domain located at the C-terminus of the protein (Pharo *et al.*, 2012). However, attached carbohydrate moieties ranging from 40-60% means that the size of N-glycosylated CTI and ELP ranges from 10-21 kDa (Cechova, 1976; Joss *et al.*, 2007; Joss *et al.*, 2009; Laskowski and Laskowski, 1951; Laskowski *et al.*, 1952; Pineiro *et al.*, 1975; Tschesche *et al.*, 1975). Like BPTI (Huber *et al.*, 1974), Cape fur seal CTI is predicted to bind to and inhibit trypsin by the reversible formation of a 1:1 complex (CTI : trypsin) via the P_1 reactive site residue (Pharo *et al.*, 2016). Although the P_1 reactive site residue is conserved in eutherian CTI, i.e. either a basic lysine or arginine residue with a positively charged side chain, this is not so for marsupial ELP. The marsupial proteins have a range of P_1 residues ranging from: Asp (acidic with a negatively charged side-chain), Ser, or Asn (polar uncharged side chain) (Pharo *et al.*, 2016). The very small volumes of marsupial milk secreted during early lactation (Phase 2A), has thus far hindered the isolation of a sufficient quantity of the purified protein for functional studies and the production of an antibody to marsupial ELP (Piotte and Grigor, 1996; Simpson *et al.*, 1998). Therefore, the development of a recombinant protein expression system would enable the large-scale production and characterization of tammar ELP (tELP).

There are a number of different methods available to produce recombinant proteins including bacterial systems: e.g. *Escherichia coli*, *Pseudomonas fluorescens* and eukaryotic systems: unicellular microalgae, yeast, insect, plant and mammalian expression systems (see reviews) (Baneyx and Mujacic, 2004; Boer *et al.*, 2007; Celik and Calik, 2012; Demain and Vaishnav, 2009; Gräslund S, 2008; Ko *et al.*, 2008; Overton, 2014; Potvin and Zhang, 2010; van Oers *et al.*, 2015; Xiao *et al.*, 2014; Zhu, 2012). However, each of these methods has advantages and disadvantages. Bacterial expression systems are readily used for recombinant protein production as they provide the cheapest and quickest method for large-scale recombinant protein production (Demain and Vaishnav, 2009). Yet, *E. coli* is not adept at disulphide bond formation, often producing insoluble, misfolded, biologically inactive proteins (Baneyx, 1999; Kingston and Brett, 2005; LaVallie and McCoy, 1995; Sorensen and Mortensen, 2005; Wall and Pluckthun, 1995). Glycosylation of bacterial proteins was thought to be rare (Baneyx and Mujacic, 2004; Kingston and Brett, 2005; Wall and Pluckthun, 1995), but recent studies have shown that both pathogenic and commensal microorganisms have active N-linked and O-linked glycosylation pathways (see reviews) (Nothaft and

Szymanski, 2010; Nothaft and Szymanski, 2013).

Mammalian expression systems, while ideal for the production of correctly modified eukaryotic proteins, are expensive and generally produce significantly lower recombinant protein yields (Baneyx and Mujacic, 2004; Chapple *et al.*, 2006; Demain and Vaishnav, 2009; Hartley, 2006; Jana and Deb, 2005; Kingston and Brett, 2005; Swietnicki, 2006; Wurm, 2004). However, they are often essential for the production of a biologically active protein (Bandaranayake and Almo, 2014), particularly for those whose activity is dependent on the correct post-translational modification(s) *in vitro* (Zhu, 2012). Mammalian cell lines commonly used to produce recombinant proteins include CHO (Chinese hamster ovary), HEK293, Per.C6 (from human retinoblastoma cells) and Cap-T (from human amniocyte lines) (Almo and Love, 2014; Bandaranayake and Almo, 2014; Zhu, 2012).

The aims of this work were to evaluate mammalian and bacterial expression systems for the *in vitro* production of recombinant tELP. The pCMVSport6.ccdb and pcDNA3.1 mammalian expression vectors were trialed for tELP production in conditioned medium by HEK293T cells using a transient transfection strategy. The bacterial (*E. coli*) production of recombinant tELP was appraised using the pMAL Protein and Fusion Purification System which produces the protein of interest (tELP, C-terminus) fused to maltose binding protein (MBP, N-terminus).

Materials and Methods

Mammalian expression of marsupial ELP and eutherian CTI

Preparation of ELP and CTI mammalian expression constructs

The pCMVSport6 and pcDNA3.1 vectors were used to express selected marsupial ELP and eutherian CTI proteins. The transcripts to encode these proteins were produced using a PCR based cloning approach or sourced from a normalised tammar mammary gland library (Chapter 2). Clones from a commercially prepared, normalised tammar mammary gland EST library in the pCMVSport6.ccdb expression vector, hereafter referred to as pCMVSport6 (Chapter 2; Invitrogen/Life Technologies, USA; Lefèvre *et al.*, 2007) were screened by *in silico* analysis to identify full-length (precursor protein-encoding) *ELP* and *cystatin C* (*CST3*) clones. Tammar CST3 was used as a positive control for recombinant protein production as it is ubiquitously expressed (Elizabeth Pharo, unpublished data) and like ELP, has disulphide bonds. The pcDNA3.1(+) vector (V790-20, Invitrogen) was also used to express tammar ELP, opossum ELP and bovine

and seal CTI. Constructs were generated by PCR of mammary gland cDNA using gene-specific primers. Complementary DNA was generated from mammary gland total RNA from the following species and reproductive stages: tammar wallaby early lactation (d80L); South American opossum early lactation (d15L); Holstein-Friesian cow, late pregnancy (at ~8 months of pregnancy/1 month prepartum) (Pharo *et al.*, 2012); Cape fur seal, on-shore, lactating (~2 months postpartum) (Pharo *et al.*, 2016). Primers and PCR conditions used to amplify the tammar and opossum *ELP* and bovine *CTI* transcripts (Chapter 3; Pharo *et al*., 2012) and Cape fur seal *CTI* (Chapter 4; Pharo *et al.*, 2016) have been described previously. PCR products were ligated into pGEM-T Easy (Chapter 2), sub-cloned into *Nhe*I and *Bam*HI digested pcDNA3.1 and construct integrity verified by sequencing and restriction enzyme digestions (Chapter 2). The pEGFP-C1 vector which ubiquitously expresses enhanced green fluorescent protein under the control of the cytomegalovirus (CMV) promoter was used as a positive control for transfection.

Human Embryonic Kidney (HEK293T) cell culture

The ELP and CST3 recombinant proteins were produced using HEK293T cells as they stably express the SV40 large T antigen (Simian Vacuolating Virus 40 large Tumour Ag) which can bind to the SV40 enhancer in mammalian expression vectors, e.g. pcDNA3.1, and increase recombinant protein production. 293T cells were cultured in complete media as described previously (Chapter 6).

Production of ELP and CTI in conditioned media

The *ELP* and *CTI* mammalian expression vectors were transiently transfected into 293T cells in order to produce secreted recombinant proteins *in vitro*, i.e. proteins that are transcribed, translated and secreted by 293Ts into culture (conditioned) media. 293Ts were transfected as per the manufacturer's instructions (Chapter 2). Supernatants were collected 48 h post-transfection. Briefly, media was removed, centrifuged at 10,000 *g* for 5 min at 4°C and the supernatant (secreted proteins) collected and stored at -20°C until use. Supernatants were concentrated (~5X) in three centrifugation steps using Nanosep 3K Omega spin columns (OD003C4, Pall Corporation). The spin columns were sterilised with 70% EtOH (500 µL) by spinning at 14,000 *g* for 20 min at 4°C, rinsed with filter sterilized (0.2 µm) MQ H_2O (500 µL) and spun again. Supernatants (500 µL) were applied to the columns and spun a third time to concentrate the samples. The protein contents of both unconcentrated and ~5x concentrated supernatants were determined by the TCA-ponceau S assay (Chapter 2) and confirmed by SDS-PAGE (Chapter 2). Gene transcription was verified by Northern analysis of

293T total RNA and hybridisation with gene-specific cDNA probes (Chapter 2).

Preparation of HEK293T cell lysates

Cell lysates were prepared in order to determine whether the recombinant ELP and CTI proteins were sequestered within 293T cells, i.e. produced as intracellular proteins rather than secreted into the conditioned media (extracellular secreted proteins) as expected. Briefly, adherent 293Ts grown in 6-well plates (35 mm diameter wells) were gently washed twice with ice-cold 1X PBS and lysed with 500 µL ice-cold modified Radioimmunoprecipitation (RIPA) buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1.0% NP-40 (Nonidet P-40), 0.25% sodium deoxycholate, 1 mM EDTA, pH 7.4). Phosphatase inhibitors (1 mM activated sodium orthovanadate and 1 mM sodium fluoride) and a Protease Inhibitor Cocktail (50 µL): [4-(2-]Aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF), aprotinin, bestatin, E-64-[N-(trans-Epoxysuccinyl)-L-leucine 4-guanidinobutylamide], leupeptin and pepstatin A (P8340, Sigma) were added immediately before use. Cells were scraped, transferred to 1.5 mL centrifuge tubes, shaken for 15 min at 4°C and then centrifuged at 14,000 *g* for 15 min at 4°C. The intracellular proteins were then analysed by SDS-PAGE.

Sulforhodamine B colorimetric assay: measurement of cell biomass production

The sulforhodamine B (SRB) assay was used to measure cell growth based upon biomass/density (Monks *et al.*, 1991; Rubinstein *et al.*, 1990; Skehan *et al.*, 1990; Voigt, 2005). The assay incorporates *in situ* fixation of cells with the anionic protein-binding SRB dye (S1402, Sigma) and so measures basic amino acids only. Most importantly, the assay has a linear relationship with both cell number and cellular protein content as measured by both the Bradford and Lowry assays and also works at a wide range of cell densities, from sparse to multilayered, supraconfluent cells (Monks *et al.*, 1991; Rubinstein *et al.*, 1990; Skehan *et al.*, 1990; Voigt, 2005). Briefly, cells were fixed to 96-well tissue culture plates by gently adding 50 µL of cold 50% TCA (trichloroacetic acid) on top of the growth media to give a final TCA concentration of 10%. Plates were incubated at 4°C for 1 h, rinsed 5x with dd H_2O to remove TCA, growth media, low molecular weight metabolites and serum. Plates were inverted and then air-dried. Fixed cells were stained with 0.4% (w/v) SRB dissolved in 1% acetic acid (100 µL) on a shaking platform for 30 min. Unbound dye was removed by four rinses with 1% acetic acid which was quickly poured over the plates to prevent desorption of the dye. Excess solution was removed by flicking the plates. Plates were inverted and air-dried. Bound SRB dye was solubilised in 100 µL of 10 mM unbuffered tris(hydroxymethyl)aminomethane (Tris, pH~10.5) for 5 min on a shaking platform.

The optical density of the solubilised cell proteins was measured at $A_{540\ nm}$ using a microtitre plate reader (Multiskan Ex, Thermo Electron Corporation, Vantaa, Finland).

Bacterial expression of tammar ELP

Preparation of bacterial expression constructs to produce recombinant tammar ELP

Recombinant tELP was produced with the pMAL Protein and Fusion Purification System (E800S, New England Biolabs; Figure 7.1) (di Guan *et al.*, 1988; Maina *et al.*, 1988). This system was chosen as the recombinant protein of interest (tELP, C-terminus) is fused to maltose binding protein (MBP, N-terminus). The expression of a recombinant protein attached to a fusion partner may improve both protein solubility and yield (di Guan *et al.*, 1988; Maina *et al.*, 1988). Furthermore, the strong "tac" promoter and the MBP (*malE*) gene translation initiation signals characteristic to the pMAL vectors produce high levels of the recombinant fusion protein. The pMAL system also provides the option of directing fusion protein expression to either the *E.coli* cytoplasm or periplasm, as well as the ability to optimise each step of production to maximise yield. Moreover, the fusion protein can be affinity purified on an amylose resin column and the recombinant protein the cleaved from MBP and purified (di Guan *et al.*, 1988; Maina *et al.*, 1988).

The pMAL-c2X and pMAL-p2X bacterial expression vectors were evaluated for the production of MBP-tELP and MBP-tCST3. The pMAL-c2X vector directs fusion protein production to the *E. coli* cytoplasm and pMAL-p2X, to the periplasm (di Guan *et al.*, 1988; Maina *et al.*, 1988). Cystatin C was chosen as a control protein, as like ELP, it contains disulphide bonds which can be problematic to reproduce in *E. coli* bacterial cells (Baneyx, 1999).

Tammar ELP and CST3 were PCR-amplified by PCR (high-fidelity, proof-reading Platinum *Taq*) from the pcDNA3.1-tELP and pCMVSport6-tCST3 (SGT20o3_C05) vectors respectively using primers which incorporated *Eco*RV and *Bam*HI restriction sites at the 5' and 3' ends of the construct respectively (Table 7.1). The region encoding the ELP/CST3 mature protein, stop codon and polyadenylation signal were amplified by PCR (Figure 7.2). PCR products were cloned into pGEM-T Easy, sequenced (Chapter 2), double-digested from pGEM-T Easy with *Eco*RV and *Bam*HI (blunt and sticky ends respectively) and ligated into the *Xmn*I and *Bam*HI-digested (compatible ends) pMAL-c2X and pMAL-p2X vectors.

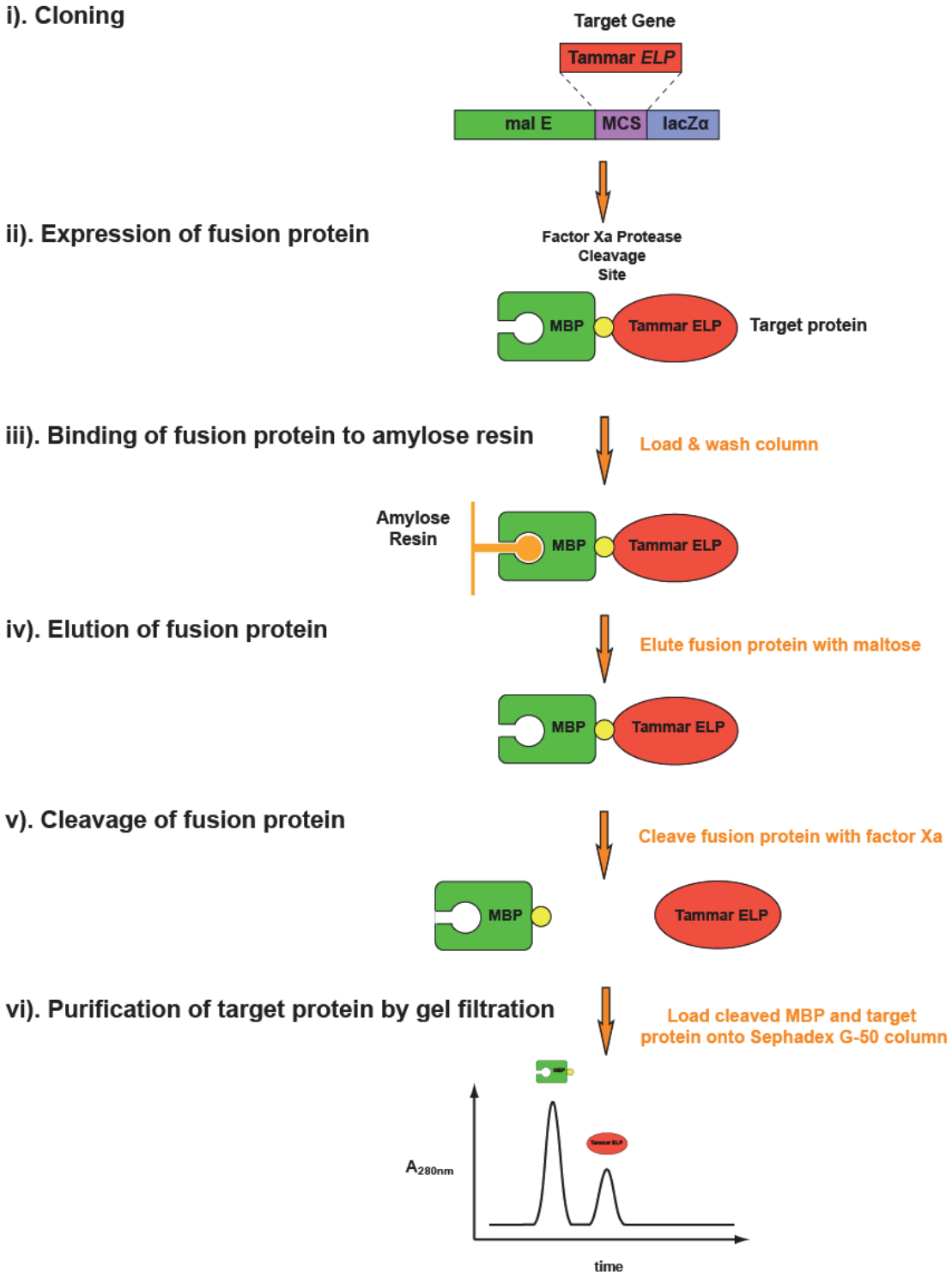


Figure 7.1. Production and purification of recombinant tammar ELP using the pMAL Protein Fusion and Purification System

The target gene (e.g. tammar *ELP*) is cloned into the multiple cloning site (MCS) of the pMAL-c2X, or pMAL-p2X vectors and expressed in *E. coli* to produce recombinant tammar ELP (tELP) fused to maltose binding protein (MBP). The MBP-tELP fusion protein is purified from *E. coli* proteins on an amylose resin column and the fusion protein eluted with maltose. Factor Xa is used to cleave the fusion protein, and then MBP and tELP separated based upon their size, by gel filtration. Adapted from the pMAL manual (E8000S, NEB).

Table 7.1 Primers and conditions used to amplify the mature protein coding region and 3' UTR of tammar *ELP* and *CST3* for expression in bacterial cells

Transcript	Primer F *EcoR*V* (5' ..... 3')	Primer F *BamH*I* (5' ..... 3')	PCR Product Size (bp)	PCR Conditions
Tammar *ELP*	GGCATGACC*GATATC*GAAAAATGTTTAGAC	*GGATCC*TGTTAATGCTTTATTGTACCAG	405	94°C for 2 min; 35 cycles of [94°C for 30 s; 56°C for 30 s; 68°C for 45s]; 68°C for 10 min.
Tammar *CST3*	CTTCGCCGTC*GATATC*GAGCGCCTG	GA*GGATCC*TTTATTTCCTTCTCCTGCCAGC	543	94°C for 2 min; 35 cycles of [94°C for 30 s; 56°C for 30 s; 68°C for 45s]; 68°C for 10 min.

*Restriction enzyme sites (underlined) incorporated into the forward and reverse primers respectively for cloning into the pMAL-c2X and pMAL-p2X expression vectors


A
polylinker
lacZα
malE
rrnB terminator
P tac
pMAL-c2X
pMAL-p2X
Amp r
lacI q
M13 ori
pBR322 ori
B
pMAL-c2X, pMAL-p2X Polylinker
Xmn I (blunt-end)
EcoR I
BamH I
Xba I
Sal I
Pst I
Hind III
malE...ATC GAG GGA AGG ATT TCA GAA TTC GGA TCC TCT AGA GTC GAC CTG CAG GCA AGC TTG...lacZα
Ile Glu Gly Arg
Factor Xa cleavage site
C
Recombinant ELP
Factor Xa cleavage site
MBP-polylinker
Recombinant tammar ELP
Maltose Binding Protein I E G R I E K C L D . . .
Tammar ELP EcoRV Forward primer
EcoR V (blunt-end)
5'-ggcatgaccgatatcgaaaaatgtttagac-3'
Mature endogenous tammar ELP
...cagcctggctggcatgaccagctca
cDNA
gaaaaatgtttagaccaaatccaagtgaactccctggagaacctatctctgttggttcca
E K C L D Q I Q V N S L E N L S L L V P
tctttatgtctgctcccacctgtgaggggtaactgcagttctcaaattcttcactatttc
S L C L L P P V R G N C S S Q I L H Y F
tacaacacaacctctcgaacctgtgagactttcatctacagtggctgcaatggcaataagg
Y N T T S R T C E T F I Y S G C N G N R
aacaattttaactcggaggaatactgcttaaagacctgcagacgcaacaagaacaggaac
N N F N S E E Y C L K T C R R N K N R N
aataataactgaagctgatgttgaccattctgggagcaaccacatcacttccttgaccag
N N N -
tcctactgccatccaaactctggtgatgctgtcctgtttctttccttctatgattccctc
ttctggtacaataaa
BamH I
5'-ctggtacaataaagcattaacaggatcc-3'
Tammar ELP BamHI Reverse primer (reverse complement)
D
Recombinant Cystatin C
Factor Xa cleavage site
MBP-polylinker
Recombinant tammar cystatin C
Maltose Binding Protein I E G R I E R L P R . . .
Tammar CST3 EcoRV Forward primer
EcoR V (blunt-end)
5'-cttcgccgtcgatatcgagcgcctg-3'
Mature endogenous tammar cystatin C
...ggccctggccttcgccgtccaagcc
cDNA
gagcgcctgccccgcctcctgggcggcctcatggaggccaacgagaacgaggagggcgtg
E R L P R L L G G L M E A N E N E E G V
caaaaagcgctcggcttcgccctgagagaattcaacaaggccagcaacgacatgtaccac
Q K A L G F A L R E F N K A S N D M Y H
agccgagtgtccagggtggtgcgagtcaggaaacagattgtgtctggtgtaaaatactat
S R V S R V V R V R K Q I V S G V K Y Y
attgaagctgagcttggtagaaccacctgcaccaagtctatggctaaccttgaaagctgc
I E A E L G R T T C T K S M A N L E S C
ccctatcatgaagacccaactctgcaaaagcactccgtctgtgaatttgtggtgtatact
P Y H E D P T L Q K H S V C E F V V Y T
gttccttggttgggcaaaaccaccctcttgaagaatgaatgcaaagatgcctaggcatat
V P W L G K T T L L K N E C K D A -
cctggtacacatctgtaaccagtctgagctgcagctgttttctccttttgaaaaaaaga
gcagtaacacaacctatttgattcattattgcttaattttaatgctgtgcaaggtaatga
tgtaatcttctccctatgctggcaggagaaggaataaaggcttgtcaaa
BamH I
Tammar CST3 BamHI Reverse primer (reverse complement)
5'-gctggcaggagaaggaaataaaggatcctc-3'

Figure 7.2. Design of ELP and cystatin C constructs for expression of the MBP-tELP and MBP-tCST3 fusion proteins in the *E. coli* cytoplasm (pMAL-c2X) and periplasm (pMAL-p2X)

A. Vector maps of pMAL-c2X and pMAL-p2X, **B.** Polylinker region. *EcoR*V (blunt-ended) and *Bam*HI-digested *ELP* and *CST3* PCR products were cloned into the *Xmn*I (blunt-ended) and *Bam*HI-digested pMAL-c2X and pMAL-p2X vectors. **C.** PCR amplification and cloning strategy used to create the pMAL-c2-tELP and pMAL-p2-tELP vectors to express recombinant mature tELP. One additional amino acid was incorporated at the tELP N-terminus (Ile) due to the incorporation on an *EcoR*V cut site for cloning purposes. The PCR primers used and the translated cDNA encoding the mature protein are shown. The stop codon is indicated by red text and the polyadenylation signal (aataaa) shown in italicized red text. Rare *E. coli* codons are boxed in blue. **D.** Like ELP, the recombinant mature tammar cystatin C protein was produced using a PCR-based cloning approach and expression of cystatin C in *E.coli* cells using the pMAL-c2-tCST3 and pMAL-p2-tCST3 vectors. As for ELP, an additional amino acid was incorporated at the tELP N-terminus (Ile) due to the incorporation on an *EcoR*V cut site for cloning purposes.

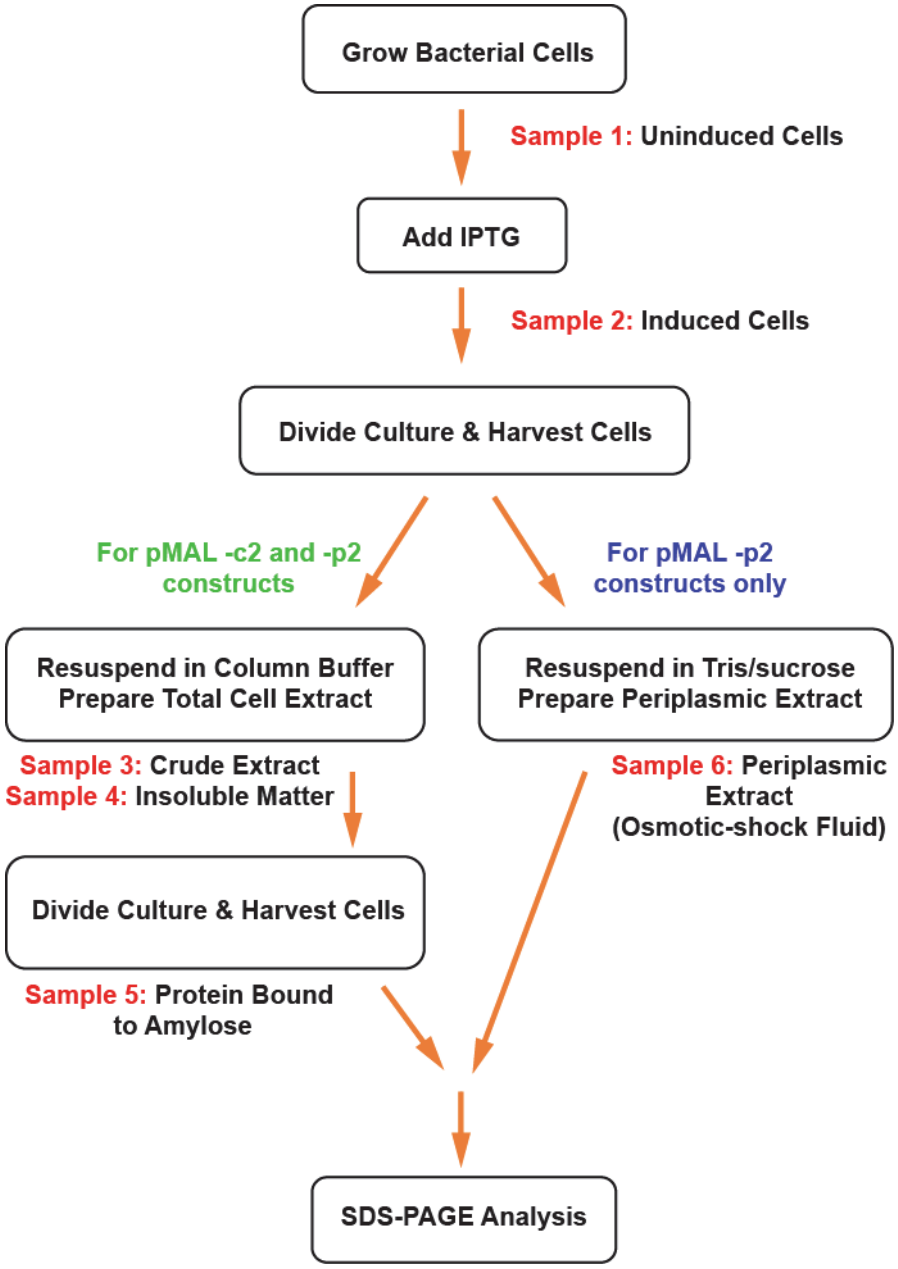


Figure 7.3. Optimization of recombinant fusion protein production with the pMAL Protein Fusion and Purification System

The target gene (e.g. tammar *ELP*) is cloned into the multiple cloning site (MCS) of the pMAL-c2X, or pMAL-p2X vectors and expressed in *E. coli* to produce recombinant tammar ELP (tELP) fused to maltose binding protein (MBP). The MBP-tELP fusion protein is purified from *E. coli* proteins on an amylose resin column and the fusion protein eluted with maltose. Factor Xa is used to cleave the fusion protein, and then MBP and tELP separated based upon their size, by gel filtration. Adapted from the pMAL manual (E8000S, NEB).

Different *E. coli* K12 strains (TB1 and UT5600; E4122S and E4129S respectively, NEB) were evaluated for the production of MBP-tELP and MBP-tCST3. The *E.coli* K12 TB1 strain contains common *E. coli* proteases which may degrade a recombinant protein produced *in vitro*. In contrast, the *E.coli* K12 UT5600 strain lacks the outer membrane protein T (*ompT*) periplasmic protease that can degrade both cytoplasmic and periplasmic recombinant proteins during the bacterial lysis step. Briefly, ligated plasmids (2 µL) were electroporated (Appendix III) into electro-competent *E. coli* K12 TB1 cells (50 µL) or *E. coli* K12 UT5600 cells (50 µL). LB broth (950 µL) was added, the bacterial cells grown for 1 h at 37°C on a shaking platform (300 rpm), and the cells (50 µL or 200 µL) then spread onto LB agar/ampicillin (100 µg/mL) plates and incubated overnight at 37°C. The next day, colonies were picked onto a 'master' LB/ampicillin plate and replicated on to a second LB/ampicillin plate containing 80 µg/mL bromo-chloro-indolyl-galactopyranoside (X-gal) and 0.1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and incubated overnight at 37°C. Successful transformants were identified on the X-gal/IPTG plate (blue/white selection), but single colonies 'picked' from the master plate and inoculated into LB broth (5 mL) containing ampicillin (100 µg/mL). Tubes were incubated overnight (37°C) on a shaking platform (300 rpm), plasmid DNA extracted (Chapter 2) and analysed by restriction enzyme digestion and sequencing with the *malE* primer (5'-GGTCGTCAGACTGTCGATGAAGCC-3', S1237S, NEB).

Optimisation of MBP-tELP fusion protein production

Pilot experiments were undertaken to determine the optimal conditions for the production of recombinant MBP-tELP and MBP-tCST3 (Figure 7.3). Protocols used included those provided by the manufacturer (E800S, New England Biolabs), as well as additional steps sourced from published literature. The effect of induction temperature, IPTG concentration and *E. coli* strain on the viability, solubility and yield of the fusion proteins were investigated as described (Appendix V). These include pilot experiments and the optimisation of fusion protein solubility and Factor Xa cleavage of MBP-tELP.

SDS-PAGE, Western blotting and immunodetection of MBP-tELP

Proteins were separated by SDS-PAGE (Chapter 2) under reducing conditions using a discontinuous buffer system (Laemmli, 1970) and stained with Coomassie (Chapter 2). For Western blotting, only 10% of the sample was applied to the gel. Following electrophoresis, the stacking gel was discarded and the resolving gel and Blot Absorbent Filter Paper (2X 6.5 x 9 cm; 170-3968, Bio-Rad) soaked for 15-30 min in protein transfer buffer [Tris base (0.303% w/v), glycine (1.44% w/v) and methanol (20% v/v),

Appendix II]. The Hybond ECL (enhanced chemiluminesence) Nitrocellulose Membrane (RPN203D, GE Healthcare) was cut (6.5 x 9 cm), pre-wet in MQ H_2O and equilibrated in protein transfer buffer for at 10 min and the gel electrophoretically transferred to the membrane for 15 min at 10V using the Trans-Blot Semi-Dry electrophoretic transfer cell (170-3940, Bio-Rad). The membrane was rinsed briefly in 1X PBS, air-dried, covered in plastic wrap and stored at 4°C until use.

For detection of MBP-tELP or MBP-tCST3, the nitrocellulose membrane was blocked overnight at 4°C, or for 1 h at RT in TBS-T [tris-buffered saline, pH 7.6 (Tris base (0.242% w/v) and NaCl (0.8% w/v, Appendix II) containing Tween 20 (0.1% v/v; P5927, Sigma) and fish gelatin (1% v/v; G7765, Sigma)]. All incubations were carried out on an orbital shaker unless otherwise indicated. The membrane was washed for 3X 5 min periods with TBS-T, and incubated for 1 h with primary antibody (1:10,000 diluted anti-MBP antiserum; E8030S, NEB). Unbound primary antibody was removed by 3x 5 min washes with TBS-T and the membrane incubated for 1 h with secondary antibody (1:10,000 diluted horseradish peroxidase-conjugated goat anti-rabbit IgG antibody, AP132P, Chemicon International). Excess antibody was removed by 3x 5 min washes with TBS-T. Immunodetection was carried out in a darkroom with Amersham ECL Western Blotting Detection Reagents (RPN2109, GE Healthcare), according to the manufacturer' protocol. Briefly, the filter was placed in a zip-lock bag, with the protein side of the membrane uppermost. ECL Detection Reagents 1 (500 µL) and 2 (500 µL) were added to the membrane and incubated for 1 min at RT with gentle mixing. Excess reagent was removed and the blot exposed to Hyperfilm ECL (RPN3103K, GE Healthcare) for exposure periods ranging from 1 to 5 min.

Purification of recombinant tELP from its cleaved MBP fusion-partner by gel filtration

Sephadex G-50 coarse resin (17-0044-02, Amersham Biosciences, GE Healthcare) in a 1.6 cm x 55 cm column (bed volume ~120 mL) was used to separate MBP and tELP. Gel filtration conditions for the separation of MBP (43 kDa) and tELP (~9.58 kDa) were first optimised using ovalbumin (44.3 kDa, A5503, Sigma) and bovine α-lactalbumin (14.17 kDa, L5385, Sigma). The column was equilibrated with degassed buffer [30 mM ammonium bicarbonate (NH_4HCO_3), 10 µM calcium chloride ($CaCl_2$)]. Glucose (10 mg) and 30 mM NH_4HCO_3/10 µM $CaCl_2$ buffer (0.75 mL) were added to the Factor Xa cleaved and concentrated MBP and tELP proteins and applied to the column. Proteins (2 mL fractions) were eluted using a flow rate of 10 mL/h and monitored by A_{280nm} readings.

Purification of insoluble MBP-tELP from E. coli inclusion bodies

Due to the lack of solubility and proteolysis of MBP-tELP in the *E. coli* periplasm, an alternate method to produce the MBP-tELP was used. Biologically active proteins have been successfully produced by 'refolding' an insoluble recombinant protein produced with *E. coli.* This involves the isolation of the recombinant protein from the insoluble *E. coli* protein fraction (inclusion bodies). The protein is then slowly 'refolded' *in vitro* in the presence of oxidising and reducing agents to promote the correct formation of disulphide bonds (Baneyx, 1999; Cabrita *et al.*, 2004; Eiberle and Jungbauer, 2010; Kingston and Brett, 2005).

MBP-tELP was produced in the *E. coli* cytoplasm with the pMAL-c2-tELP vector and insoluble MBP-tELP was isolated from inclusion bodies, solubilised and refolded based on the Protein Refolding Kit (70123-3, Novagen) protocol, but with several modifications. MBP-tELP production was induced at 28°C for 8 h with 0.8 mM IPTG, the bacterial cells harvested by centrifugation at 3220 g for 30 min at 4°C and the supernatant discarded. The pellet was resuspended in a 0.1X culture volume of modified inclusion body (IB) Wash Buffer (20 mM Tris-HCl, pH 7.5, 10 mM EDTA, 1% Triton X-100), with the addition of 1 mM sodium azide and 10 mM β-ME. Lysosyme (100 μg/mL) was added, the cells placed on ice for ~40 min with occasional mixing by swirling and then sonicated as previously described.

E. coli inclusion bodies were collected by centrifugation at 10,000 g for 10 min, the supernatant discarded and the pellet resuspended in modified IB Wash Buffer (0.1X culture vol) and the centrifugation step repeated. The supernatant was discarded and the cell pellet resuspended in modified IB Wash Buffer (0.1X culture vol), in a tube of known weight, and the centrifugation step repeated. The supernatant was discarded, the remaining liquid removed with blotting paper and the tube and cell pellet weighed. The pellet was resuspended in RT 1X IB Solubilisation Buffer [50 mM CAPS (3-(cyclohexylamino)-1-propanesulfonic acid), pH 11.0 (C2632, Sigma)] to a concentration of 10-20 mg/mL and then 0.3% (w/v) N-lauroylsarcosine (L5000, Sigma) and 0.1 M DTT (dithiothreitol, D9779, Sigma) added. The resuspended pellet was gently mixed, incubated for 15 min at RT, centrifuged at 10,000 g for 10 min at RT and the supernatant collected. The solubilized protein was dialysed for 6 h at 4°C against one change of dialysis buffer (20 mM Tris-HCl, pH 8.5, at least 50 vol) containing 0.1 M DTT. Dialysis was continued for an additional 6 h period against one change of dialysis buffer, but with DTT omitted. Protein re-folding (disulphide bond formation) was promoted by dialysis overnight at 4°C in 1X dialysis buffer (at least 25 vol)

containing 1 mM reduced glutathione (G6529, Sigma) and 0.2 mM oxidised glutathione (G4376, Sigma). The fusion protein was then cleaved with factor Xa and analysed by SDS-PAGE.

Results

Evaluation of the pCMVSport6 and pcDNA3.1 mammalian expression vectors for ELP and CTI production

Two pCMVSport6 EST library clones encoding the complete tammar ELP precursor protein and with a Kozak consensus sequence (GCCRCCAUGG, where R is a purine (A/G) and AUG is the translation start site) and a polyadenylation signal (AATAAA), (SGT20p3_C08, GenBank: EX200229 and SGT20e4_A04, GenBank: EX196710) were identified by *in silico* analysis of a 15,001 clone tammar mammary gland EST library (Lefèvre *et al.*, 2007). A full-length tammar *CST3* clone (SGT20o3_C05, GenBank: EX199939) encoding cystatin C, a ubiquitously expressed (Elizabeth Pharo, unpublished data) cysteine protease inhibitor with two disulphide bonds (Turk *et al.*, 2008), as well as a pCMVSport6 vector without a cDNA insert (SGT20i2_G03, negative control) were identified.

Expression vectors were transiently transfected into 293Ts and secreted proteins collected in conditioned media (Opti-MEM I) 48 h post-transfection. However, only the tammar *CST3* clone (SGT20o3_C05) produced a prominent band at its predicted molecular weight (~13.32 kDa) that was visible in both the silver-stained (Figure 7.4) and the less-sensitive Coomassie-stained gel (data not shown). The tammar *ELP* clones were predicted to produce a non-glycosylated protein of ~9.58 kDa and/or an N-glycosylated protein of 10-20.5 kDa (Joss *et al.*, 2007; Joss *et al.*, 2009; Simpson, 1998; Simpson *et al.*, 1998). However, no protein bands of the correct size range were detected visually (SGT20p3_C08 clone shown only, Figure 7.4). Indeed, the tammar *ELP* clones appeared identical to the negative controls (the empty vector and mock- and non-transfected cells, the latter treatment not shown). In contrast, the detection of EGFP by fluorescence microscopy indicated that transfections were successful (data not shown).

Due to the apparent lack of success in expressing tELP using the pCMVSport6-tELP clones, *ELP* was cloned into the pcDNA3.1 mammalian expression vector. Tammar and opossum ELP and bovine and Cape fur seal CTI pcDNA3.1 expression constructs were prepared, expressed in 293Ts and secreted proteins and cell lysates collected. While, tammar cystatin C (SGT20o3_C05 clone) was successfully produced at high levels *in vitro* (Figure 7.5), none of the *ELP* or *CTI* pcDNA3.1 expression constructs produced

obvious protein bands in either the secreted or non-secreted protein fractions (Figure 7.5; seal CTI data not shown). However, Northern analysis of total RNA extracted from transfected 293Ts confirmed that the tammar *ELP* and *CST3* and bovine *CTI* constructs were transcribed (Appendix VI). Preliminary studies of HC11 MECs and Kato-III gastric carcinoma cell line treated with CST3 conditioned medium suggested that cystatin C had no effect on cell growth (data not shown).

The apparent inability to produce recombinant ELP *in vitro* with mammalian expression vectors and the inability to detect the protein (no antibody available) necessitated a new strategy.

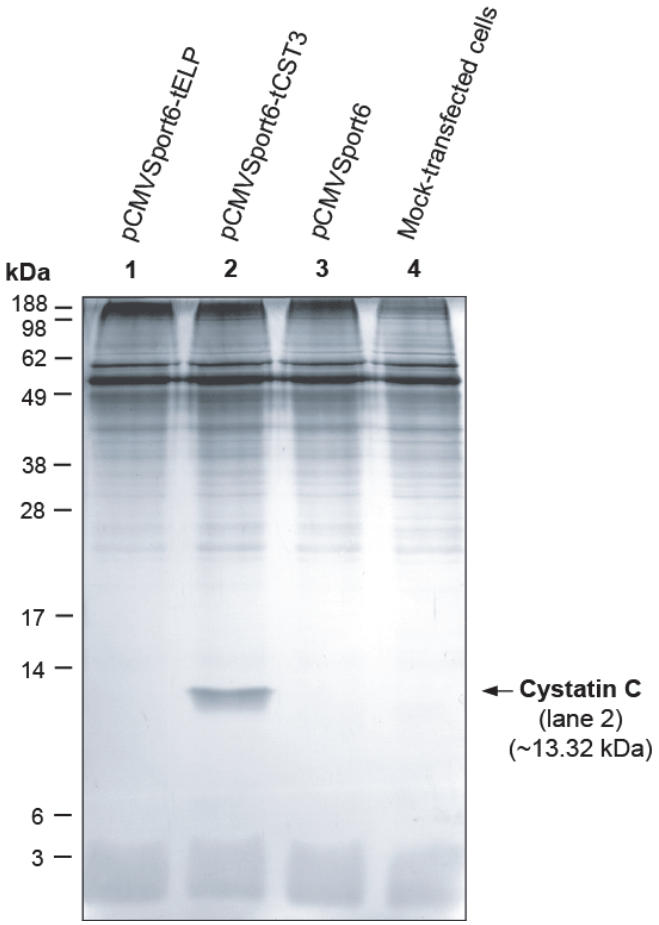
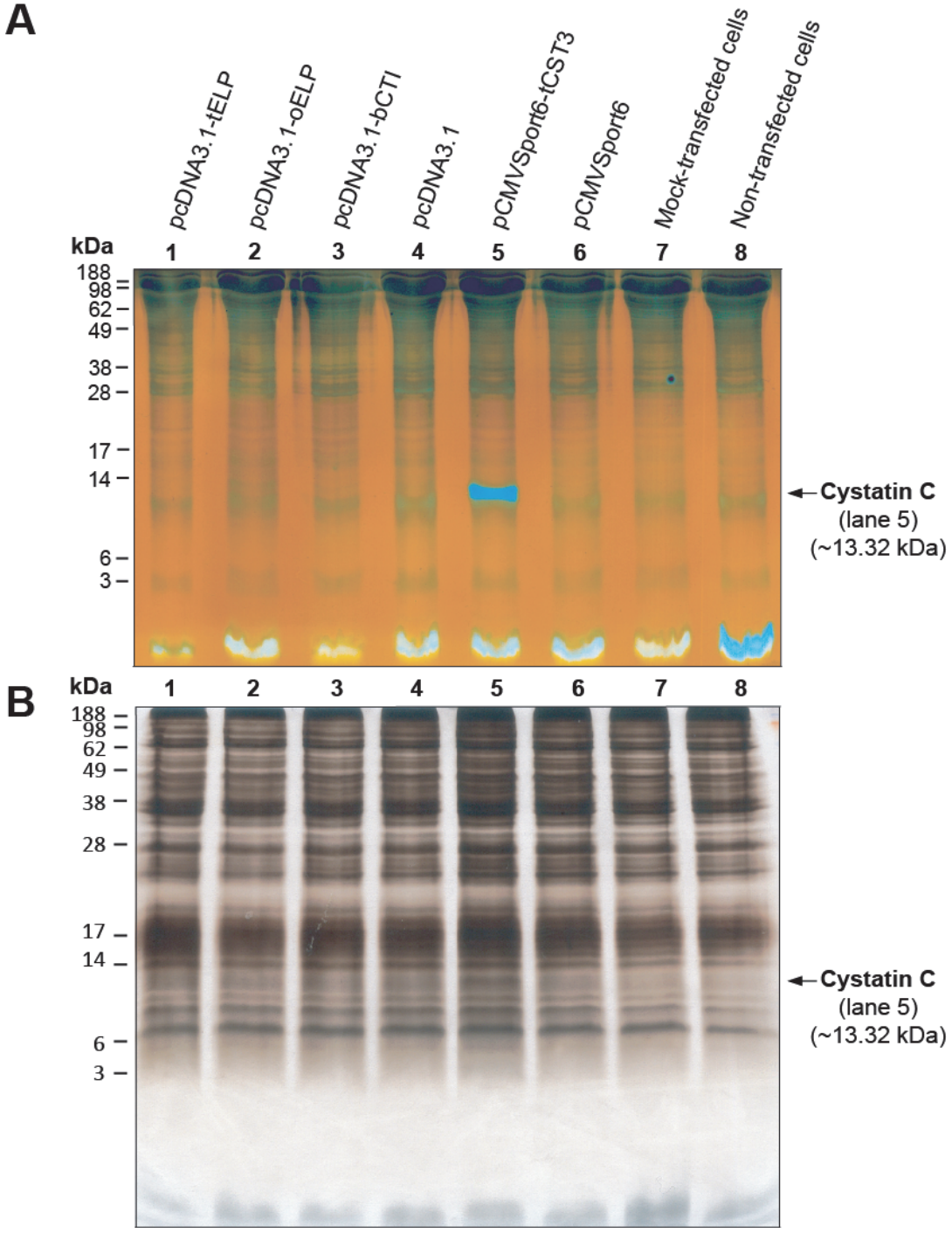


Figure 7.4. Production of tammar ELP and Cystatin C in conditioned media

Mammalian expression constructs were transfected into HEK-293T cells cultured in Opti-MEM I reduced serum medium and supernatants collected 48 h post-transfection. Supernatants (20 µL) were analysed by SDS-PAGE through a 15% gel and the gel silver-stained. Lane 1. tammar ELP, pCMVSport6-tELP, SGT20p3_C08; 2. tammar cystatin C (pCMVSport6-tCST3, SGT20o3_C05; 3. Empty vector, pCMVSport6, SGT20i2_G03 and 4. mock transfected cells. Tammar ELP was not visually detected (Lane 1). In contrast, a protein the expected size of cystatin C (~13.32 kDa) was secreted into the conditioned media (Lane 2). Approximate protein sizes are indicated based on the SeeBlue Plus2 molecular weight marker (Invitrogen), marker not shown.

Figure 7.5. Production of tammar and opossum ELP, bovine CTI and Cystatin C in conditioned media

Expression constructs were transfected into HEK 293T cells cultured in Opti-MEM I reduced serum medium and supernatants collected 48 h post-transfection. **A.** Concentrated (~5X) supernatants (~100 ng, 20 µL) and **B.** Cell lysates (20 µL) were analysed by SDS-PAGE (15% gel) and silver-stained. Lane 1. pcDNA3.1-tELP (tammar ELP), 2. pcDNA3.1-oELP (opossum ELP), 3. pcDNA3.1-bCTI (bovine CTI), 4. pcDNA3.1 (empty vector), 5. pCMVSport6-tCST3 (tammar CST3, SGT20o3_C05), 6. pCMVSport6 (SGT20i2_G03, empty vector), 7. Mock-transfected cells and 8. Non-transfected cells. None of the ELP or CTI expression vectors produced visually detectable secreted or non-secreted proteins. Consistent with Fig 7.3, tammar cystatin C was detected in the supernatant (A, lane 2), with a faint band also present in the cell lysate (B, lane 2). Approximate protein sizes are indicated based on the SeeBlue Plus2 molecular weight marker (Invitrogen), marker not shown.


A
pcDNA3.1-tELP
pcDNA3.1-oELP
pcDNA3.1-bCTI
pcDNA3.1
pCMVSport6-tCST3
pCMVSport6
Mock-transfected cells
Non-transfected cells
kDa
1 2 3 4 5 6 7 8
188
98
62
49
38
28
17
14
6
3
Cystatin C
(lane 5)
(~13.32 kDa)
B
kDa
1 2 3 4 5 6 7 8
188
98
62
49
38
28
17
14
6
3
Cystatin C
(lane 5)
(~13.32 kDa)

Overexpression of recombinant tammar ELP in *E. coli* cells with the pMAL system

The mature protein-coding regions of tammar *ELP* and *CST3* were successfully amplified by PCR (Table 7.1) and cloned into the pMAL-c2X and pMAL-p2X expression vectors to create *mal*E-*ELP* and *mal*E-*CST3* expression constructs. Hence, induction with IPTG produces MBP-ELP and MBP-CST3 respectively (Figure 7.2).

Optimisation MBP-tELP fusion protein production

All facets of ELP and cystatin C production were optimised using pilot experiments. Both tammar ELP (pMAL-c2-tELP) and cystatin C (pMAL-c2-tCST3) were successfully expressed as MBP fusion proteins in the *E. coli* K12 TB1 strain after a 2 h induction with either 0.3 mM, 1 mM or 2 mM IPTG at 37°C and 30°C (data not shown). IPTG (1 mM) was optimal at both 37°C and 30°C, but fusion protein production was greater at a reduced temperature (data not shown). However, MBP-tELP and MBP-tCST3 were predominantly produced as insoluble proteins, with minimal fusion protein in the crude (soluble) fraction. Both fusion proteins were also degraded by *E. coli* cytoplasmic proteases. As both MBP-tELP and MBP-tCST3 were insoluble, the pMAL-p2-tELP and pMAL-p2-tCST3 vectors (periplasmic fusion protein production) were tested using conditions identical to those for cytoplasmic protein production. Although both fusion proteins were expressed in the periplasm, they too were degraded by *E. coli* proteases, particularly MBP-tELP, which was very susceptible to proteolysis, with a loss of ~70% estimated by Western analysis (data not shown).

Expression of MBP-tELP in the *ompT* protease-deficient *E. coli* K12 UT5600 strain – reduced proteolytic degradation

In order to reduce fusion protein degradation, MBP-tELP and MBP-tCST3 were produced using the protease-deficient *E. coli* K12, UT5600. This strategy improved both cytoplasmic and periplasmic production of MBP-tELP and MBP-tCST3 when induced with 1 mM IPTG at 37°C for 2 h (MBP-tELP: Figures 7.6A, 7.6B, and MBP-tCST3: Figures 7.7A, 7.7B). Intact MBP-tELP (~52.2 kDa) was produced in the cytoplasm, but mainly as insoluble inclusion bodies (Insol) and the soluble fusion protein (Sol) was degraded (Figure 7.6A). The periplasmic MBP-tELP (Shock fraction) was degraded, with less than ~20% of the fusion protein intact (Figure 7.6 B). In contrast, there was very little degradation of MBP-tCST3 in the *E. coli* cytoplasm but the majority of MBP-tCST3 was insoluble (Figure 7.7A). Periplasmic production of MBP-tCST3 (Figure 7.7B) was more successful than for MBP-tELP as there was little fusion protein degradation, but most of MBP-tCST3 was insoluble. Unlike MBP-tELP, there was minimal MBP-tCST3 degradation in the *E. coli* Shock fraction.

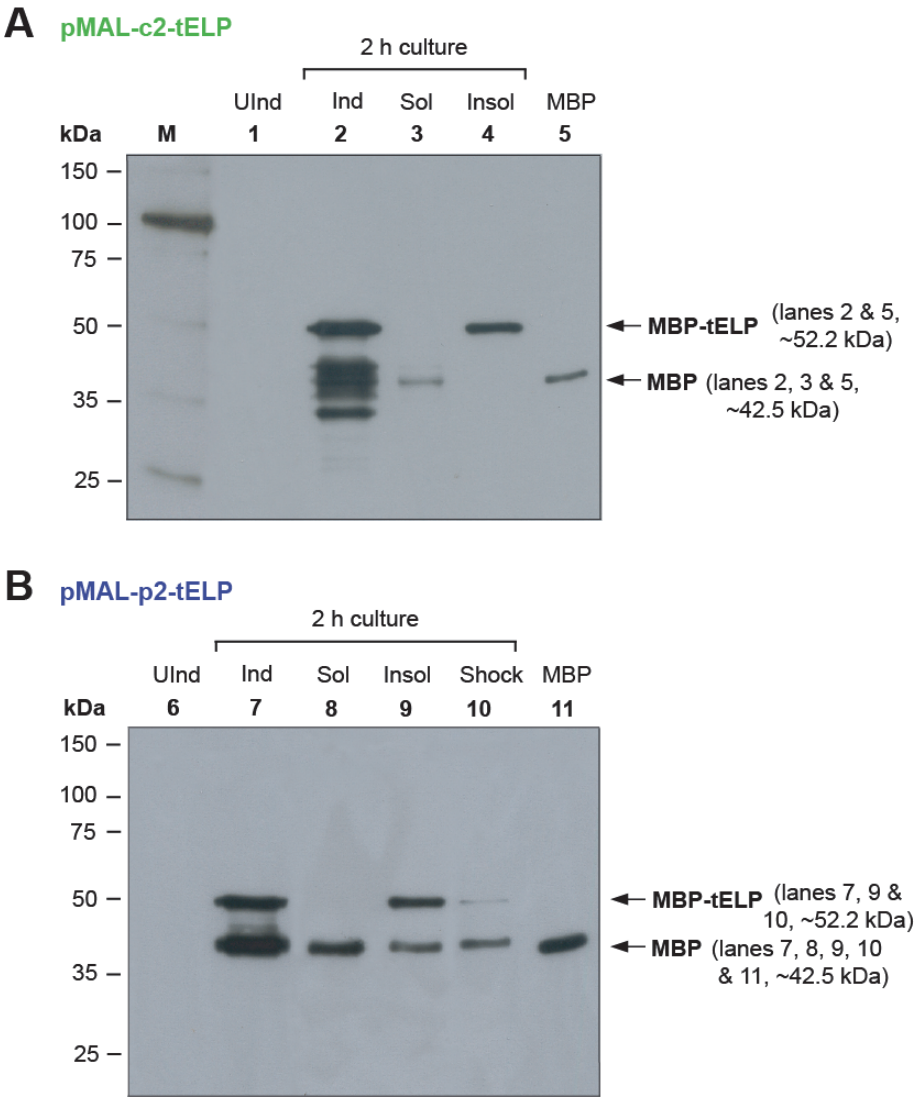


Figure 7.6. Western analysis of MBP-tELP fusion protein production in the *E. coli* K12 UT5600 *ompT* protease-deficient bacterial strain

MBP-tELP fusion protein production was induced with 1 mM IPTG at 37°C for 2 h for **A.** pMALc2-tELP (cytoplasmic protein) and **B.** pMALp2-tELP (periplasmic protein). Samples were analysed by SDS-PAGE (10% gel), transferred to Hybond ECL membrane and MBP-tELP detected with anti-MBP antiserum. **A.** Cytoplasmic MBP-tELP: Lane 1. Uninduced (UInd) bacterial culture at t_0; 2-4: 2 h induction [2. total protein (Ind); 3. soluble (Sol); and 4. insoluble (Insol) protein]; and 5. MBP. **B**. Periplasmic MBP-tELP: 6. Uninduced (UInd) culture at t_0; 7-10: 2 h induction [7. total protein (Ind); 8. soluble (Sol); 9. insoluble (Insol); 10. periplasmic (Shock) protein]; and 11. MBP. MBP-tELP (~52.2 kDa) and MBP (~42.5 kDa) are arrowed. M. Approximate protein sizes are indicated (ECL DualVue Western Blotting Markers, GE Healthcare). Soluble MBP-tELP was degraded in both the cytoplasm (A, lane 3) and the periplasm (B, lane 8). Insoluble MBP-tELP was produced intact in the cytoplasm (A, lane 4), but partly degraded in the periplasm (B, lanes 9, 10).

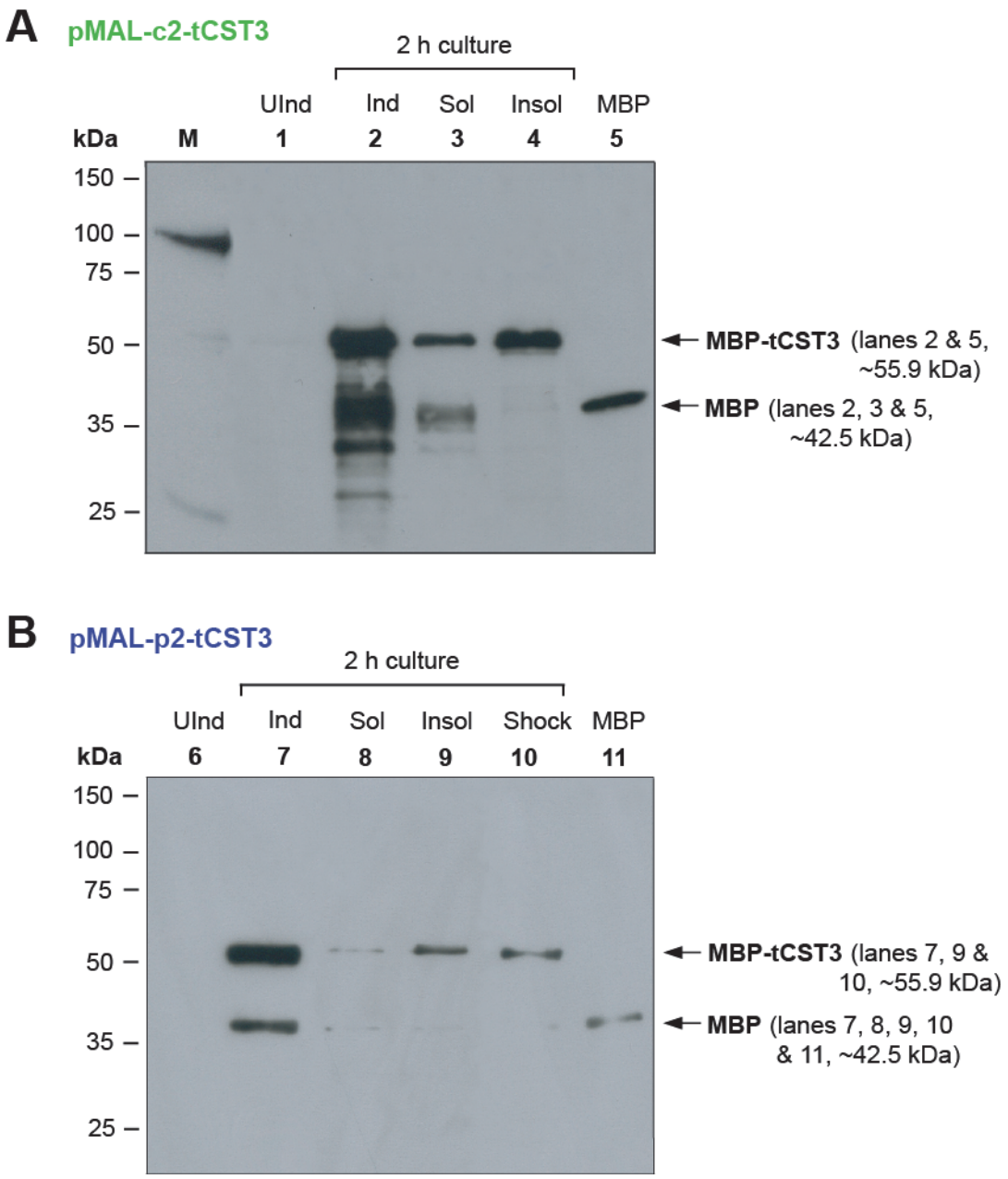


Figure 7.7. Western analysis of MBP-tCST3 protein production in the *E. coli* K12 UT5600 cells

MBP-tCST3 production was induced with 1 mM IPTG at 37°C for 2 h for **A.** pMAL-c2-tCST3 and **B.** pMAL-p2-tCST3. **A.** Cytoplasmic MBP-tCST3: Lane 1. Uninduced (UInd) bacterial culture at t_0; 2-4: 2 h induction [2. total protein (Ind); 3. soluble (Sol); and 4. insoluble (Insol) protein]; and 5. MBP. **B**. Periplasmic MBP-tCST3: 6. Uninduced (UInd) culture at t_0; 7-10: 2 h induction [7. total protein (Ind); 8. soluble (Sol); 9. insoluble (Insol); 10. periplasmic (Shock) protein]; and 11. MBP. MBP-tCST3 (~55.9 kDa) and MBP (~42.5 kDa) are arrowed. M: ECL DualVue Western Blotting Markers (GE Healthcare) show approximate protein sizes.

As the yield of periplasmic fusion proteins (pMAL-p2X) is moderate (~1-5% of *E. coli* total protein) and cytoplasmic production (pMAL-c2X) is much higher (~20-40%) (NEB, 2006), it was decided to optimise the cytoplasmic production of both MBP-tELP and MBP-tCST3 and therefore maximize recombinant protein yield by improving fusion protein solubility and reducing proteolytic degradation.

Improved solubility of MBP-tELP in the *E. coli* cytoplasm

A reduced induction temperature and IPTG concentration can improve recombinant protein solubility (Bishai *et al.*, 1987; Fox and Waugh, 2003; Niiranen *et al.*, 2007; Reddy *et al.*, 1998). Therefore, MBP-tELP and MBP-tCST3 were induced with 0.8 mM IPTG at 28°C. To compensate for a lower induction temperature growth periods of 1, 2, 3 and 4 h were trialed (induction time should be doubled if the temperature is reduced by 7°C) (NEB, 2006). The 4 h induction period produced the highest yield of MBP-ELP (Figures 7.10A, 7.10B) and MBP-CST3 (Figures 7.11A, 7.11B), with minimal soluble fusion protein degradation, so induction periods were increased to 4, 6 and 8 h.

Protein solubility was further improved by the addition of lysozyme (100 μg/mL) to resuspended cell pellets and incubation on ice for ~30-45 min prior to sonication, plus the addition of Triton X-100 non-ionic detergent to bacterial cells immediately before sonication (0%, 0.05%, 0.1% and 0.2% (v/v) Triton X-100; Figures 7.10A and 7.10B). MBP-tELP production and solubility (lane 16, Figure 7.10A) was optimal after an 8 h induction at 28°C with 0.8 mM IPTG, plus the addition of lysozyme (100 μg/mL) and Triton X-100 (0.2%). Even at 0.2% detergent, the maximum amount recommended (NEB, 2006), MBP-tELP was not completely soluble (lane 16, Figure 7.10A and lane 29, Figure 7.10B). Notably, the addition of detergent, but not lysozyme, was critical for improved fusion protein solubility. Without detergent, MBP-tELP was generally insoluble (lanes 5, 9 and 13, Figure 7.10A and lanes 18, 22 and 26 Figure 7.10B).

In contrast to MBP-tELP, MBP-tCST3 production was optimal after a 4 h induction with 0.8 mM IPTG at 28°C, plus the addition of lysozyme (100 μg/mL) and 0.1% Triton X-100 prior to sonication (Figure 7.11A). MBP-tCST3 production was stopped at this stage to focus on the purification of MBP-tELP.

MBP-tELP amylose resin affinity binding - ~40% of MBP-tELP is lost

Detergent affects the binding of the MBP fusion protein to the amylose resin (NEB, 2006). While MBP-tELP was bound to the resin in the presence of 0.05, 0.1 or 0.2% Triton X-100, Western analysis suggested that the binding affinity was around 60% (data for the 8 h induction only shown, Figure 7.12).

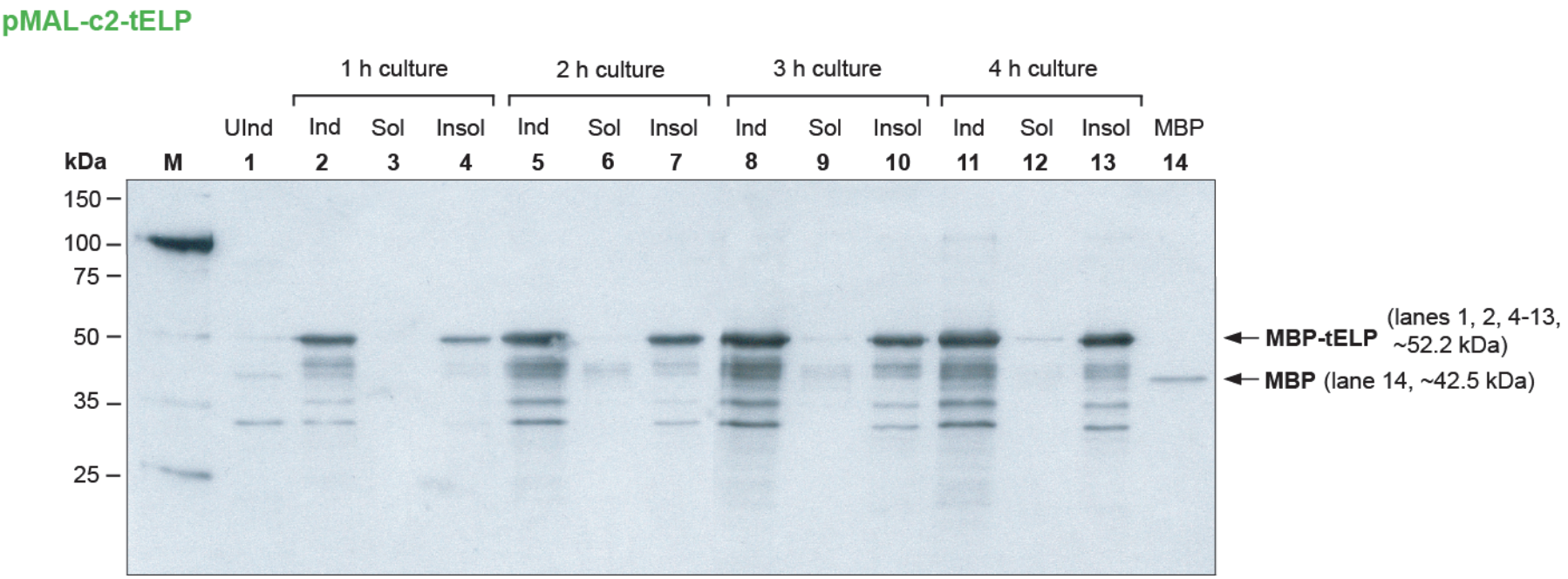


Figure 7.8. Western analysis of MBP-tELP yield and solubility over an increasing induction period

A 4 h induction period produced the highest protein yield of MBP-tELP (lane 11), but the majority of the fusion protein was insoluble (lane 13). MBP-tELP expression was induced with 0.8 mM IPTG at 28°C using the pMALc2-tELP vector in the *E. coli* K12, UT5600 strain. Induction periods of 1, 2, 3 and 4 h were tested. Samples were analysed by SDS-PAGE (10% gel), transferred to Hybond ECL membrane and MBP-tELP detected with anti-MBP antiserum. 1. Uninduced (UInd) culture at t_0; 2-4: 1 h induction [2. total protein (Ind); 3. soluble (Sol); and 4. insoluble (Insol) protein]; 5-7: 2 h induction; 8-10: 3 h induction; 11-13: 4 h induction; and 14. MBP. MBP-tELP (~52.2 kDa) and MBP (~42.5 kDa) are arrowed. M. Approximate protein sizes are indicated (ECL DualVue Western Blotting Markers, GE Healthcare).

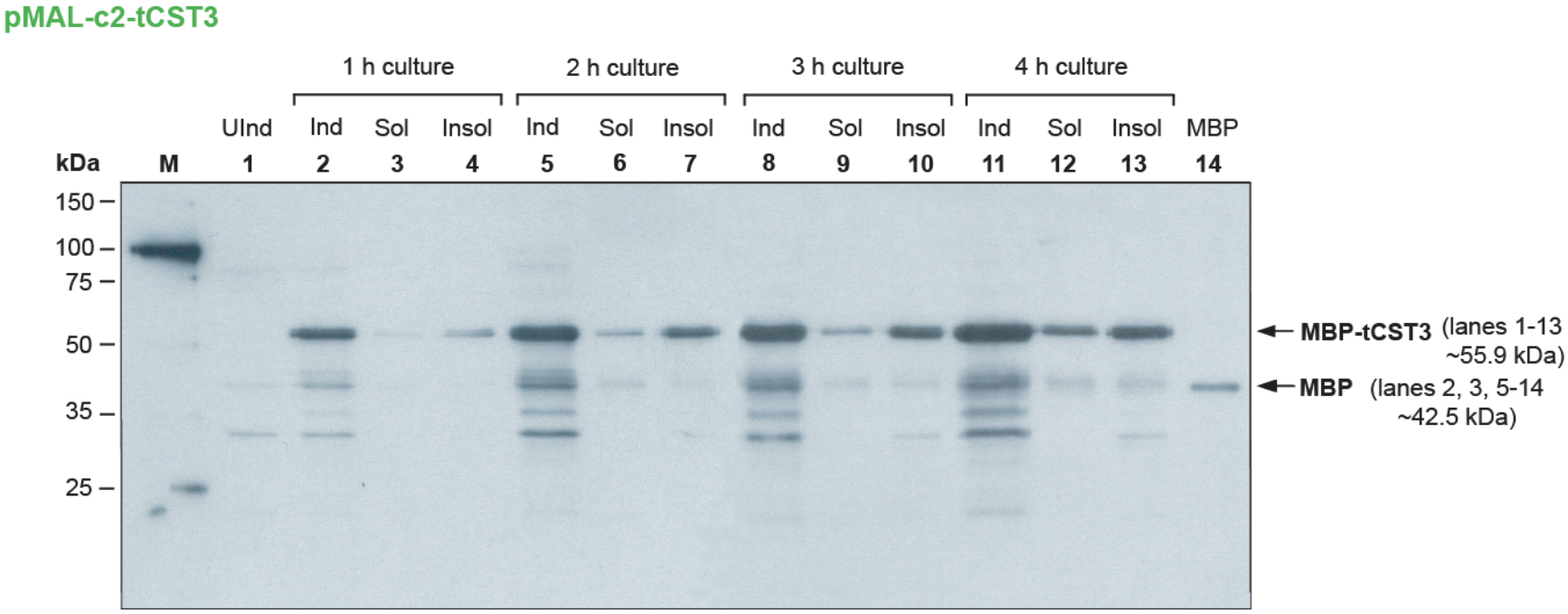


Figure 7.9. Western analysis of the effect of an increasing induction period on MBP-tCST3 yield and solubility

The yield of MBP-tCST3 was highest after induction a 4 h induction with 0.8 mM IPTG at 28°C in the UT5600 strain (lane 11). However, most of the fusion protein was insoluble (lane 13). Samples were analysed by SDS-PAGE (10% gel), transferred to Hybond ECL membrane and MBP-tCST3 detected with anti-MBP antiserum. 1. Uninduced (UInd) culture at t_0; 2-4: 1 h induction [2. total protein (Ind); 3. soluble (Sol); and 4. insoluble (Insol) protein]; 5-7: 2 h induction; 8-10: 3 h induction; 11-13: 4 h induction; and 14. MBP. MBP-tCST3 and MBP are arrowed. M. Approximate protein sizes are indicated (ECL DualVue Western Blotting Markers, GE Healthcare).

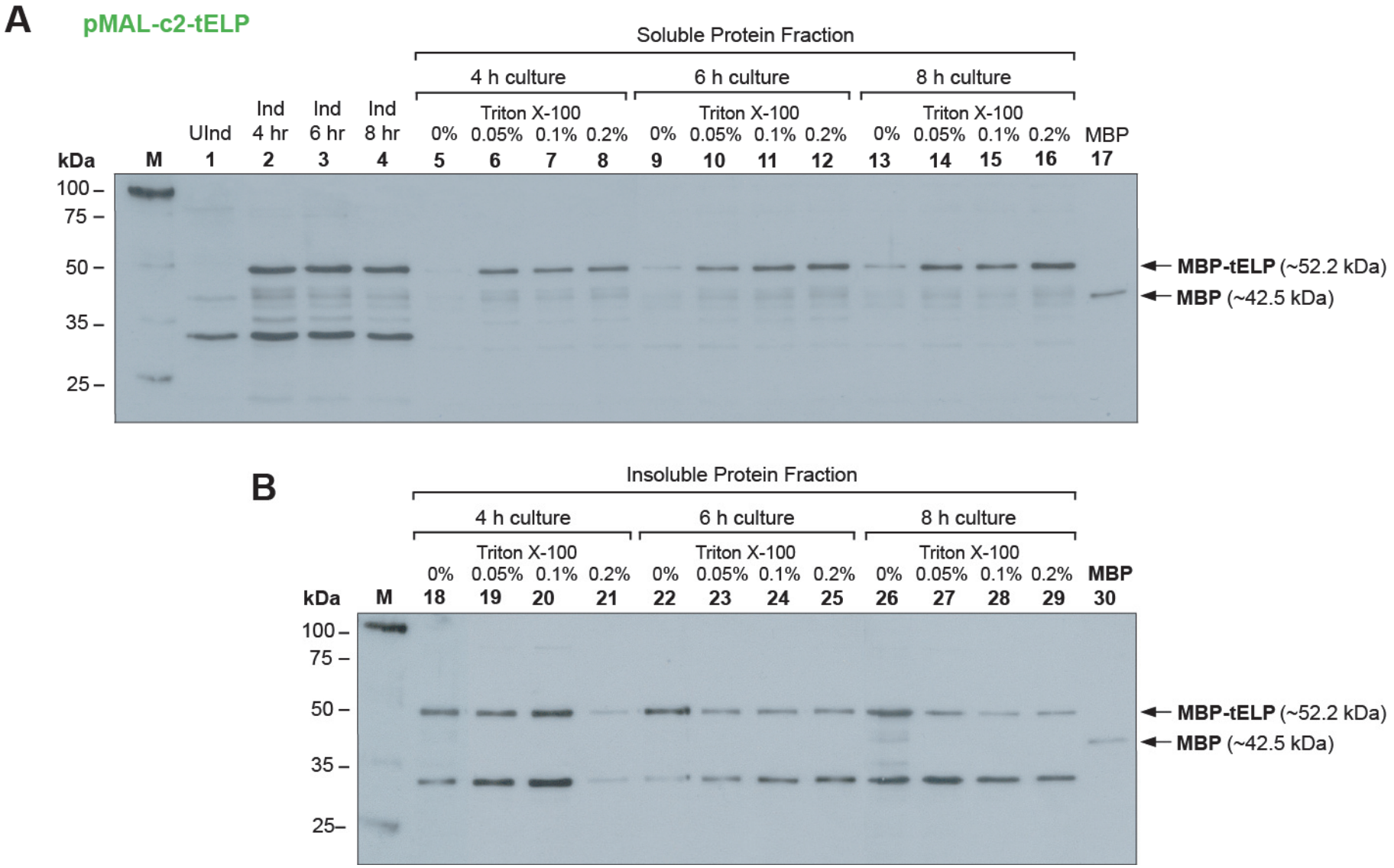

A
pMAL-c2-tELP
Soluble Protein Fraction
4 h culture
6 h culture
8 h culture
Triton X-100
0% 0.05% 0.1% 0.2%
UInd
Ind 4 hr
Ind 6 hr
Ind 8 hr
MBP
kDa
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
100 –
75 –
50 –
35 –
25 –
MBP-tELP (~52.2 kDa)
MBP (~42.5 kDa)
B
Insoluble Protein Fraction
4 h culture
6 h culture
8 h culture
Triton X-100
0% 0.05% 0.1% 0.2%
MBP
kDa
M 18 19 20 21 22 23 24 25 26 27 28 29 30
100 –
75 –
50 –
35 –
25–
MBP-tELP (~52.2 kDa)
MBP (~42.5 kDa)

Figure 7.10. Western analysis of the effect of induction time and lysosyme and detergent treatment on the yield and solubility of MBP-tELP

Detergent improves MBP-tELP solubility as increasing Triton X-100 concentration increases **A.** soluble MBP-tELP and decreases **B.** insoluble MBP-tELP. MBP-tELP expression was induced with 0.8 mM IPTG at 28°C using the pMALc2-tELP vector in the *E. coli* K12, UT5600 strain. Induction periods of 4, 6, and 8 h were tested. Lysosyme (100 µg/mL) and Triton X-100 detergent [0%, 0.05%, 0.1% and 0.2% (v/v)] were added to cells prior to sonication. Samples were analysed by SDS-PAGE (10% gel) and transferred to Hybond ECL membrane. MBP-tELP was detected with anti-MBP antiserum. Detergent improves MBP-tELP solubility, as increasing Triton X-100 concentration increases **A.** soluble MBP-tELP and decreases **B.** insoluble MBP-tELP. **A.** Lane 1. Uninduced (UInd) bacterial cells at t_0; 2. Induced (Ind) total protein (4 h); 3. Ind (6 h); and 4. Ind (8 h); 5-8: Soluble protein fraction (4 h); 9-12. Soluble protein (6 h); and 13-16. Soluble protein (8 h); 17. MBP. **B.** 18-21. Insoluble protein fraction (4 h); 22-25. Insoluble protein (6 h); 26-29 Insoluble protein (8 h); 30. MBP. Arrows indicated the MBP-tELP fusion protein (~52.2 kDa) and MBP (~42.5 kDa). Approximate protein sizes are shown (ECL DualVue Western Blotting Markers, GE Healthcare).

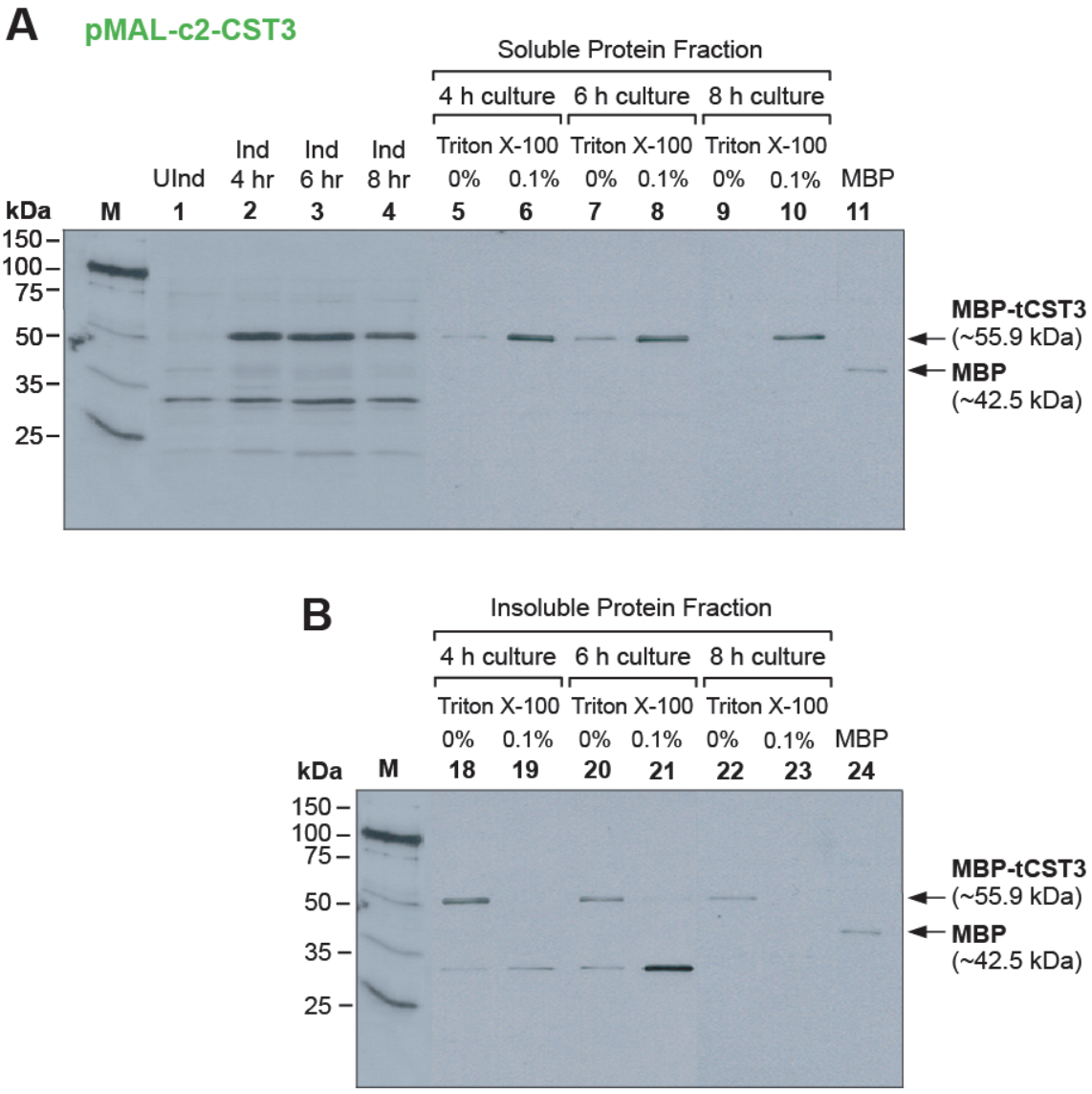


Figure 7.11. The effect of induction time and lysosyme and detergent treatment on the yield and solubility of MBP-tCST3

Detergent improves MBP-tCST3 solubility as increasing Triton X-100 concentration increases **A.** soluble MBP-tCST3 and decreases **B.** insoluble MBP-tCST3. MBP-tCST3 production was optimal after a 4 h induction with 0.8 mM IPTG at 28°C in the *E. coli* K12 UT5600 strain. Lysosyme (100 µg/mL) and Triton X-100 detergent [0%, 0.05%, 0.1% and 0.2% (v/v)] were added to cells prior to sonication. Samples were analysed by SDS-PAGE (10% gel) and transferred to Hybond ECL membrane. MBP-tCST3 was detected with anti-MBP antiserum. **A.** Lane 1. Uninduced (UInd) bacterial cells at t_0; 2. Induced (Ind) total protein (4 h); 3. Ind (6 h); and 4. Ind (8 h); 5-8: Soluble protein fraction (4 h); 9-12. Soluble protein (6 h); and 13-16. Soluble protein (8 h); 17. MBP. **B.** 18-21. Insoluble protein fraction (4 h); 22-25. Insoluble protein (6 h); 26-29 Insoluble protein (8 h); 30. MBP. Arrows indicated the MBP-tCST3 fusion protein (~55.9 kDa) and MBP (~42.5 kDa). Approximate protein sizes are shown (ECL DualVue Western Blotting Markers, GE Healthcare).

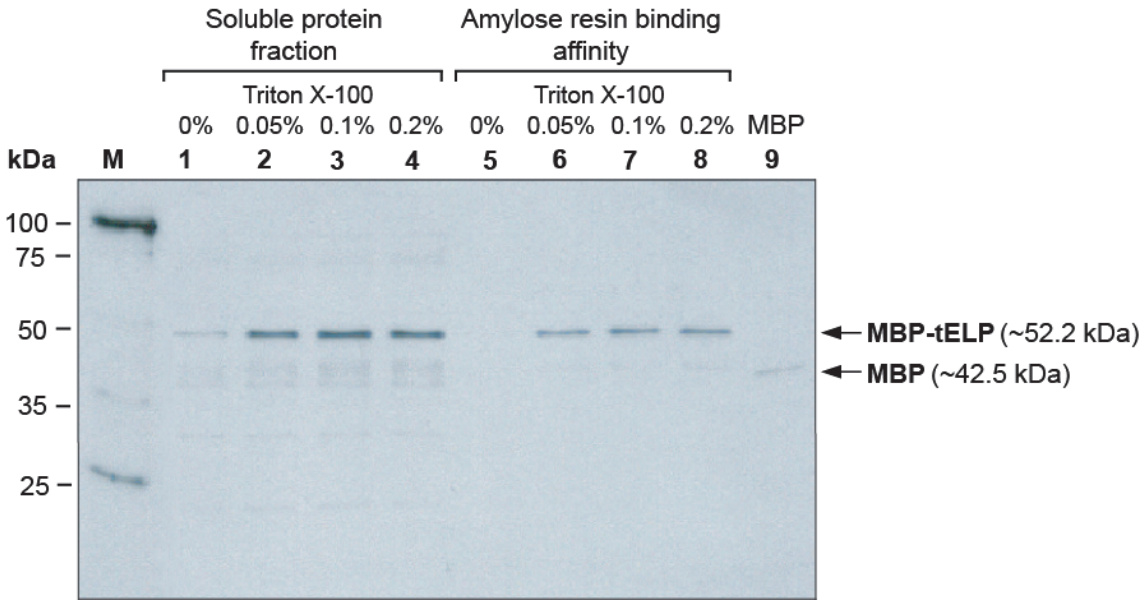


Figure 7.12. Western analysis of the effect of detergent on the amylose resin binding affinity of MBP-tELP

Detergent increases the solubility and hence the amylose resin binding affinity of MBP-tELP, but the affinity is <100%. The affinity of MBP-tELP for the amylose resin is shown in the presence of 0%, 0.05%, 0.1% and 0.2% (v/v) Triton X-100. Samples were analysed by SDS-PAGE (10% gel), transferred to Hybond ECL membrane and MBP-tELP detected with anti-MBP antiserum. Lanes 1-4. Soluble protein fraction with increasing amounts of detergent; 5-8. Amylose resin binding affinity of MBP-tELP in the presence of increasing amounts of detergent; 9. MBP. Arrows indicate MBP-tELP (~52.2 kDa) and MBP (~42.5 kDa). M. ECL DualVue Western Blotting Markers (GE Healthcare) show approximate protein size.

Factor Xa cleavage of MBP-tELP is optimal but incomplete with added detergent

Factor Xa cleavage of MBP-tELP (1 mg factor Xa/100 mg MBP-tELP fusion protein) was poor after 2, 4, 8 and 24 h incubations at RT (data not shown). Changes in temperature, buffers and salt conditions have no effect on Factor Xa enzyme activity (NEB, 2006), but the addition of SDS detergent [0.005% to 0.05% (w/v)] may improve fusion protein cleavage (Ellinger *et al.*, 1991; NEB, 2006). This was evident for MBP-tELP, with Factor Xa cleavage optimal, but incomplete after 36 h incubation with added SDS (0.01%, Lane 4, Figure 7.13A). Notably, at > 0.01% SDS, Factor Xa activity was inhibited (Lanes 5 and 6, Figure 7.13A). An increased incubation period of 5 days at RT with 0.01% SDS improved enzymatic cleavage, but fusion protein cleavage was still incomplete (Lane 2, Figure 7.13B).

Denaturation (20 mM Tris-HCl, 6 M guanidine hydrochloride, pH 7.4) and refolding (column buffer) of an MBP fusion protein may improve Factor Xa cleavage (NEB, 2006). This was true for MBP-tELP, a high proportion of which was cleaved (1 mg factor Xa/100 mg MBP-tELP) after 48 h, without SDS addition (Figure 7.14). However, MBP-tELP was cleaved at two different positions, producing two proteins at the predicted size of tELP (~9.7 kDa).

Scaled-up production of MBP-tELP - amylose resin affinity purification

Production of MBP-tELP in the *E. coli* cytoplasm was scaled up to 1 L culture volumes and fusion protein production induced with 0.8 mM IPTG for 8 h at 28°C. Lysosyme (100 μg/mL) and Triton X-100 (0.2%) were added to the harvested bacterial cells and the crude protein collected. The yield of total *E. coli* crude proteins prior to affinity purification was ~360 mg/L culture. MBP-tELP was successfully eluted (3 mL fractions) from the amylose resin with column buffer containing 10 mM maltose, collected and analysed by SDS-PAGE and Western analysis (Figure 7.15). Fractions 5 to 9 inclusive indicated the presence of the MBP-tELP fusion protein and were pooled and concentrated to 1 mg/mL under N_2. For 2 L culture volumes, MBP-tELP was eluted in fractions 4 to 15 inclusive (data not shown). Possible degradation of MBP-tELP fusion protein was observed, with the presence of two smaller bands detected in fractions 6 and 7, one consistent with MBP size and the other slightly larger. Purified MBP-tELP (~10.3 mg/L of bacterial culture) was eluted from the amylose resin column, at the lower range (3.6%) of total *E. coli* proteins. Fractions containing MBP-tELP were pooled, concentrated to 1 mg/mL and then cleaved with Factor Xa.

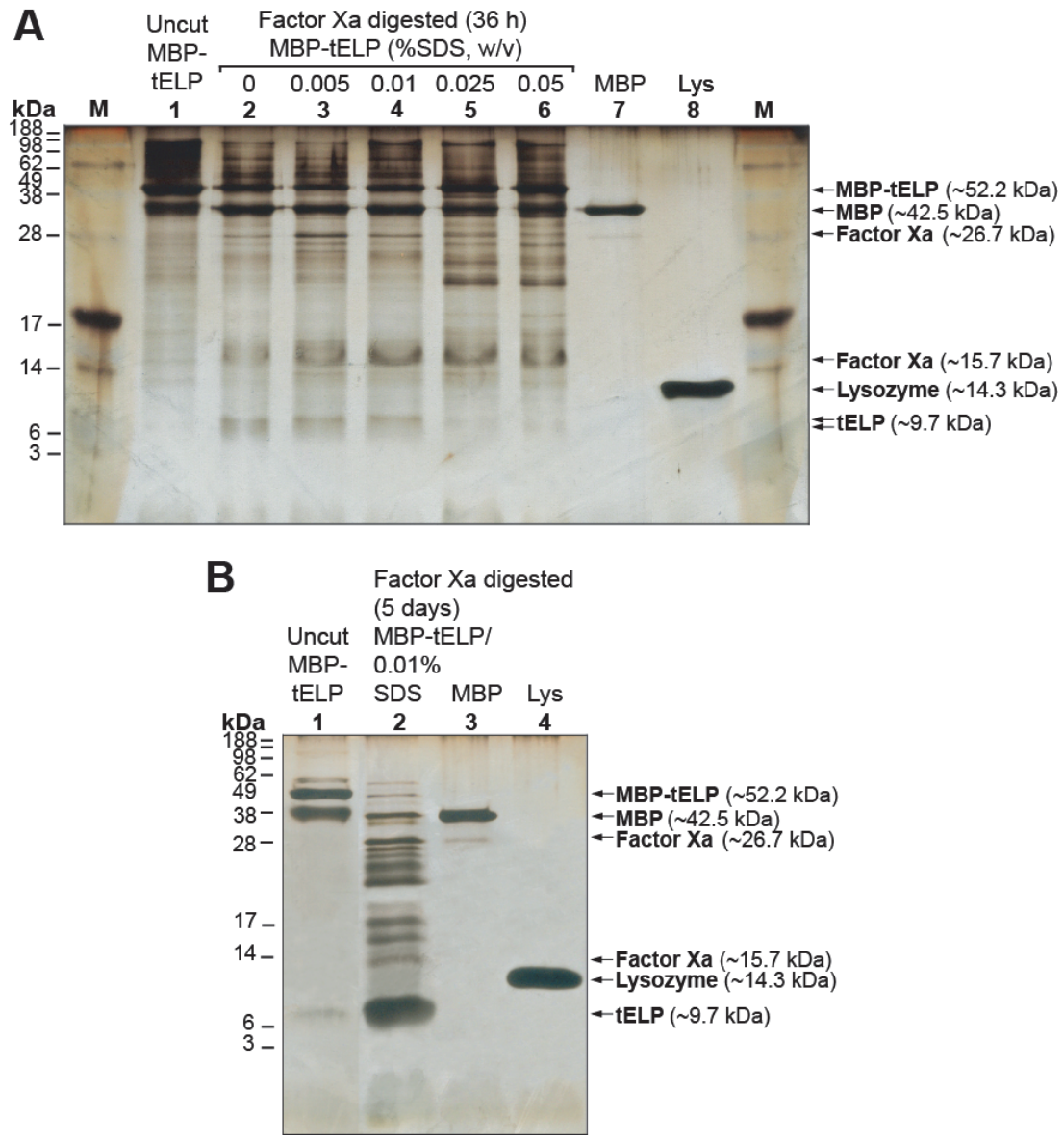


Figure 7.13. Optimisation of Factor Xa cleavage of MBP-tELP

A. MBP-tELP (1 mg/mL) was digested with Factor Xa (1 mg/100 mg fusion protein) in column buffer. Varying amounts of SDS detergent [0%, 0.005%, 0.01%, 0.025% and 0.05% (w/v)] were added to improve enzymatic cleavage over a 36 h incubation period at RT. Cleavage was optimal, but incomplete with 0.01% SDS. Protein samples (5 µg) were analysed by SDS-PAGE (15% gels) and the gel silver-stained. 1. Uncut MBP-tELP (~52.2 kDa), 2-6. Factor Xa digested MBP-tELP treated with SDS; 7. MBP (42.5 kDa); 8. Lysosyme (Lys); M. Approximate protein sizes are shown (SeeBlue Plus2 marker, Invitrogen). Uncut MBP-tELP, MBP, the Factor Xa chains (~15.7 kDa and ~26.7 kDa), lysosyme (14.3 kDa) and cleaved tELP (~9.7 kDa) are arrowed. **B**. Increasing the incubation period to 5 days at RT with 0.01% SDS added improved Factor Xa cleavage of MBP-tELP, but cleavage was incomplete. 1. Uncut MBP-tELP; 2. Factor Xa digested MBP-tELP (0.01% SDS); 3. MBP; and 4. Lysosyme. The presence of MBP and a protein corresponding to predicted size of tELP in the uncut sample suggested that MBP-tELP was degraded by *E. coli* proteases, at, or near the linker site.

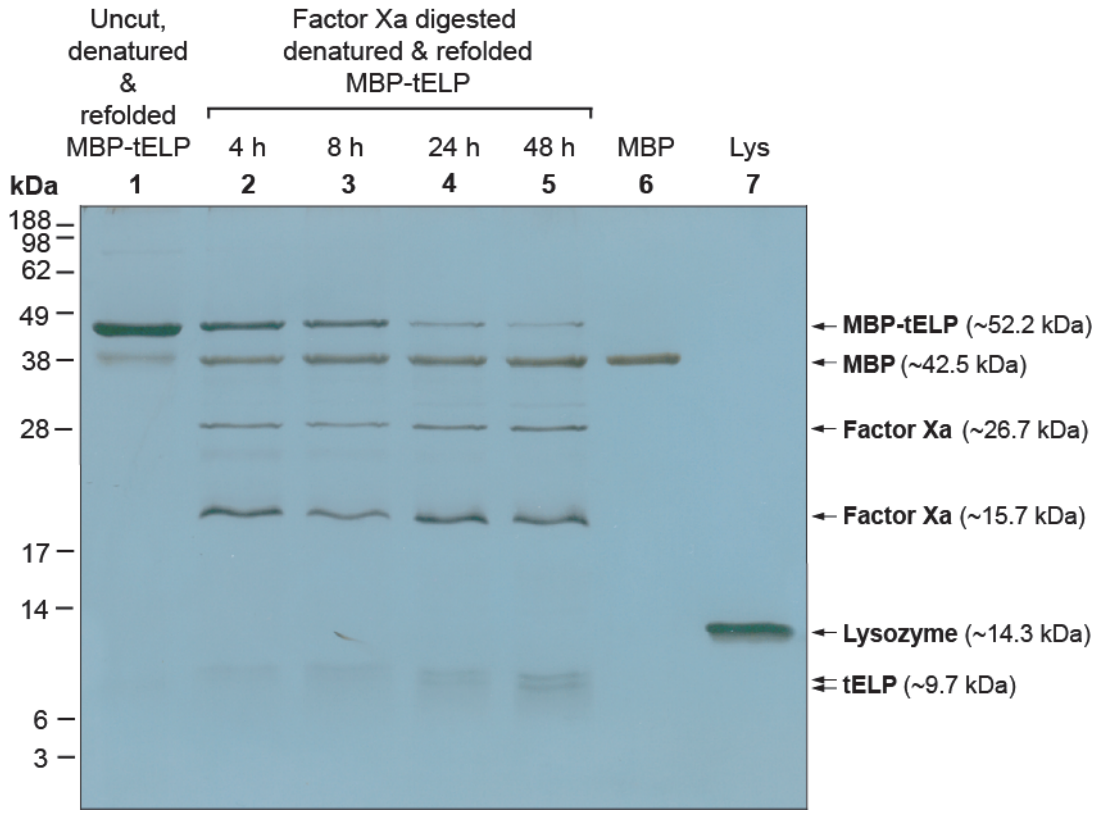


Figure 7.14. Optimisation of Factor Xa cleavage of the denatured and refolded MBP-tELP fusion protein

Denaturation and refolding and a 48 h digestion period improved Factor Xa cleavage of MBP-tELP. Two putative recombinant tELP proteins of slightly different sizes (~9.7 kDa) were produced. Amylose resin affinity-purified MBP-tELP was denatured in 20 mM Tris-HCl, 6 M guanidine hydrochloride, pH 7.4, refolded in column buffer and digested with Factor Xa (1 mg/100 mg fusion protein) for 2, 4, 8, 24 and 48 h. Samples were analysed by SDS-PAGE (10%) and the gel silver-stained. 1. Uncut denatured and refolded MBP-tELP; 2-5. Factor Xa digested MBP-tELP (2. 2 h; 3. 8 h; 4. 24 h; and 5. 48 h), 6. MBP; and 7. Lysosyme (Lys). Cleavage of MBP-tELP was maximal after 48 h (lane 5). Uncut MBP-tELP, MBP, the Factor Xa chains (~15.7 kDa and ~26.7 kDa), lysosyme (14.3 kDa) and cleaved tELP (~9.7 kDa) are arrowed. M. Approximate protein sizes are indicated (SeeBlue Plus2 marker, Invitrogen).

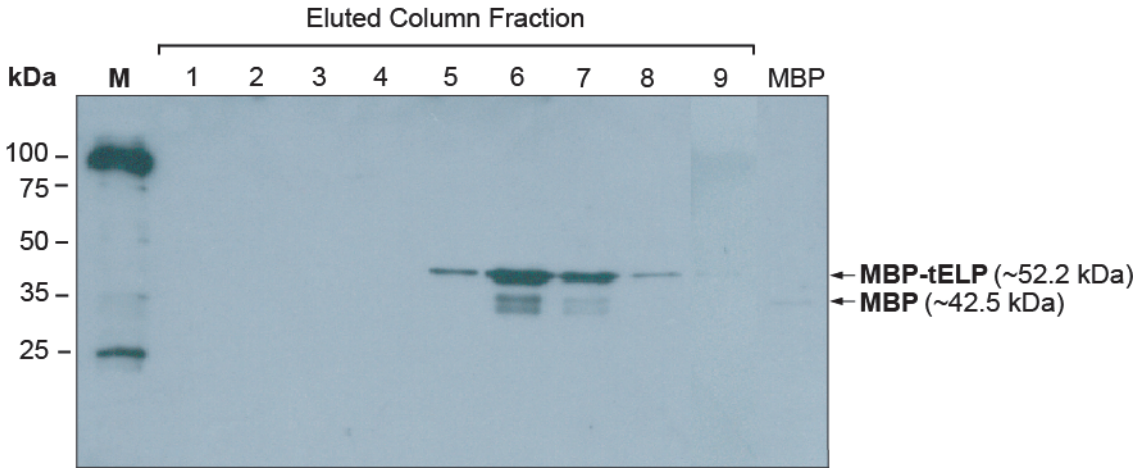


Figure 7.15. Amylose resin affinity-purification of MBP-tELP

MBP-tELP was affinity-purified from crude protein extract on a 2.5 x 10 cm amylose resin column equilibrated with column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM sodium azide, 10 mM β-ME. Fractions (3 mL) were eluted with column buffer supplemented with 10 mM maltose. Samples were analysed by SDS-PAGE (10% gel), transferred to Hybond ECL membrane and MBP-tELP detected with anti-MBP antiserum. MBP-tELP was eluted in fractions 5 to 9 inclusive. MBP-tELP (~52.2 kDa) and MBP protein (42.5 kDa) (lane 10) are arrowed. Putative MBP-tELP degradation products are present in elution fractions 6 and 7. M. ECL DualVue Western blotting markers (GE Healthcare) indicate protein sizes.

Gel purification of recombinant tammar ELP

Factor Xa cleaved MBP-tELP was concentrated to ~10 mg/mL and purified by gel filtration on a 1.6 x 55 cm Sephadex G-50 coarse resin column with a bed volume of 120 mL (Figure 7.16). Three major peaks were observed, consistent with the successful separation of MBP from its fusion partner (Ashraf *et al.*, 2004; Buck *et al.*, 2003). Based on these studies and the elution profiles of ovalbumin (44.3 kDa) and α-LA (14.2 kDa) in column calibration tests (data not shown), the first peak includes MBP (42.4 kDa)/Factor Xa (42.4 kDa); the second, recombinant tELP (~9.7 kDa); and the third peak, degraded tELP and other *E. coli* protein contaminants. Recombinant tELP protein was eluted as a doublet (inset, Figure 7.16), consistent with MBP-tELP Factor Xa cleavage (Figure 7.14). However, protein loss for the gel filtration step was ~30%, reducing the tELP yield to a very low yield of <0.5 mg/L bacterial culture.

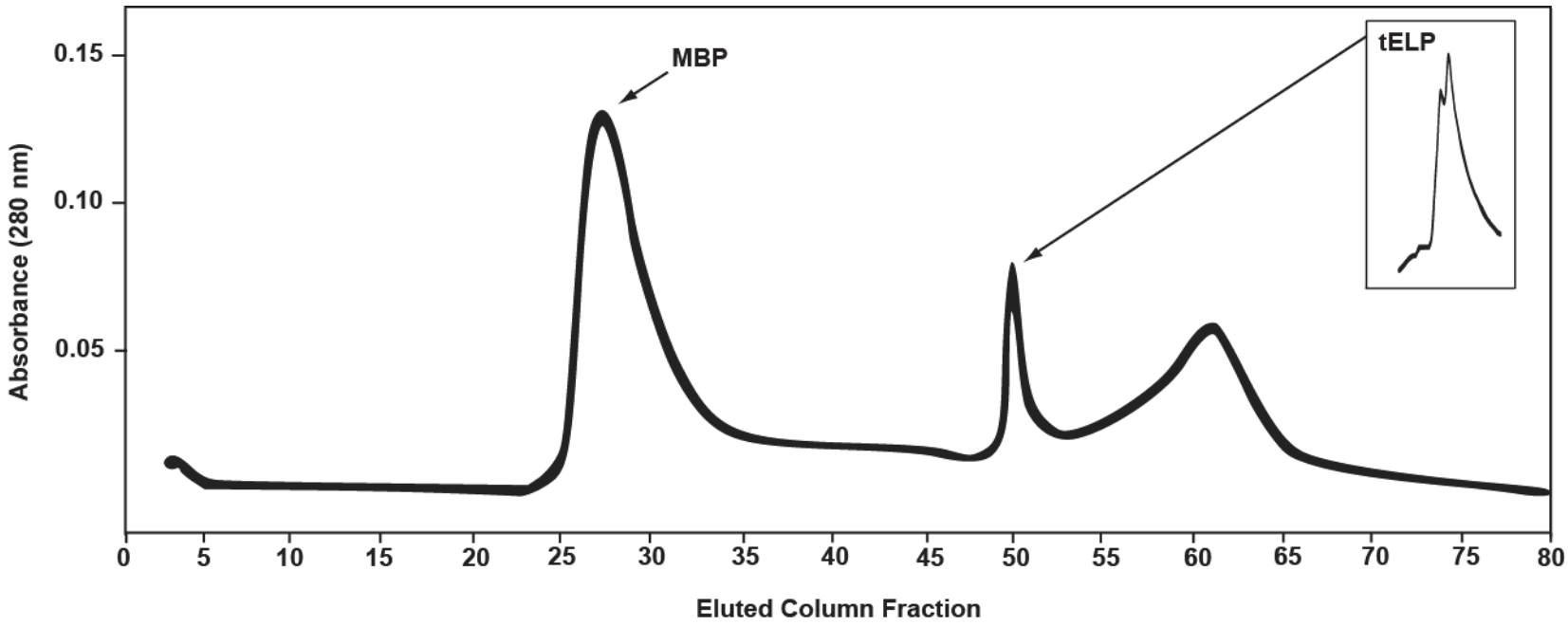


Figure 7.16. Purification of tELP by gel filtration

tELP was purified on a 1.6 x 55 cm Sephadex G-50 coarse resin column equilibrated with degassed 30 mM ammonium bicarbonate buffer containing 10 μM calcium chloride. A flow rate of 10 mL/h was used and 2 mL fractions collected. Three peaks were eluted: 1. MBP (42.5 kDa) and Factor Xa; 2. recombinant tammar ELP [~9.7 kDa, eluted as a doublet (inset)]; and 3. an additional peak that most likely included *E. coli* proteins and degradation products.

Production and isolation of the MBP-tELP fusion protein from inclusion bodies

The predominantly insoluble MBP-tELP fusion protein was successfully isolated from inclusion bodies from of 2 L cultures of the UT5600 *E. coli* K12 strain containing the pMAL-c2-tELP (Figure 7.17). Although, the yield of total protein isolated from inclusion bodies (~25 mg/L) was less than one tenth that isolated from the soluble protein fraction (~360 mg/L), there were fewer impurities.

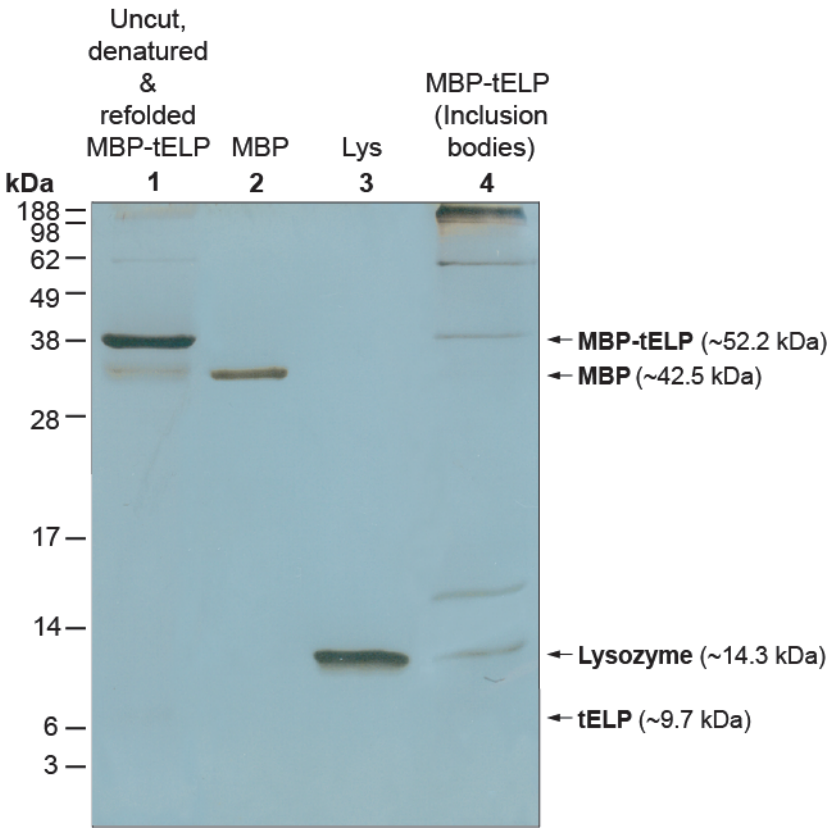


Figure 7.17. Production of MBP-tELP in *E. coli* inclusion bodies

MBP-tELP was overexpressed with pMALc2-tELP in the cytoplasmic fraction of *E. coli* K12 UT5600 bacterial cells. The fusion protein was isolated from inclusion bodies, solubilised and refolded in dialysis buffer (20 mM Tris-HCl, pH 8.5) containing 1 mM reduced glutathione and 2 mM oxidised glutathione. Total protein (5 µg) was analysed by SDS-PAGE (10% gel) and the gel silver-stained. Lane 1. Uncut denatured and refolded MBP-tELP; 2. MBP; 3. Lysosyme (Lys); and 4. MBP-tELP isolated from *E. coli* inclusion bodies, solubilised and refolded. MBP-tELP (~52.2 kDa), MBP (42.5 kDa), lysosyme (14.3 kDa) and tELP (very faint band ~9.7 kDa, lane 1) are indicated by arrows. Approximate protein sizes are based on the SeeBlue Plus2 marker, Invitrogen.

Discussion

The availability of recombinant tammar ELP would provide a better understanding of the function of this temporally expressed, putative Kunitz inhibitor which is secreted into milk only during the early lactation period; a critical time for PY development and survival. Unfortunately, the mammalian and bacterial protein expression systems used to express recombinant tammar ELP were not ideal. In particular, the lack of an antibody to detect tELP *in vitro* and the absence of a visible secreted protein in the mammalian expression systems used meant that the bacterial expression system was the best method for the production of significant (mg) quantities of the MBP-tELP fusion protein. MBP is one of the best fusion partners for maximising fusion protein solubility (Fox *et al.*, 2003). However, both MBP-tELP and MBP-tCST3 were preferentially expressed as insoluble inclusion bodies in the *E. coli* cytoplasm and periplasm when expressed with the pMAL-c2X and pMAL-p2X vectors, respectively. However, the solubility and hence yield of MBP-tELP was the improved by the addition of lysozyme (100 μg/mL) and Triton X-100 detergent (0.2%) prior to sonication. Bacterial proteolysis of both MBP-tELP and MBP-tCST3 was a common problem and degradation was reduced but not eliminated by the use of the *E. coli* K12 UT5600 *ompT* protease-deficient strain and the optimisation of culture conditions. These included induction with 0.8 mM IPTG at a 28°C for 8h for MBP-tELP, but only 4h for MBP-tCST3. Cleavage of tELP from MBP with Factor Xa was poor, but improved by the addition of SDS (0.01% w/v) and an extended 5-day incubation period. MBP and tELP were separated by gel filtration, but whether tELP was produced in a biologically active form is unknown, particular as the function of tELP is yet to be determined.

Successful production of tammar CST3, but not ELP and CTI in mammalian expression systems

Mammalian expression systems provide one of the best methods to produce correctly folded and biologically active eukaryotic proteins, particularly for glycosylated proteins, or those with disulphide bonds (Chapple *et al.*, 2006; Hartley, 2006; Sethuraman and Stadheim, 2006; Swietnicki, 2006). However, not all proteins are expressed and secreted at high levels in these systems (Demain and Vaishnav, 2009). While tammar CST3 was successfully expressed *in vitro* and secreted into conditioned media, the lack of visible tammar and opossum ELP or bovine and seal CTI recombinant proteins was unexpected, particularly as ELP and CTI were transcribed *in vitro*. Moreover, all expression constructs had a 5' UTR Kozak consensus sequence, an initiation codon, a stop codon and the complete 3' UTR up to and including the polyadenylation signal.

Hence, the absence of ELP and CTI secretion *in vitro* may be due to post-transcriptional regulatory mechanisms. Alternatively, HEK293Ts may have failed to export the ELP and CTI transcripts to the cytoplasm for protein assembly on the ribosomes.

The introduction of an additional methionine codon preceding the correct start codon in both of the pCMVSport6-tELP clones during library preparation may also have affected *in vitro* protein production. All tammar mammary gland transcripts were cloned into a 5' EcoRV (5'-G*AT*/ATC-3') half-site. For *tELP* this produced *AU*GCACCAUGA for SGT20e4_A04 and *AU*GUGGCACCAUGA for SGT20p3_C08 (vector sequence is italicised and the correct *tELP* methionine codon underlined). Translation of these clones from the introduced methionine codon would produce out-of-frame, non-secreted proteins, with predicted molecular weights of 6.2 kDa and 6.0 kDa, respectively. The absence of products of this size in either the supernatants or cell lysates suggested that post-translational mechanisms may regulate *tELP* expression. Although an additional methionine codon was also incorporated into tammar CST3 (SGT20o3_C05), *in silico* analysis showed that this in-frame mutation would produce an identical ~13.32 kDa mature cystatin C protein.

Mammalian expression systems have been used by many researchers to express and purify eukaryotic proteins for functional studies (Chapple *et al.*, 2006; Kawaguchi *et al.*, 1997; Shimomura *et al.*, 1999). However, without the use of an epitope-tag, such as the FLAG, HA, c-myc, 6-His or AU1 tags (Zhang and Chen, 1998) and without an antibody to ELP or CTI, the detection and purification of these proteins was problematic. The small octopeptide FLAG epitope (DYKDDDDK) (Hopp *et al.*, 1988) would provide an excellent method for both the detection and purification of a FLAG-tELP recombinant protein by FLAG affinity chromatography (Brizzard *et al.*, 1994; Chubet and Brizzard, 1996). Although the FLAG protein purification resin has a low capacity and is expensive, it does produce a high-purity protein (Lichty *et al.*, 2005). Furthermore, an enzyme cleavage site could be included between the FLAG and tELP peptides which would facilitate the removal of the highly-charged FLAG peptide prior to functional studies of tELP. Alternatively, the use of the shorter FLAG peptide DYKD, which is recognized with similar affinity to the full-length FLAG-tag (Knappik and Pluckthun, 1994) may negate the need to remove the FLAG peptide for future applications.

Successful production of the MBP-tELP and MBP-tCST3 fusion proteins

While the production of recombinant tELP was unsuccessful in the pCMV-Sport6 and pcDNA3.1 vectors, the pMAL vectors proved far more effective. However, both

MBP-tELP and MBP-tCST3 were predominantly produced as insoluble inclusion bodies in both the cytoplasmic and periplasmic fractions and were prone to degradation by *E. coli* proteases. Although fusion protein degradation was reduced, but not eliminated by the use of the *E. coli* K12 UT5600 strain to produce MBP-tELP and MBP-tCST3, the addition of protease inhibitors prior to sonication (Guan *et al.*, 2002) may eliminate proteolysis. Furthermore, the *E. coli* B strains including BL21 and its derivatives, lack the *lon* and *ompT* proteases which degrade misfolded; and cytoplasmic and periplasmic proteins respectively (Grodberg and Dunn, 1988; Phillips *et al.*, 1984) and hence may provide better alternatives for the *in vitro* production of intact recombinant MBP-tELP and MBP-tCST3. Additional factors, such as: expression construct sequence and design; disulphide bond formation and hence 3D-structure and Factor Xa cleavage, may also reduce bacterial expression and yields of the fusion and recombinant proteins, but could be improved, given additional time and resources.

Replacement of rare *E. coli* codons in the protein of interest

Recombinant protein yields can be reduced by over 20% due to the presence of rare *E. coli* codons in the coding sequence of the target protein (see reviews) (Gustafsson *et al.*, 2004; Overton, 2014; Rosano and Ceccarelli, 2014; Sorensen *et al.*, 1989). Rare codons may produce mistranslations, e.g. Arg to Lys and Arg to Gln (McNulty *et al.*, 2003), frameshifts (Spanjaard and van Duin, 1988) and truncated proteins because ribosomes stall at rare codons, uncoupling translation from transcription (Bonekamp and Jensen, 1988; Del Tito *et al.*, 1995; Kane, 1995; Zahn, 1996). The tammar *ELP* and *CST3* constructs included 5 (4X Arg, 1X Leu) and 6 (4X Arg, 2X Pro) rare codons respectively and proteins intermediate in size between the MBP-fusion and MBP were detected, indicative of truncation. In order to prevent premature termination, the Rosetta or Rosetta 2 (Novagen) *E. coli* strains could be used for recombinant ELP and CST3 production as they carry the pRARE and pRARE2 plasmids that encode 6 and 7 of the 8 rare *E. coli* codons respectively. The Rosetta strain would be sufficient for production of both tELP and tCST3 as it encodes all of the rare codons present in these transcripts.

Improvement of MBP fusion protein solubility

MBP acts like a molecular chaperone, enhancing the solubility and promoting the correct folding of the passenger protein into its native, biologically active conformation (Dyson *et al.*, 2004; Fox *et al.*, 2003; Gopal and Kumar, 2013). Yet, it was only by addition of Triton X-100 detergent that MBP-tELP and MBP-tCST3 solubility was improved but this strategy was not as effective for MBP-tELP as more than 20% of the

protein remained insoluble. Other strategies which could improve fusion protein solubility include: a reduced growth temperature (less than 28°C); a plasmid with a weaker promoter (ptac is a strong promoter) and/or a low copy number plasmid, reduced IPTG concentration for induction, changing the growth media and/or adding chaperones, foldases and/or isomerases (Arya *et al.*, 2015; Correa and Oppezzo, 2015; de Marco, 2013; Rosano and Ceccarelli, 2014). However, not all MBP-fusions can be made soluble (Raran-Kurussi and Waugh, 2012; Reddy *et al.*, 1998) and not all soluble MBP-fusions are biologically active (Dyson *et al.*, 2004; Fox and Waugh, 2003; Sachdev and Chirgwin, 1998).

Disulphide bond formation in *E. coli*

As the native tammar ELP and CST3 have three and two disulphide bonds respectively, future strategies to improve the cytoplasmic production of the biologically active fusion proteins should use an *E. coli* host strain that has mutations in both the thioredoxin reductase (*trx*B) and glutathione reducase (*gor*) genes. These mutations enhance disulphide bond formation and promote correct protein-folding (Bessette *et al.*, 1999; Derman *et al.*, 1993; Prinz *et al.*, 1997; Stewart *et al.*, 1998) are present in the Origami 2, Origami B and Rosetta-gami 2 BL21 strains (Novagen) and the SHuffle strains (NEB) (de Marco, 2009; Lobstein *et al.*, 2012). Although the SHuffle strain constitutively expresses a chromosomal copy of DsbC (disulphide bond isomerase), which acts as a chaperone and repairs incorrectly formed disulphide bonds (Berkmen, 2012; Lobstein *et al.*, 2012; Nozach *et al.*, 2013), it lacks the rare codon-encoding of the Rosetta-gami 2 BL21 strain and hence the latter would be the best alternative for MBP-tELP and MBP-tCST3 production trials.

While the oxidative environment of the *E. coli* periplasm is adept at disulphide bond formation (Berkmen, 2012), the periplasmic volume is small and so recombinant protein yields are much lower than for cytoplasmic protein production (Fox and Waugh, 2003; Sachdev and Chirgwin, 2000). However, inclusion bodies have a higher density and the periplasm can protect the target protein against proteolysis (Eiberle and Jungbauer, 2010). Many proteins with disulphide bond-containing Kunitz domains have been successfully produced in the *E. coli* periplasm, eg. BPTI (Ghrayeb *et al.*, 1984; Goldenberg, 1988; Marks *et al.*, 1986; Zhang and Goldenberg, 1993); the Kunitz domains from β-amyloid protein precursor (Oyama *et al.*, 1993) and mamba dendrotoxins K, I and E (Smith *et al.*, 1995). Biologically active periplasmic cystatin C has also been produced (Dalboge *et al.*, 1989). The resistance of MBP-tCST3 to proteolysis in the *E. coli* periplasm (shock fraction) suggests that this method is feasible

for the production of recombinant tammar cystatin C. Though MBP-tELP was susceptible to periplasmic proteases, fusion protein production in an alternate *E. coli* strain, or the addition of protease inhibitors may prevent proteolysis and improve protein yield.

Isolation of MBP-tELP from insoluble inclusion bodies using denaturation and refolding reduces *E. coli* protein impurities

While the yield of MBP-tELP from inclusion bodies was much lower than for amylose resin affinity-purification, the denaturation and re-folding process simplifies *in vitro* protein production, reducing *E. coli* protein impurities and often producing a biologically active recombinant protein (Baneyx, 1999; Cabrita *et al.*, 2004; Eiberle and Jungbauer, 2010; Kingston and Brett, 2005). Functional Kunitz domain proteins that have been successfully extracted from inclusion bodies include: BPTI and BPTI mutants (Buczek *et al.*, 2002; Bulaj *et al.*, 2004; Krokoszynska *et al.*, 1998; Zhang and Goldenberg, 1993), TFPI (Diaz-Collier *et al.*, 1994), the first domain of TFPI2 (Chand *et al.*, 2004), HAI-1B (Shia *et al.*, 2005) and the *Naja atra* chymotrypsin inhibitor (NACI) from Taiwan cobra venom (Cheng *et al.*, 2005). Encouragingly, the lack of precipitation of MBP-tELP after denaturation and refolding suggested that tELP may have the correct 3D conformation *in vitro* (Eisenmesser *et al.*, 2000; Donnelly *et al.*, 2006). This could be verified using circular dichroism (CD), or Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid), DTNB). DTNB reacts with free cysteine residues and therefore indicates the number of free thiols (and hence the number of disulphide bonds) in a protein (Ellman, 1959).

Improving Factor Xa enzymatic cleavage of MBP-tELP

The Kunitz domains of both BPTI (Wearne, 1990), which most-likely evolved from the common ancestral gene of *ELP*/*CTI* (Pharo *et al.*, 2012), and TFPI domain 2 inhibit factor Xa (Burgering *et al.*, 1997; Girard *et al.*, 1989). Although the cleavage rate of MBP-tELP was low (indicative of Factor Xa inhibition), the production of two products at the expected size of recombinant tELP (~9.7 kDa) showed that Factor Xa was active. Denaturation and refolding of MBP-tELP and the addition of 0.01% SDS did improve Factor Xa activity, implying that tammar ELP was misfolded. The use of pMAL vectors with either Genenase I or enterokinase cleavage sites was considered as an alternative to factor Xa. However, Genenase I is predicted to cleave mature tELP and enterokinase is a gut serine protease and therefore its activity may be inhibited by tELP. A future strategy for MBP-tELP cleavage could involve the incorporation of an alternate enzyme cleavage site, for example tobacco etch virus (TEV) protease, a cysteine protease which

is unlikely to be inhibited by tELP.

Production of N-glycosylated tammar ELP

The pMAL bacterial expression system provided an effective method for producing MBP-tCST3 and MBP-tELP, but whether *E. coli* or indeed other bacteria can produce N-glycosylated forms of tELP characteristic to the mammary gland and milk (Joss *et al.*, 2009; Wongtrakul-Kish *et al.*, 2013) is unknown. Glycoproteins with attached oligosaccharides, particularly sugars with multiple outer branches tend to be bulky and hydrophilic, altering the physical properties of the protein, both at the local attachment site and globally (Dwek, 1996). Therefore, it is critical to determine the function of both non-glycosylated and N-glycosylated tELP (and its multiple glycoforms (Joss *et al.*, 2009)), both *in vivo* and *in vitro*. While yeasts produce complex glycoproteins, the attached oligosaccharides tend to be homogeneous (Bobrowicz *et al.*, 2004) and hence not ideal for tELP production. Furthermore, the cell-specificity of N-glycosylation suggests that MECs, particularly those from the tammar mammary gland should be used to express recombinant tELP *in vitro*. Immortalised tammar MECs are yet to be developed, but they could be produced by the overexpression of TERT (telomerase reverse transcriptase), the expression of the SV (simian virus) 40 large T antigen, the mutation (or siRNA silencing) of the *p53* and *pRb* (retinoblastoma) cell cycle control (and tumour suppressor) genes, the expression of viral oncogene(s), e.g. human papillomavirus (HPV) oncogenes E6 and E7, or a combination of these strategies (see reviews) (Maqsood *et al.*, 2013; Noguchi and Kobayashi, 2013). TERT increases telomere length, stabilising chromosomes overcoming the Hayflick limit (the number of times a normal cell can divide) and replicative senescence and making cells immortal (Shay and Wright, 2000).

To tag or not to tag?

Future strategies for *in vitro* tELP and tCST3 production should involve the use of affinity tag(s) which enable both protein detection and purification. Although the *in vitro* production of tELP and tCST3 fused to MBP was successful, the 4-5X greater size of MBP compared to tELP and tCST3 reduces target protein yield (Wood, 2014). Therefore, the use of alternate protein and/or peptide tags may improve the production of these small (~9 kDa and ~13 kDa) proteins respectively. The HaloTag Protein Purification System (Promega Corporation) enhances protein solubility (Hata and Nakayama, 2007; Motejadded *et al.*, 2010) and can be used to study protein-protein and protein-DNA interaction *in vivo* and *in vitro* (Daniels *et al.*, 2014; Urh *et al.*, 2008).

This would permit the identification of factor(s) that interact with tELP and tCST3 in the mammary gland, milk and in the gut of the young by pull-down assays (Daniels *et al.*, 2014; Guan and Kiss-Toth, 2008; Ratner, 1974). The recently developed Flag-Acidic-Target Tag (FATT) (Sangawa *et al.*, 2013) may be an ideal tag for tELP and tCST3 detection and purification. The FATT system includes the FLAG tag and a hyper-acidic, highly-charged segment from the extracellular region of human amyloid precursor protein (APP). The FATT system has high expression in *E. coli*, improves target protein solubility and correct folding, facilitates the proper refolding of misfolded fusion partners that contain disulfide bonds and offers a single-step purification on a conventional anion exchange chromatography resin making it affordable (Sangawa *et al.*, 2013). The use of compound tags for target protein purification and detection (see reviews (Gräslund S, 2008; Wood, 2014) could also be used in future, particularly for tELP, for which no antibody is currently available.

Conclusions

MBP-tELP and MBP-tCST3 and recombinant tELP were successfully produced *in vitro* with the pMAL system. However, significant improvements could be made by using an appropriate *E. coli* strain to produce correctly folded recombinant tammar proteins, the use of protease inhibitors and a the addition of a second tag for enhanced detection and purification. Ideally, the function of the N-glycosylated and non-glycosylated tammar ELP should be determined. This could involve the expression of tELP in mammalian and bacterial expression systems respectively, or the removal of oligosaccahrides from N-glycosylated tELP using Peptide-*N*-Glycosidase F (PNGase F). Furthermore, the production of an antibody to either the native tammar ELP milk protein, or a synthetic antigenic peptide would facilitate future investigations of tELP function *in vitro* and *in vivo*.

Chapter Eight:

General discussion and future directions

Discussion

Marsupial lactation is unique. The tammar wallaby has a relatively short pregnancy of ~26.5 days after diapause and gives birth to a ~400 mg altricial 'exteriorised fetus' (Menzies *et al.*, 2012; Renfree *et al.*, 1989). During the first ~200 days pp, the PY remains in its mother's pouch, effectively an *ex utero* gestation (Tyndale-Biscoe and Renfree, 1987). During this period, the PY is permanently (0-100 days pp; Phase 2A) and then intermittently (~100-200 days pp; Phase 2B) attached to its mother's teat (Tyndale-Biscoe and Renfree, 1987) and its growth and development are driven entirely by its mother's milk (Menzies *et al.*, 2012; Nicholas *et al.*, 1997; Tyndale-Biscoe and Renfree, 1987). In contrast, during Phase 3 (~200-350 days pp) the PY begins to ingest herbage while still depending on milk, more like the post-natal eutherian pattern of early nutrition (Nicholas *et al.*, 1997; Tyndale-Biscoe and Janssens, 1988). In order to provide appropriate nutrition for the considerable growth and development of its pouch young, the tammar mother progressively alters milk composition and individual milk constituents throughout the long and complex 300-350 day lactation period (Green and Merchant, 1988; Nicholas *et al.*, 1997; Tyndale-Biscoe and Renfree, 1987). Not all milk components are conserved in all mammals (Oftedal, 2013; Oftedal, 2012), with specific constituents tailored for the needs of the young.

Early lactation protein (*ELP*) was first identified in the brushtail possum (Piotte and Grigor, 1996) and then in the tammar wallaby (Nicholas *et al.*, 1997; Simpson *et al.*, 1998) and striped-faced and fat-tailed dunnarts (De Leo *et al.*, 2006). The presence of a Kunitz domain, indicative of serine protease inhibition and its expression in the early lactation mammary gland suggested that *ELP* and *colostrum trypsin inhibitor* (*CTI*) may be orthologous genes. In this thesis, I examined the evolutionary history, expression and regulation of the tammar wallaby *ELP* gene, identified the tammar *ELP* promoter and evaluated methods to produce recombinant tammar ELP *in vitro*. Additionally, the eutherian orthologue of marsupial *ELP* was identified by comparative genomics analyses.

The Evolution of the *Early Lactation Protein* gene

ELP was originally thought to be marsupial-specific (Lefèvre *et al.*, 2007). However, my research has shown that marsupial *ELP* and the eutherian *CTI* gene evolved from a common ancestral gene (Figure 3.9) prior to the divergence of marsupials and eutherians over 160 million years ago (Luo *et al.*, 2011). Comparative genomic analysis

of *ELP* and *CTI* confirmed that these genes are true orthologues as they share the same three exon structure (Figures 3.1 and 4.3; Additional files 3.1 and 4.4), the conserved single-copy *PIGT* and *WFDC2* flanking genes (Figures 3.4 and 3.5) and common repeat elements (Additional files 3.3, 3.10 and 3.11).

While *ELP* and *CTI* are not present in amphibians, birds, reptiles and fish, the status of these genes in extant monotremes is inconclusive (Figure 3.9; Chapter 3). The current platypus genome assembly (OANA5) (Warren *et al.*, 2008) has a gap downstream from the *PIGT* gene where *ELP*/*CTI* is predicted to be located. However, the sequencing of the echidna genome as part of the Genome 10K Project (Koepfli *et al.*, 2015) may finally resolve the question as to whether monotremes have the *ELP*/*CTI* gene, or whether it evolved after the divergence of ancestral eutherians and marsupials from the monotreme ancestor.

It is possible that monotremes do not have an *ELP*/*CTI* gene, particularly if the gene evolved in viviparous species (therians) after the divergence of ancestral therians from prototherians. The duplication of the Kunitz domain has occurred many times over the past 500 MYA, as has its specialisation. For example, the trophoblast specific *TKDP1-5* genes, for which *CTI* is the ancestral gene, have rapidly evolved to become ruminant-specific genes that are expressed only in the placenta (Chakrabarty *et al.*, 2006a; Chakrabarty and Roberts, 2007; MacLean *et al.*, 2003; MacLean *et al.*, 2004). The absence of *ELP*/*CTI* in extant oviparous species, i.e., reptiles and birds (descendents of sauropsids) suggests that the ELP/CTI protein may not be required *in ovo* by the young. However, this can only be confirmed by the presence or absence of the ancestral *ELP*/*CTI* gene in extant monotremes (platypus and echidna, descendents of synapsids) and the expression of the gene in the monotreme egg and/or mammary gland.

The presence of *ELP*/*CTI* in species that predominantly transfer maternal immunoglobulins via milk suggests a possible protective role during early lactation. Neither the mode of maternal immunoglobulin transfer, nor the maturation of the puggle acquired immune system have been characterised (Miller and Hansen, 2014). However, maternal immunoglobulin transfer takes place *in ovo* for birds (Brambell, 1970; Kowalczyk *et al.*, 1985), fish (Swain and Nayak, 2009), reptiles (Schumacher *et al.*, 1999; Zimmerman *et al.*, 2010) and amphibians (Poorten and Kuhn, 2009), none of which have the *ELP*/*CTI* gene (Figure 3.4; Supplementary file 4.3; Chapter 3). Furthermore, maternal immunoglobulin transfer in the early-Synapsid mammalian ancestor, the Pelycosaurs, most likely occurred via the egg (Miller and Hansen, 2014).

The conservation of a functional, protein-coding *ELP* gene in all marsupials investigated thus far (species from 3 of the 7 marsupial orders were surveyed) (Figure 3.1; Supplementary file 4.5) (De Leo *et al.*, 2006; Nicholas *et al.*, 1997; Piotte and Grigor, 1996), suggests that ELP is vital to the maruspial young. In contrast, the loss of a functional *CTI* gene in most eutherians that have been studied (humans, mice and rats, Thoroughbred horse, elephant and others (Figures 3.4 and 3.5; Supplementary files 3.5 and 4.5) suggests that CTI has become redundant in these species. *CTI* gene loss may have occurred due to the transposition or deletion of regions within the gene by DNA transposons and/or endogenous retroviruses (Additional file 3.6) and/or mutations within the coding region of the gene (Supplementary files 3.5, 4.2 and 4.5).

Interestingly, a functional CTI protein is present in the land-based and aquatic mammals of the Laurasiatherian Orders Cetariodactyla, e.g. cow, sheep, camel, dolphin, whales, etc., and Carnivora including the dog, cat, Cape fur seal, walrus, Polar bear and others (Figure 3.1, Supplementary file 4.5). The exception is the Weddell seal (true seals; family Phocidae) which has a *CTI* pseudogene (Supplementary files 4.2 and 4.5). Whether all Phocids have a *CTI* pseudogene is unknown.

The conservation of the *ELP* and *CTI* genes and their respective exon sizes (Table 8.1; Supplementary file 4.4), plus a trend towards purifying selection within the 51 aa Kunitz BPTI 2 motif suggests that particular amino acid changes are not tolerated (Table 3.2) and that the Kunitz domain has an important role in ELP and CTI function. In contrast, the identification of bovine *CTI* as the ancestral gene of *PTI*, *STI* and the *TKDP1-5* genes highlighted the effects of gene and domain duplication and the role of postitive selection, whereby amino acids are free to change (Figure 3.6; Additional files 3.6, 3.7, 3.8, 3.9, 3.10, 3.11). The ubiquitous expression of *PTI* and *STI* but the trophoblast-specific expression of the *TKDP*s shows that gene regulatory mechanisms and protein function may also change due to the redundancy created by gene duplication. While PTI is a ubiquitous protease inhibitor (Ascenzi *et al.*, 2003), not all of the placenta-specific TKDPs have retained their protease inhibitor function (Chakrabarty *et al.*, 2006b; MacLean *et al.*, 2003). Furthermore, the conservation of *PTI*, *STI* and the *TKDP1-5* genes in ruminants suggests that they evolved after the divergence of ruminants ~25-35 MYA (Bininda-Emonds *et al.*, 2007).

Table 8.1. Summary of marsupial *ELP* and eutherian *CTI* gene structure

Gene	Gene size# (bp)	Exon 1 (bp)	Exon 2 (bp)	Exon 3 (bp)	Transcript size (CDS*)
ELP	1,387-4,818	73	216 210 (koala) 219 (Tas. devil)	142-158	291-315
CTI	2,513-3,283	76 64 (Cape fur seal & walrus)	216	124-136	291-315

#From the start of translation (ATG) to the end of the polyadenylation signal (AATAAA), inclusive; *CDS (coding DNA sequence) from the start to the stop codon inclusive

The ancestral gene that gave rise to *ELP* and *CTI* is unknown (this study). In particular, the very high conservation of nucleotides encoding the present-day Kunitz domains has made this analysis problematic (Elizabeth Pharo, unpublished data). This may change in future with improved genome sequence coverage sequencing and assembly. Hence the ancestral *ELP*/*CTI* gene may be identified by comparative genomics and retrotransposon analyses of the regions within and flanking the Kunitz domain encoding genes.

The regulation of tammar *ELP*

While the precise timing of *ELP* gene induction was unknown (Nicholas *et al.*, 1997; Simpson *et al.*, 1998), this study has shown that tammar *ELP* is induced at parturition (Figure 3.7) like the major milk-specific casein (*CSN1* and *CSN2*) and whey protein genes (*LALBA*, *LGB*) (Nicholas *et al.*, 1997; Sharp *et al.*, 2009). However, while *CSN1*, *CSN2*, *LALBA* and *LGB* are expressed throughout lactation, tammar *ELP* is down-regulated ~100-125 days pp (Nicholas *et al.*, 1997; Simpson, 1998; Elizabeth Pharo, unpublished data). In contrast, *WAP* is upregulated and expressed during Phase 2B-early Phase 3 and *LLPA* and *LLPB* are expressed only during late Phase 2B/Phase 3 and Phase 3 respectively (Simpson *et al.*, 2000; Trott *et al.*, 2005; Trott *et al.*, 2002).

The minimal, mammary gland-specific expression of *ELP* during pregnancy and its induction and high level expression only during the early lactation period (Figures 3.7 and 3.8) suggests that its regulation is complex (Chapter 5, Chapter 6). *ELP* regulatory mechanisms include lactogenic hormones (Figures 5.2, 5.2, 5.3, 5.4, 6.3; Supplementary files 6.9, 6.11, Table 8.3), autocrine and paracrine factors and mechanisms intrinsic to the mammary gland (Figures 5.5, 6.4, 6.5, 6.8; Supplementary files 5.4, 5.5, 6.12, 6.15), as shown using the 3D tammar mammary explant culture (Table 8.2) and 2D cell

culture systems. *ELP* expression is also controlled by the developmental stage of the mammary gland (Simpson, 1998) as for the *WAP* (Phase 2B) (Simpson *et al.*, 2000), *LLPA*, and *LLPB* genes (Trott *et al.*, 2005; Trott *et al.*, 2002). This explains the inability to induce *ELP* expression with lactogenic hormones in explants from mammary gland tissue in which the gene is not expressed *in vivo*, i.e. from mid- and late-lactation (Figures 5.5; Supplementary files 5.4, 5.5).

Table 8.2. Summary of *ELP* response^ to hormones *in vitro* in the 3D mammary explant culture system

HORMONE TREATMENT*	PREGNANCY	LACTATION		
	Phase 1	Phase 2A	Phase 2B	Phase 3
NH	- (5, 10 days)	- (8 days) (Elizabeth Pharo, preliminary data)	ND	ND
I	+ (4 days) ++ (8 days) ++++ (12 days)	+ (4 days) - (14 days)	- (4 days) - (8 days)	- (Elizabeth Pharo,; preliminary data)
I+HC	+ (4 days) ++ (8 days) ++++ (12 days)	+ (4 days) - (4 days I; 10 days I+HC)	- (4 days) - 4 days I; 4 days I+HC)	ND
PRL	++ (8 days) (Simpson, 1998)	+ (4 days)	- (4 days) - (4 days I, 4 days PRL)	ND
IHCPRL	+++ (4 days) +++++ (8, 12 days)	++ (4 days) ++ (4 days I; 10 days IHCPRL)	- (4 days) - (4 days I; 4 days IHCPRL)	- (Elizabeth Pharo, preliminary data)
IHCPRLT_3E_2	+++ (4 days) +++++ (8 days)	++ (4 days)	- (4 days) - (4 days I; 4 days IHCPRLT_3E_2)	ND
IHCPRLT_3E_2 IGF-IEGF	+++ (4 days) ++++ (8 days)	ND	- (4 days) - (4 days I; 4 days IHCPRLT_3E_2IGF-IEGF)	ND
IProgE_2EGF	ND	ND	- (3, 6, 10 days)	- (3, 6, 10 days)

^*ELP* expression level, +(low) to +++++ (high), - no expression detected by Northern analysis; *hormone treatment of Phase 1, 2A, 2B and 3 mammary explants unless stated otherwise, ND, not determined.

The maximal response of *ELP* to IHCPRL *in vitro* in mammary explants from Phase 1 (Figures 5.2, 5.3; Simpson *et al.*, 1998) and Phase 2A (Supplementary file 5.4) mammary gland tissue was characteristic to the lactogenic hormone-responsiveness of the tammar *LGB*, *LALBA*, *CSN1*, *CSN2*, *WAP*, *LLPA* and *LLPB* genes *in vitro* (Collet *et al.*, 1992; Nicholas *et al.*, 1997; Nicholas and Tyndale-Biscoe, 1985; Nicholas *et al.*, 1995; Simpson *et al.*, 2000; Trott, 1999; Trott *et al.*, 2002). However, unlike the major marsupial and eutherian milk protein genes (Qian and Zhao, 2014; Rosen *et al.*, 1999; Topper and Freeman, 1980), with the exception of rat *LALBA* which is responsive to I+HC (Nicholas *et al.*, 1981), *ELP* expression could be maintained for 4, 8 and 12 days *in vitro* in Phase 1 (pregnancy) explants supplemented with I and I+HC but without prolactin (Figures 5.2, 5.3). In contrast, when cultured without hormones for 5 and 10-day periods, neither *ELP* nor *LGB* expression could be maintained in explants *in vitro* (Figure 5.3). *ELP* and *LGB* were also suppressed by 8-day NH treatments in Phase 2A explants (Elizabeth Pharo, preliminary data).

Although explants harvested from Phases 2A, 2B and 3 were responsive to hormones, the high-level expression of *ELP*, *WAP* and *LGB in vivo* could generally not be maintained *in vitro* (Figures 5.4, 5.5; Supplementary files 5.4, 5.5; Elizabeth Pharo, preliminary data). This may reflect the inability to maintain the metabolic activity of the mammary gland tissue *in vitro* (Nicholas and Tyndale-Biscoe, 1985; Sheehy *et al.*, 2004) and the lack of milk secretion and hence milk removal *in vitro*.

The maximal induction of tammar *ELP* in Phase 1 explants supplemented with IHCPRL (Figure 5.2, 5.3; Simpson, 1998) and the maintenance of *ELP* in either I, I+HC, or IHCPRL suggests that insulin, glucocorticoids and prolactin act synergistically to activate *ELP*, similar to other milk protein genes (Groenen and van der Poel, 1994; Topper and Freeman, 1980; Vonderhaar and Ziska, 1989). These hormones may also indirectly influence *ELP* expression through elements downstream of the PI3K/AKT and/or MAPK/ERK signalling pathways (as discussed in Chapter 5), while changes in hormone-sensitivities during pregnancy and lactation, as for oxytocin (Lincoln and Renfree, 1981) may also affect gene expression. The PI3K/AKT pathway is involved in the synthesis of lactose, lipids and glucose transport (Oliver and Watson, 2013; Rudolph *et al.*, 2003), but most importantly, induces autocrine prolactin production in MECs, STAT5A activation, terminal differentiation and secretory differentiation via Akt1 (Chen *et al.*, 2012).

In the mammary gland, PRL signalling occurs via the PRLR (Figure 1.10; Chapter 1) (Hughes and Watson, 2012; Villarino *et al.*, 2015). Tyrosine phosphorylated STAT5A homodimers translocated to the nucleus bind to GAS motifs in milk protein gene promoters and activate transcription (Schmitt-Ney *et al.*, 1991; Watson *et al.*, 1991). Other STATs also form homo- and heterodimers, e.g. STAT5A-STAT5B (Liu *et al.*, 1996) and can bind to GAS motifs producing heterogeneous signalling (Schindler *et al.*, 2007; Villarino *et al.*, 2015). Although *ELP* was maximally responsive to lactogenic hormones, the lack of a conserved STAT5 consensus site in all *ELP* and *CTI* promoters (Figure 6.5; Supplementary file 6.15) was surprising, but may indicate that STAT5 sites located in distal and/or intragenic regions activate gene transcription.

The Phase 2A maturation of the GH-IGF-I growth axis in the pouch young (Menzies *et al.*, 2012) and GH/GHR signalling via STAT5B (Schindler *et al.*, 2007) in mammary parenchyma and stroma (Gallego *et al.*, 2001; Liu *et al.*, 1998) suggests that GH may be involved in the Phase 2A-specific expression of *ELP*. GH has a direct effect on milk protein gene expression (*CSN1S1*, *CSN1S2*, *CSN2*, *LALBA*) and lipid synthesis *in vitro* (Johnson *et al.*, 2010; Zhou *et al.*, 2008), but also acts indirectly via the insulin and IGF-I receptors (Forsyth *et al.*, 1999; Plath-Gabler *et al.*, 2001). Insulin is essential for eutherian milk protein gene transcription (Menzies *et al.*, 2010; Menzies *et al.*, 2009). Insulin signal transduction occurs via the binding of the insulin, IGF-I, or IGF-II ligands to the respective receptors, but cross-talk between ligands and receptors also occurs (Baumrucker and Erondu, 2000). Hence the maintenance of tammar *ELP* expression *in vitro* in Phase 1 insulin-treated explants *in vitro* (Figures 5.2, 5.3) may be due to the effects of either GH and/or insulin. While insulin acts via the PI3K/AKT pathway during bovine lactation (Lemay *et al.*, 2007), the mode of insulin action in the tammar mammary gland has yet to be characterised.

In contrast to the receptor-mediated signalling of PRL, GH, insulin, IGF-I and IGF-II, glucocorticoids enter MECs via diffusion (Groner, 2002; Oakley and Cidlowski, 2011). The intracellular GR is activated by the binding of glucocorticoids and the dissociation of heat shock proteins (hsp90) (Bamberger *et al.*, 1996; Gross and Cidlowski, 2008). The 'activated' GR is then phosphorylated and translocated to the nucleus where it binds to either GREs or ½GREs in milk protein promoters (Groner, 2002; Oakley and Cidlowski, 2011). The importance of glucocorticoids and PRL in the synergistic activation of milk protein genes has been demonstrated for *CSN2* (Doppler *et al.*, 2001; Lechner *et al.*, 1997; Schmitt-Ney *et al.*, 1991). The binding of STAT5, ½GREs and C/EBPs to the *CSN2* promoter are essential for expression in HC11s; the

latter critical for MEC-specific *CSN2* expression (Lechner *et al.*, 1997; Raught *et al.*, 1995). The synergistic response of tammar *ELP* to IHCPRL in explants (Figures 5.2, 5.3; Supplementary file 5.4) suggests that glucocorticoids are also important for *ELP* transcription. As for STAT5, the lack of conserved consensus GREs in the proximal *ELP* and *CTI* promoters (Figure 6.5; Supplementary file 6.15) was puzzling, but suggests that GR activation of *ELP* occurs through the indirect binding of GR to the tammar *ELP* promoter.

Thyroid hormones are essential for secretory activation, copious milk secretion and the expression of milk protein genes (Ziska *et al.*, 1988). In contrast, progesterone withdrawal is essential for eutherian (Hartmann *et al.*, 1973; Leung *et al.*, 2004; Neville and Morton, 2001), but not marsupial lactation (Neville *et al.*, 2002; Renfree, 1994; Ward and Renfree, 1984). Estrogen withdrawal may not be necessary for successful lactation (Hartmann *et al.*, 1973). Notably, only the addition of T_3 and E_2 to lactogenic hormones overrides the repression of the Phase 2B-specific tammar *WAP* gene *in vitro* in Phase 1 explants. In contrast, this treatment tends to reduce *ELP* expression in Phase 1 explants compared to lactogenic hormones alone (Figures 5.2, 5.3). Whether T_3, E_2 or both hormones affect *ELP* expression is unknown. Apart from the 'classical' receptor-mediated signalling, estrogen can also act via 'non-classical' pathways through kinases and crosstalk between AP1, NF-B and SP1, with the orphan estrogen-related receptors (ERRs) also involved (Jarzabek *et al.*, 2009). The identification of two conserved ERRα/β-like TFBS in the tammar *ELP* and eutherian *CTI* promoters (Figure 6.5; Supplementary file 6.15) suggests that these genes may be activated or repressed by estrogen and warrants further investigation.

The identification of the two tammar *ELP* TSSs and proximal promoter(s) (Figures 6.1, 6.2; Supplementary file 6), the temporal, mammary gland-specific expression of *ELP* and *CTI* (Pharo *et al.*, 2012; Simpson *et al.*, 1998; Veselsky *et al.*, 1978) and the conservation of putative MAF (ETS), AP2γ, C/EBP, NF-κB, ERRα/β TFBS within the *ELP* and *CTI* proximal promoters (Figure 6.5; Supplementary file 6.15) suggests these transcription factors play a major role in the hormone-independent transcription of *ELP* (Figure 6.4) and *CTI* during Phase 2A and colostrum production respectively. The strict control of *ELP* (Chapter 5, Chapter 6) and other Phase-specific milk protein genes (Simpson, 1998; Trott, 1999) suggests that the 3D organisation of the genome may control tammar mammary gland gene expression, as in other species (Devinoy and Rijnkels, 2010; Kress *et al.*, 2010; Lemay *et al.*, 2013; Rijnkels *et al.*, 2013; Rijnkels *et al.*, 2010). Therefore, *ELP* and *CTI* expression may be controlled by local and/or global

epigenetic modifications (histone and DNA methylation and histone acetylation) (Barutcu *et al.*, 2016; Dixon *et al.*, 2012) involving enhancers or silencers tens of kilobases and up to several megabases away, or located on other chromosome(s) (Barutcu *et al.*, 2016; Dekker, 2008; Dixon *et al.*, 2012).

During pregnancy, the minimal expression of *ELP* (Figure 3.7) is most likely due to the binding of NF-κB and/or an ETS repressor element to the proximal promoter (Chapter 6; Figure 6.8), plus the bivalent nature of chromatin. The hormone-independent activation of *ELP* at parturition and high level expression during Phase 2A only is promoted by MAF, C/EBP and/or AP2γ, facilitated by "open" chromatin. In contrast, "closed" chromatin may block transcription factor access to the *ELP* gene and suppress transcription in the Phase 2A, Phase 2B, virgin, non-pregnant and involuting mammary glands, as well as in all other tissues (Figure 6.8; Chapter 6). These hypotheses could be tested using DNase I assays (Schmitt-Ney *et al.*, 1991), electrophoretic mobility shift assays (EMSAs), supershift assays (Dey *et al.*, 2012; Hellman and Fried, 2007). Global regulatory mechanisms could be investigated with chromatin conformation capture (Barutcu *et al.*, 2016; Cattoni *et al.*, 2015), chromatin immunoprecipitation and sequencing of bound DNA (ChIP-Seq) (Hoffman and Jones, 2009) and methyl-DNA immunoprecipitation followed by next generation sequencing to map DNA methylation (MeDIP-Seq) (Pomraning *et al.*, 2009). RNA-Seq (Wang *et al.*, 2009) would also be invaluable for determining genes with a Phase 2A-specifc expression and whether non-coding RNAs affect *ELP* and other tammar milk protein genes.

Local control mechanisms play a vital role in the tammar mammary gland (Chapter 5), as shown by the inability to induce *ELP* expression with lactogenic hormones in Phase 2B mammary explants, primary cells isolated from the Phase 1 tammar mammary gland and the lack of activity of *ELP* promoter reporter constructs *in vitro*. This is also evident in the tammar *in vivo*. Although all four mammary glands prepare for lactation (Findlay, 1982; Stewart, 1984), only the sucked gland continues to produce milk (Sharman, 1962; Stewart, 1984; Tyndale-Biscoe and Renfree, 1987) and so *ELP* expression only occurs in that gland (Chapter 3). Consequently, the three non-sucked glands enter lactational quiescence in preparation for the next pregnancy or oestrous cycle post-lactation (Findlay, 1982; Sharman, 1962) and *ELP* is down-regulated in them (Chapter 3) in the tammar, and in the brushtail possum (Demmer *et al.*, 1998).

The concurrent asynchronous lactation phenomenon (Griffiths *et al.*, 1972; Nicholas, 1988) also demonstrates the strict control and independence of the four tammar mammary glands. It allows the marsupial mother to simultaneously produce two different milks in two different glands, for a young at foot and neonate in the pouch (Lemon and Barker, 1967; Lincoln and Renfree, 1981; Sharman and Calaby, 1964). Hence, the mother produces low volumes of dilute Phase 2A milk which is high in oligosaccharides and lipids (and ELP) in one gland and high volumes of Phase 3 milk which is high in monosaccharides and proteins (and LLPA and LLPB), but has a low lipid content (Green and Merchant, 1988; Simpson, 1998; Trott, 1999). While oxytocin sensitivity is different in the two glands (Lincoln and Renfree, 1981), a clock mechanism(s) within each gland may control the progress of lactation and the differential expression and secretion of milk constituents. This may explain the inability to induce the temporally expressed *ELP*, *WAP*, *LLPA* and *LLPB* genes *in vitro* with lactogenic hormones in tissues in which these genes are not expressed *in vivo* (this study; Simpson, 1998; Trott, 1999); with the exception of *WAP* induction in Phase 1 IHCPRLT_3E_2-treated explants (Simpson *et al.*, 2000).

Cross-talk between the stroma (ECM proteins, adipocytes, fibroblasts, endothelial and innate immune cells (Muschler and Streuli, 2010)) and parenchyma (alveoli and ducts consisting of MECs and myoepithelial cells) is essential to mammary gland function (Bonnans *et al.*, 2014; Gjorevski and Nelson, 2011), and the expression of *ELP* (Chapter 5) and other Phase-specific tammar milk protein genes (Wanyonyi *et al.*, 2013a; Wanyonyi *et al.*, 2013b). It also explains the difference in *ELP* expression and/or *ELP* promoter reporter activity in 3D tissue architecture (Phase 1 and 2A mammary gland explants; Chapter 5) versus 2Dn cell culture (primary cells harvested from the mid- and late-pregnancy tammar mammary gland, or murine HC11 and KIM2 cell lines; Chapter 6).

CTI, like *ELP* may also be under epigenetic control, because in the eutherians investigated thus far, the gene is only expressed for the first 1-2 days pp during the period of colostrum secretion (Laskowski and Laskowski, 1951; Veselsky *et al.*, 1978). The prolonged, intermittent expression of otariid *CTI* for at least 2-3 months pp (Figure 4.3) provides an intriguing exception. As mentioned previously (Chapter 6), both of the at-sea lactating Cape fur seal females exhibited no signs of involution in their mammary glands as shown by histological analysis (Cane, 2005), indicative of lactation. Hence, *CTI* expression may be cyclical, i.e. up-regulated on shore and down-regulated at-sea. Further research should address this question. It is tempting to speculate that the

periodic expression of otariid *CTI* on-shore and the >80% reduction in milk production as the fur seal forages at seal (Arnould and Boyd, 1995) are due to changes in energy-partitioning. During lactation, thyroid hormones (and deiodinases) regulate the partitioning of metabolic activity in the mammary gland by controlling the conversion of T_4 (thyroxine) to the more-active T_3 (triiodothyronine) (Capuco *et al.*, 1999; Jack *et al.*, 1994; Kahl *et al.*, 1991). Therefore, mammary gland activity increases, directing maternal resources to milk secretion for the young (Jack *et al.*, 1994). In contrast, activity is reduced in peripheral maternal tissues. In contrast, as the otariid mother goes to sea to feed, metabolic activity and milk production in the mammary gland are reduced and resources are directed to the accumulation of fats and proteins in maternal tissues (Cane, 2005; Sharp *et al.*, 2005). However, when the otariid mother returns to shore, metabolic energy is once more redirected to the mammary gland and the secretion of milk for her pup (Arnould and Boyd, 1995).

Marsupial ELP and eutherian CTI: putative protease inhibitors

The presence of BPTI Kunitz domain motifs 1 and 2 within maruspial ELP and eutherian CTI (Figures 3.2, 4.4; Supplementary files 4.11, 4.12) (Cechova *et al.*, 1969; Piotte and Grigor, 1996; Simpson *et al.*, 1998) suggests that like bovine PTI (Huber and Bode, 1978) (Figure 1.6) and CTI (Wagner *et al.*, 1978), Cape fur seal CTI (Figure 4.5) and tammar ELP (Simpson *et al.*, 1998) form a globular, pear-shaped protein stabilised by three disulphide bonds. Structural homology modelling suggests that Cape fur seal CTI inhibits walrus (and hence Cape fur seal) trypsin in a 1:1 canonical mechanism (Figure 4.5) like the TFPI KD2-porcine trypsin (Burgering *et al.*, 1997) and BPTI-bovine trypsin complexes (Bode *et al.*, 1984). As for BPTI, the archetypal Kunitz family member (Rawlings *et al.*, 2014), Cape fur seal CTI has a basic P_1 'warhead' residue that is inserted into the S_1 cleft of the serine protease, thus inhibiting trypsin (Figure 4.5).

The lack of a basic P_1 residue in marsupial ELP, with the P_1 residue type not conserved (Figures 3.2, 4.4; Supplementary file 4.11) suggests that these proteins are inactive, or have only weak activity against trypsin. However, tammar, koala and dunnart ELP have a small uncharged P_1 residue (e.g. Ser, Asn) and so may inhibit elastase-like enzymes, e.g. neutrophil elastase (Laskowski and Kato, 1980). In contrast, possum and opossum ELP with a negatively charged P_1 residue (Asp), (which is rare amongst Kunitz domains (Benson *et al.*, 2015; Cunningham *et al.*, 2015)), may, like

TKDP2, possess minimal inhibitory activity (MacLean *et al.*, 2004). Alternatively, the protease reactive site may be located elsewhere in these proteins, as for the tick anticoagulant peptide which inhibits Factor Xa (Jordan *et al.*, 1992). While bovine CTI inhibits trypsin and plasmin *in vitro* (Feeney *et al.*, 1969; Laskowski and Laskowski, 1951), the inhibitory potential of marsupial ELP is yet to be determined, primarily due to the very small quantities of ELP secreted in early lactation tammar milk (Simpson, 1998).

Milk is the end-product of a long process of evolution, having been transformed from a non- or minimal nutritive secretion to a major conveyer of nutrients to the young (Oftedal, 2013). Therefore, many genes have been co-opted into lactation. Some genes have gained, or altered function(s), whilst others have become pseudogenes (Chapter 3). For example, the *LGB* and *WAP* genes have become non-functional in humans, but remain functional in many other mammals. Hence, it is quite possible that ELP and/or CTI may have lost their protease inhibitor activity and may now have primarily a non-essential, nutritional role, for example as a source of sulphur-containing amino acids for the mammalian young.

The production of recombinant tammar ELP *in vitro* in both mammalian and bacterial expression systems proved to be challenging and problematic (Figures 7.3-7.4, 7.6, 7.8, 7.10, 7.12-7.17 inclusive; Chapter 7), particularly without an antibody to detect tammar *ELP*. In contrast to cystatin C, the absence of visually detectable recombinant ELP and/or CTI proteins in mammalian cells (Figures 7.3, 7.4), despite the expression of *ELP* and *CTI in vitro* (Appendix VI) suggests that post-transcriptional mechanisms may influence *in vitro* protein production.

Although the pMAL bacterial expression system permitted the production of purified recombinant tammar ELP (Figure 7.16), the protein may be misfolded and biologically inactive, due to the lack of disulphide bond formation in the *E. coli* cytoplasm (Baneyx and Mujacic, 2004). Furthermore, native tammar ELP is predominantly secreted in milk as at least 7 different ~16-20.5 kDa N-glycosylated proteins (Joss *et al.*, 2007; Simpson *et al.*, 1998). An ~8 kDa non-glycosylated form is also secreted in brushtail possum milk (Piotte and Grigor, 1996). As the nature of attached sugar(s) affects protein stability, solubility and bioactivity (Hmiel *et al.*, 2015; Roth *et al.*, 2010), each ELP glycoprotein and indeed each N-linked sugar group may have different biological function(s). Additionally, as N-glycosylation is species- and cell type-specific (Devasahayam, 2007; Jacobs and Callewaert, 2009), the production of recombinant

tammar ELP in either immortalised or stably-transfected tammar MECs using the mammalian FLAG vector would provide the best system for future production (Brizzard *et al.*, 1994; Chubet and Brizzard, 1996).. The advantages include the production, detection and purification of correctly folded, N-glycosylated, biologically active tELP-FLAG proteins.

The MEC-specific expression of bovine *CTI* and its secretion in colostrum during the first 1-2 days pp when the calf lacks an acquired immune system (Laskowski and Laskowski, 1951; Veselsky *et al.*, 1978) mirrors the immune status of the marsupial PY during *ELP* expression and secretion (Deane and Cooper, 1988; Edwards *et al.*, 2012; Joss *et al.*, 2007). Therefore, as suggested by Laskowski and Laskowski (1951), CTI (and ELP) may prevent the degradation of immunoglobulins that are transferred from mother to young in colostrum and early milk respectively.

The low levels of trypsin, pepsin and chymotrypsin in the tammar PY gut until ~175 days postpartum (Griffiths and Barton, 1966; Janssens and Ternouth, 1987) suggests that tammar ELP may act in the mother's mammary gland. It is interesting to speculate that tammar ELP may prevent the remodelling and premature development of the mammary gland during Phase 2A. Hence, maternal milk production is programmed during this period to provide nutrients appropriate for development, rather than growth of the altricial young *per se.* The tammar mammary gland (and the PY) grows exponentially up until ~210 days pp (early Phase 3) (Stewart, 1984). Consequently, during the first 100 days pp, mammary gland weight increases ~2.5-fold from ~1 g at parturition to ~2.5 g at 100 days pp at the end of Phase 2A (Stewart, 1984). Intriguingly, after *ELP* is downregulated *in vivo*, mammary gland weight increases ~8-fold to ~30 g over the next 110 days (~day 210 pp, early Phase 3) (Stewart, 1984). The expression of *WAP* (Simpson *et al.*, 2000), an *in vitro* growth promoter (Topcic *et al.*, 2009) during this period may cause the significant increase in both mammary gland and PY weight (Stewart, 1984). ELP may also protect the tammar mammary gland against neutrophil elastase which is released from neutrophils. Although neutrophil elastase acts against enterobacteria, e.g *E. coli*, *Shigella*, *Salmonella*, *Yersinia* (Weinrauch *et al.*, 2002), *Pseudomonas aeruginosa* (Hirche *et al.*, 2008), it also degrades elastin and collagen IV within ECM (Alexander and Werb, 1991). Hence ELP in milk and the mammary gland may inhibit ECM degradation and/or remodelling, both in the mammary gland and in the gut of the young.

Oligosaccharides attached to tammar ELP may play a protective role in the tammar

PY gut as milk oligosaccharides provide the first line of defence against pathogens, preventing gastric and urinary infections for the suckling young (Eiwegger *et al.*, 2010; Morrow *et al.*, 2005). In the gut, oligosaccharides can act as decoys for pathogenic bacteria by forming soluble receptor analogues (Zopf and Roth, 1996). This prevents the binding of pathogenic microbes and their toxins to receptors on the intestinal epithelium, improving neonate health and survival (Kunz *et al.*, 2014; Newburg and Walker, 2007; Pacheco *et al.*, 2015). Notably, different glycoconjugates, i.e., glycolipids, glycoproteins, glycosaminoglycans, and mucins exhibit different activities against pathogenic bacteria and viruses (Dai *et al.*, 2000; Newburg, 1999). Therefore, the heterogeneity of milk sugars maximises neonate protection against a diverse range of microorganisms (Bode, 2006; Dallas *et al.*, 2014; Newburg, 2009).

Milk oligosaccharides also act as prebiotics (Newburg, 2013; Walker, 2013). They promote nutrient digestion and absorption by influencing the composition of gut microflora and promoting gut colonisation by beneficial bacteria (Barile and Rastall, 2013; Smilowitz *et al.*, 2014). Oligosaccharides also affect gastrointestinal development, proliferation and differentiation (Kuntz *et al.*, 2009). Interestingly, human milk oligosaccharides inhibit intestinal cell growth *in vitro* (Kuntz *et al.*, 2009; Kuntz *et al.*, 2008). Hence the oligosaccharides in Phase 2A and 2B tammar milk may maintain the undifferentiated state of the PY gut during the first ~170 days pp when the gut mucosa consists of immature gastric glandular cells permeable to macromolecules (Kwek *et al.*, 2009).

The Evolution of Lactation

The mammary glands and lactation presumably evolved ~310 MYA in synapsids, the lineage ancestral to mammals (Figure 1.1) (Lefevre *et al.*, 2010; Oftedal, 2002; Oftedal and Dhouailly, 2013). Ancestral mammaliaforms then experienced an intense period of extensive morphological and geographical diversification in the early-mid Jurassic (Close *et al.*, 2015; Luo *et al.*, 2015; Meng, 2014). Hence, the present-day monotremes, marsupials and eutherians inhabit terrestrial and/or aquatic environments in polar, temperate and tropical climates and evolved a diverse range of reproductive strategies, including specialised mammary gland anatomy and physiology, milk composition and milk constituents (Andres and Djonov, 2010; Lefèvre *et al.*, 2010; Oftedal, 2013).

The precise characteristics of ancestral ‘milk’ are unknown, but would have been determined by the original ancestral gland (Blackburn, 1991; Blackburn *et al.*, 1989).

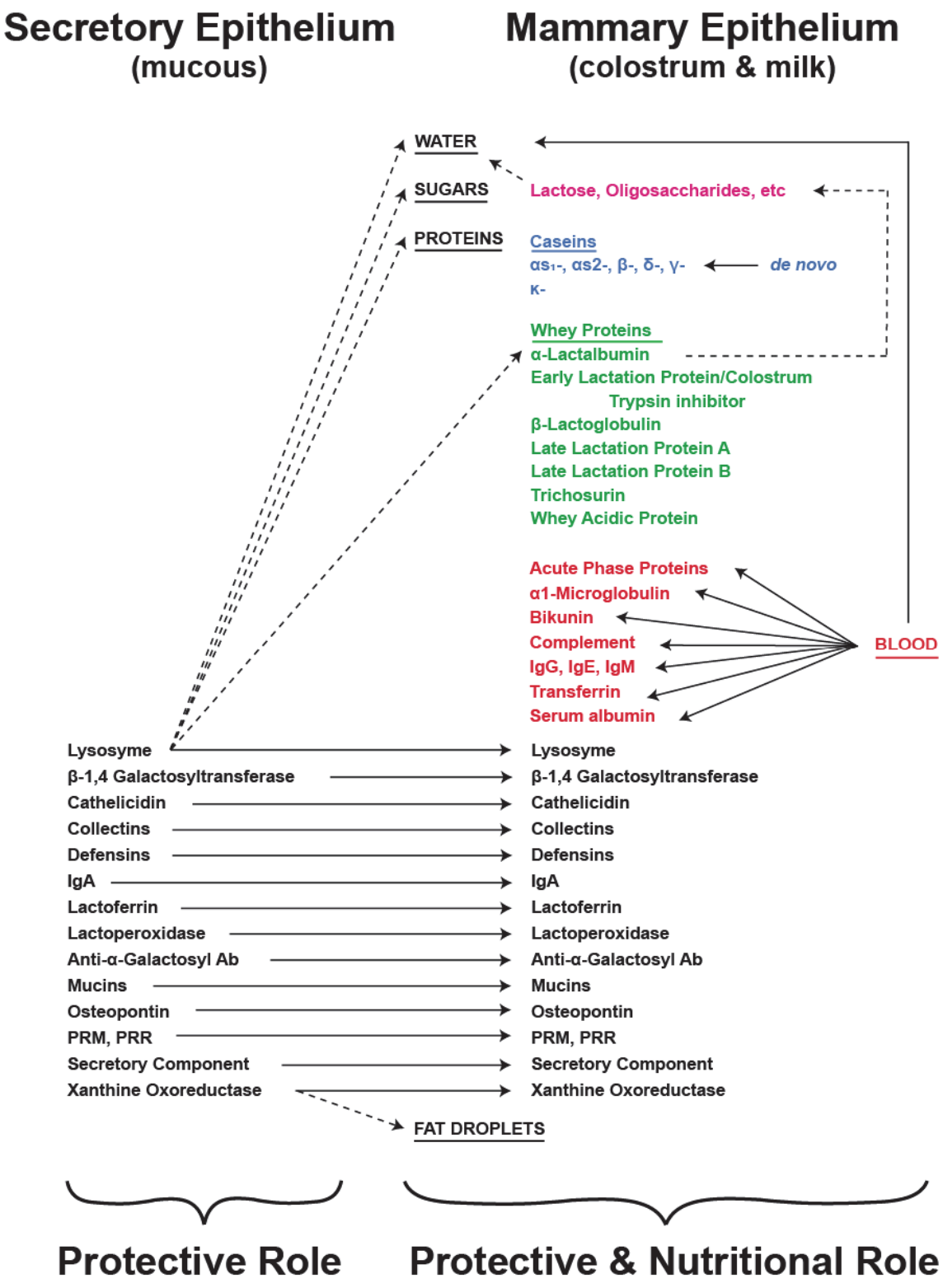
An eccrine gland origin suggests that ancestral milk was a glandular skin secretion of water and electrolytes which prevented the desiccation of the mother's egg(s) (Blackburn *et al.*, 1989; Oftedal, 2002). Hence, proteins, lipids and sugars were later co-opted into milk (Oftedal, 2012), producing a vital source of nutrition for the young (Blackburn *et al.*, 1989; Vorbach *et al.*, 2006). In contrast, for a mammary gland of epitrichial (sebaceous/apocrine) origin, ancestral milk most probably included antimicrobials as well as carbohydrates, fats and proteins (Blackburn *et al.*, 1989). Notably, recent evolution of development studies have provided strong evidence that mammary secretory tissue originated from an apocrine-like gland (Oftedal and Dhouailly 2013). This is consistent with the conservation of constituents produced by secretory epithelia (mucous formed in an inflammatory response) and mammary epithelia (milk secreted during lactation) in present-day milk (Blackburn, 1991; Blackburn *et al.*, 1989; Wheeler *et al.*, 2007). Furthermore, given the vulnerability of a nutrient-rich secretion like milk to microbial colonisation, it is not surprising that both components of the innate and adaptive immune systems are involved in lactation and protect both the mammary gland, milk and/or neonate from pathogens (Olav Oftedal, personal communication).

The innate immune system provides the non-specific, first line of defence against pathogens (Akira *et al.*, 2001; Stelwagen *et al.*, 2009). Innate immune system components include lysosyme, lactoferrin, cathelicidin, lactoperoxidase, IgA, xanthine oxoreductase, etc (Figure 8.1). Hematopoietic and non-hematopoietic cells of the innate immune system are also secreted into milk, including neutrophils, macrophages, natural killer cells, mast cells, eosinophils, basophils and dendritic cells (Hassiotou and Geddes, 2015; Hassiotou *et al.*, 2013; Turvey and Broide, 2010; Young *et al.*, 1997).

Unlike the innate immune system, the adaptive immune system which evolved ~500 MYA provides antibody-mediated protection against specific pathogens (Bayne, 2003; Flajnik and Kasahara, 2010; Travis, 2009). Interestingly, the acquired immune system is activated by the innate immune system (Iwasaki and Medzhitov, 2010; Iwasaki and Medzhitov, 2015). Dendritic cells of the innate immune system have a critical role in microbial detection. They have pattern-recognition molecules and receptors (PRMs, PRRs), e.g. Toll-like receptors (TLRs), on their surface which recognise pattern-associated molecular patterns (PAMPs) of bacterial and fungal cell walls (Iwasaki and Medzhitov, 2015; Vorbach *et al.*, 2006). PAMP recognition by PRRs leads to the activation of antigen-presenting cells, T and/or B cells and hence the acquired immune system activation (Akira *et al.*, 2006; Brubaker *et al.*, 2015; Kumagai and Akira, 2010).

Figure 8.1. Similarities between products of the secretory epithelium (mucous) and mammary epithelium (colostrum and milk)

Products of the secretory epithelium generally play a protective role in the body. Many of these constituents are also expressed in the mammary epithelium. In the mammary gland they may function as immunological agents, whilst for the young, these secreted products may act both as antimicrobial compounds and have a nutritive function in colostrum and/or milk. Some milk components, such as lactose, caseins and whey proteins are synthesized specifically by the mammary gland, whilst others are transported from the vascular system (red text) into colostrum and/or milk. Whilst α-LA evolved from lysosyme (Brew *et al.*, 1967, 1968; Prager and Wilson, 1988), the ancestral genes of other whey proteins have yet to be determined. Adapted from (Vorbach *et al.*, 2006).


Secretory Epithelium
(mucous)
Mammary Epithelium
(colostrum & milk)
WATER
SUGARS
PROTEINS
Lactose, Oligosaccharides, etc
Caseins
αs1-, αs2-, β-, δ-, γ-
κ-
de novo
Whey Proteins
α-Lactalbumin
Early Lactation Protein/Colostrum
Trypsin inhibitor
β-Lactoglobulin
Late Lactation Protein A
Late Lactation Protein B
Trichosurin
Whey Acidic Protein
Acute Phase Proteins
α1-Microglobulin
Bikunin
Complement
IgG, IgE, IgM
Transferrin
Serum albumin
BLOOD
Lysosyme
β-1,4 Galactosyltransferase
Cathelicidin
Collectins
Defensins
IgA
Lactoferrin
Lactoperoxidase
Anti-α-Galactosyl Ab
Mucins
Osteopontin
PRM, PRR
Secretory Component
Xanthine Oxoreductase
FAT DROPLETS
Protective Role
Protective & Nutritional Role

Adaptive immune system components include lymphocytes (T-cells and B-cells) which develop in the thymus and bone marrow respectively, the T cell receptor, B cell receptor (immunoglobulins), the major histocompatibility complex (MHC), T-helper type 1 (Th1) and Th2 cells (Bartl *et al.*, 2003; Flajnik and Kasahara, 2001; Goldman, 2002; Hirano *et al.*, 2011).

Immune cells also affect mammary epithelial cell fate and function (Khaled *et al.*, 2007; Watson, 2009). MECs secrete different T-cell regulatory cytokines in the undifferentiated and differentiated glands. In undifferentiated cells, Th1 cytokines such as IFNγ, TNFα and IL-12α, which promote cellular immunity, are secreted (Goldman, 2002; Khaled *et al.*, 2007). However, when treated with lactogenic hormones (insulin, glucocorticoids and prolactin), MECs differentiate and switch to secreting (Th2) cytokines such as IL-4, IL-5 and IL-13 (Khaled *et al.*, 2007), promoting the development of antibody-mediated immunity in the eutherian neonate (Goldman, 2002). The inability of the tammar pouch young to mount a specific immune response during Phase 2A and the maturation of the PY immune system only by the end of early lactation suggests that the maternal mammary gland secretes Th2 cytokines during this period.

The present-day mammary gland has extremely complex host-defence systems with the innate and acquired immune systems co-ordinated to protect both mother and young (Blackburn, 1991; Goldman *et al.*, 1990; Peaker, 2002; Stelwagen *et al.*, 2009; Wang *et al.*, 2011). Most importantly, two-way signalling occurs between the mother and her young via the 'bacterial entero-mammary pathway'. Hence antibodies from the maternal gut are translocated to the mammary gland and then to the neonate via milk (Jost *et al.*, 2014; Kleinman and Walker, 1979; Salmon *et al.*, 2009). Furthermore, if the mother ingests a pathogen, she can make antibodies to it and transfer these to the young via milk (Cowie, 1972). Additionally, the nursing mother can secrete antibodies in milk in response to a specific infection in the young (Hassiotou and Geddes, 2015). This immune system feedback process almost certainly occurs in the tammar via inducing defecation and urination by licking and coprophagy, which the mother practises for up to ~190 days postpartum (Tyndale-Biscoe and Renfree, 1987). This process would be critical to the survival of the young, particularly during early lactation when the altricial PY lacks antibody-mediated immunity.

General Conclusions

Marsupial lactation is the most complex in the mammalian lineage. The maternal control over growth and development of the pouch young is mediated entirely via the temporal changes in the composition of the milk and must have evolved in the last common ancestor of mammals before the divergence of monotremes and therians, since monotremes also have a changing milk composition. This study, focussing on one of the key proteins, namely marsupial *Early Lactation Protein* sheds light on its co-evolution with the eutherian *Colostrum Trypsin Inhibitor* gene from a common ancestral gene over 160 million years ago. The secretion of these two proteins that are putative Kunitz protease inhibitors during a period of immunoglobulin transfer from mother to young via milk suggests that they play a vital role in neonate health and pouch young survival. *ELP* regulation is complex, involving lactogenic hormones, transcription factors and mechanisms intrinsic to the gland. The timing of its secretion correlates with the most vulnerable period of pouch young life in the first third of lactation and thus underpins the success of the marsupial mode of reproduction.

Bibliography

Adamski, F.M., and Demmer, J. (2000). Immunological protection of the vulnerable marsupial pouch young: two periods of immune transfer during lactation in *Trichosurus vulpecula* (brushtail possum). *Developmental and Comparative Immunology* **24**: 491-502.

Aggeler, J., Park, C.S., and Bissell, M.J. (1988). Regulation of milk protein and basement membrane gene expression: the influence of the extracellular matrix. *J Dairy Sci* **71**: 2830-2842.

Åkerström, B., Flower, D.R., and Salier, J.P. (2000). Lipocalins: unity in diversity. *Biochim Biophys Acta* **1482**: 1-8.

Akira, S. (1999). Functional roles of STAT family proteins: lessons from knockout mice. *Stem Cells* **17**: 138-146.

Akira, S., Takeda, K., and Kaisho, T. (2001). Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Immunol* **2**: 675-680.

Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen Recognition and Innate Immunity. *Cell* **124**: 783-801.

Alexander, C., and Werb, Z. (1991). Extracellular Matrix Degradation. In: Cell Biology of Extracellular Matrix, E. Hay, ed. (Springer US), pp. 255-302.

Ali, S., Edery, M., Pellegrini, I., Lesueur, L., Paly, J., Djiane, J., and Kelly, P.A. (1992). The Nb2 form of prolactin receptor is able to activate a milk protein gene promoter. *Mol Endocrinol* **6**: 1242-1248.

Ali, S., Pellegrini, I., and Kelly, P.A. (1991). A prolactin-dependent immune cell line (Nb2) expresses a mutant form of prolactin receptor. *J Biol Chem* **266**: 20110-20117.

Allen, J.C., Keller, R.P., Archer, P., and Neville, M.C. (1991). Studies in human lactation: milk composition and daily secretion rates of macronutrients in the first year of lactation. *Am J Clin Nutr* **54**: 69-80.

Almo, S.C., and Love, J.D. (2014). Better and faster: improvements and optimization for mammalian recombinant protein production. *Current Opinion in Structural Biology* **26**: 39-43.

Altiok, S., and Groner, B. (1994). β-Casein mRNA sequesters a single-stranded nucleic acid-binding protein which negatively regulates the β-casein gene promoter. *Mol Cell Biol* **14**: 6004-6012.

Ambatipudi, K., Joss, J., and Deane, E. (2007). A comparative proteomic analysis of skin secretions of the tammar wallaby (*Macropus eugenii*) and the wombat (*Vombatus ursinus*). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **2**: 322-331.

Ambatipudi, K., Joss, J., Raftery, M., and Deane, E. (2008). A proteomic approach to analysis of antimicrobial activity in marsupial pouch secretions. *Dev Comp Immunol* **32**: 108-120.

Amsterdam, A., Tajima, K., and Sasson, R. (2002). Cell-specific regulation of apoptosis by glucocorticoids: implication to their anti-inflammatory action. *Biochem Pharmacol* **64**: 843-850.

Anderson, S., Rudolph, M.C., McManaman, J.L., and Neville, M. (2007). Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis! *Breast Cancer Research* **9**: 204.

Andres, A.C., and Djonov, V. (2010). The mammary gland vasculature revisited. *J Mammary Gland Biol Neoplasia*.

Aoki, N. (2006). Regulation and functional relevance of milk fat globules and their components in the mammary gland. *Biosci Biotechnol Biochem* **70**: 2019-2027.

Arnould, J.P., and Hindell, M.A. (1999). The composition of Australian fur seal (Arctocephalus pusillus doriferus) milk throughout lactation. *Physiol Biochem Zool* **72**: 605-612.

Arnould, J.P.Y., and Boyd, I.L. (1995a). Inter- and intra-annual variation in milk composition in Antarctic fur seals (*Arctocephalus gazella*). *Physiol Zool* **68**: 1164–1180.

Arnould, J.P.Y., and Boyd, I.L. (1995b). Temporal patterns of milk production in Antarctic fur seals (*Arctocephalus gazella*). *J Zool Lond* **237**: 1-12.

Arya, R., Sabir, J.M., Bora, R., and Saini, K. (2015). Optimization of culture parameters and novel strategies to improve protein solubility. In: Insoluble Proteins, E. García-Fruitós, ed. (Springer New York), pp. 45-63.

Ascenzi, P., Bocedi, A., Bolognesi, M., Spallarossa, A., Coletta, M., De Cristofaro, R., and Menegatti, E. (2003). The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor): a milestone protein. *Curr Protein Pept Sci* **4**: 231-251.

Ashraf, S.S., Benson, R.E., Payne, E.S., Halbleib, C.M., and Gron, H. (2004). A novel multi-affinity tag system to produce high levels of soluble and biotinylated proteins in Escherichia coli. *Protein Expression and Purification* **33**: 238-245.

Ashworth, U.S., Ramaiah, G.D., and Keyes, M.C. (1966). Species difference in the composition of milk with special reference to the Northern fur seal. *J Dairy Sci* **49**: 1206-1211.

Baintner, K., and Csapo, J. (1996). Lack of acid-resistant trypsin inhibitor in mare's colostrum: short communication. *Acta Vet Hung* **44**: 95-97.

Baintner, K. (2007). Transmission of antibodies from mother to young: Evolutionary strategies in a proteolytic environment. *Vet Immunol Immunopathol* **117**: 153-161.

Baintner, K., Jr. (1973). The physiological role of colostral trypsin inhibitor: experiments with piglets and kittens. *Acta Vet Acad Sci Hung* **23**: 247-260.

Baintner, K., Jr. (1976). The physiological role of colostral trypsin inhibitor. II. Experiments with lambs. *Acta Vet Acad Sci Hung* **26**: 405-413.

Ball, R.K., Friis, R.R., Schoenenberger, C.A., Doppler, W., and Groner, B. (1988). Prolactin regulation of β-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *EMBO J* **7**: 2089-2095.

Ballester, M., Kress, C., Hue-Beauvais, C., Kieu, K., Lehmann, G., Adenot, P., and Devinoy, E. (2008). The nuclear localization of WAP and CSN genes is modified by lactogenic hormones in HC11 cells. *J Cell Biochem* **105**: 262-270.

Bamberger, C.M., Schulte, H.M., and Chrousos, G.P. (1996). Molecular Determinants of Glucocorticoid Receptor Function and Tissue Sensitivity to Glucocorticoids. *Endocrine Reviews* **17**: 245-261.

Bandaranayake, A.D., and Almo, S.C. (2014). Recent advances in mammalian protein production. *Febs Letters* **588**: 253-260.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current opinion in biotechnology* **10**: 411-421.

Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotech* **22**: 1399-1408.

Barile, D., and Rastall, R.A. (2013). Human milk and related oligosaccharides as prebiotics. *Current opinion in biotechnology* **24**: 214-219.

Bartl, S., Baish, M., Weissman, I.L., and Diaz, M. (2003). Did the molecules of adaptive immunity evolve from the innate immune system? *Integr Comp Biol* **43**: 338-346.

Barutcu, A.R., Fritz, A.J., Zaidi, S.K., van Wijnen, A.J., Lian, J.B., Stein, J.L., Nickerson, J.A., Imbalzano, A.N., and Stein, G.S. (2016). C-ing the Genome: A Compendium of Chromosome Conformation Capture Methods to Study Higher-Order Chromatin Organization. *J Cell Physiol* **231**: 31-35.

Basden, K., Cooper, D.W., and Deane, E.M. (1997). Development of the lymphoid tissues of the tammar wallaby *Macropus eugenii*. *Reprod Fertil Dev* **9**: 243-254.

Bassam, B.J., Caetano-Anolles, G., and Gresshoff, P.M. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* **196**: 80-83.

Bassuk, A.G., Anandappa, R.T., and Leiden, J.M. (1997). Physical interactions between Ets and NF-κB/NFAT proteins play an important role in their cooperative activation of the human immunodeficiency virus enhancer in T cells. *J Virol* **71**: 3563-3573.

Bathgate, R.A., Parry, L.J., Fletcher, T.P., Shaw, G., Renfree, M.B., Gemmell, R.T., and Sernia, C. (1995). Comparative aspects of oxytocin-like hormones in marsupials. *Adv Exp Med Biol* **395**: 639-655.

Baudinette, R.V., Boontheung, P., Musgrave, I.F., Wabnitz, P.A., Maselli, V.M., Skinner, J., Alewood, P.F., Brinkworth, C.S., and Bowie, J.H. (2005). An immunomodulator used to protect young in the pouch of the Tammar wallaby, *Macropus eugenii*. *FEBS J* **272**: 433-443.

Baumrucker, C.R., Burkett, A.M., Magliaro-Macrina, A.L., and Dechow, C.D. (2010). Colostrogenesis: mass transfer of immunoglobulin G1 into colostrum. *J Dairy Sci* **93**: 3031-3038.

Baumrucker, C.R., and Erondu, N.E. (2000). Insulin-like growth factor (IGF) system in the bovine mammary gland and milk. *J Mammary Gland Biol Neoplasia* **V5**: 53-64.

Bawden, W.S., and Nicholas, K.R. (1999). Molecular genetics of milk production. In: The Gentics of Cattle, R. Fries, and A. Ruvinsky, eds. (Wallingford, UK: CABI Publishing), pp. 539-576.

Bayne, C.J. (2003). Origins and evolutionary relationships between the innate and adaptive arms of immune systems. *Integr Comp Biol* **43**: 293-299.

Belov, K., Nguyen, M.A., Zenger, K.R., and Cooper, D.W. (2002). Ontogeny of immunoglobulin expression in the brushtail possum (*Trichosurus vulpecula*). *Dev Comp Immunol* **26**: 599-602.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**: 783-795.

Benson, D.A., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2015). GenBank. *Nucleic Acids Res* **43**: D30-35.

Berger, A., and Schechter, I. (1970). Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Phil Trans R Soc Lond B* **257**: 249-264.

Berger, P.J. (1966). Eleven-month "embryonic diapause" in a marsupial. *Nature* **211**: 435-436.

Berger, P.J., and Sharman, G.B. (1969). Progesterone-induced development of dormant blastocysts in the tammar wallaby, *Macropus eugenii* Desmarest; Marsupialia. *J Reprod Fertil* **20**: 201-210.

Berkmen, M. (2012). Production of disulfide-bonded proteins in *Escherichia coli*. *Protein Expression and Purification* **82**: 240-251.

Berlanga, J.J., Garcia-Ruiz, J.P., Perrot-Applanat, M., Kelly, P.A., and Edery, M. (1997). The short form of the prolactin (PRL) receptor silences PRL induction of the β-casein gene promoter. *Mol Endocrinol* **11**: 1449-1457.

Berndt, K.D., Guntert, P., Orbons, L.P., and Wuthrich, K. (1992). Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J Mol Biol* **227**: 757-775.

Bessette, P.H., Aslund, F., Beckwith, J., and Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc Natl Acad Sci U S A* **96**: 13703-13708.

Bevan, P. (2001). Insulin signalling. *J Cell Sci* **114**: 1429-1430.

Biggin, M.D. (2011). Animal transcription networks as highly connected, quantitative continua. *Dev Cell* **21**: 611-626.

Bignon, C., Binart, N., Ormandy, C., Schuler, L.A., Kelly, P.A., and Djiane, J. (1997). Long and short forms of the ovine prolactin receptor: cDNA cloning and genomic analysis reveal that the two forms arise by different alternative splicing mechanisms in ruminants and in rodents. *J Mol Endocrinol* **19**: 109-120.

Binart, N., Bachelot, A., and Bouilly, J. (2010). Impact of prolactin receptor isoforms on reproduction. *Trends Endocrinol Metab* **21**: 362-368.

Binart, N., Imbert-Bollore, P., Baran, N., Viglietta, C., and Kelly, P.A. (2003). A short form of the prolactin (PRL) receptor is able to rescue mammopoiesis in heterozygous PRL receptor mice. *Mol Endocrinol* **17**: 1066-1074.

Bininda-Emonds, O.R., Cardillo, M., Jones, K.E., MacPhee, R.D., Beck, R.M., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., and Purvis, A. (2007). The delayed rise of present-day mammals. *Nature* **446**: 507-512.

Bird, P.H., Hendry, K.A., Shaw, D.C., Wilde, C.J., and Nicholas, K.R. (1994). Progressive changes in milk protein gene expression and prolactin binding during lactation in the tammar wallaby (*Macropus eugenii*). *J Mol Endocrinol* **13**: 117-125.

Birkenmeier, E.H., Gwynn, B., Howard, S., Jerry, J., Gordon, J.I., Landschulz, W.H., and McKnight, S.L. (1989). Tissue-specific expression, developmental regulation, and genetic mapping of the gene encoding CCAAT/enhancer binding protein. *Genes & Development* **3**: 1146-1156.

Bishai, W.R., Rappuoli, R., and Murphy, J.R. (1987). High-level expression of a proteolytically sensitive diphtheria toxin fragment in *Escherichia coli*. *J Bacteriol* **169**: 5140-5151.

Bissell, M.J., and Ram, T.G. (1989). Regulation of functional cytodifferentiation and histogenesis in mammary epithelial cells: role of the extracellular matrix. *Environ Health Perspect* **80**: 61-70.

Blackburn, D.G. (1991). Evolutionary origins of the mammary-gland. *Mammal Review* **21**: 81-96.

Blackburn, D.G. (1993). Lactation: historical patterns and potential for manipulation. *J Dairy Sci* **76**: 3195-3212.

Blackburn, D.G., Hayssen, V., and Murphy, C.J. (1989). The origins of lactation and the evolution of milk - a review with new hypotheses. *Mammal Review* **19**: 1-26.

Blatchford, D.R., Hendry, K.A., Turner, M.D., Burgoyne, R.D., and Wilde, C.J. (1995). Vectorial secretion by constitutive and regulated secretory pathways in mammary epithelial cells. *Epithelial Cell Biol* **4**: 8-16.

Bleck, G.T., Wheeler, M.B., Hansen, L.B., Chester-Jones, H., and Miller, D.J. (2009). Lactose synthase components in milk: concentrations of α-lactalbumin and β1,4-galactosyltransferase in milk of cows from several breeds at various stages of lactation. *Reprod Domest Anim* **44**: 241-247.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294**: 1351-1362.

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., and Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**: 1633-1649.

Blum, J.W., and Baumrucker, C.R. (2008). Insulin-like growth factors (IGFs), IGF binding proteins, and other endocrine factors in milk: role in the newborn. *Adv Exp Med Biol* **606**: 397-422.

Bobrowicz, P., Davidson, R.C., Li, H., Potgieter, T.I., Nett, J.H., Hamilton, S.R., Stadheim, T.A., Miele, R.G., Bobrowicz, B., Mitchell, T.*, et al.* (2004). Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. *Glycobiology* **14**: 757-766.

Bode, L. (2006). Recent advances on structure, metabolism, and function of human milk oligosaccharides. *J Nutr* **136**: 2127-2130.

Bode, W., Walter, J., Huber, R., Wenzel, H.R., and Tschesche, H. (1984). The refined 2.2-A (0.22-nm) X-ray crystal structure of the ternary complex formed by bovine trypsinogen, valine-valine and the Arg15 analogue of bovine pancreatic trypsin inhibitor. *Eur J Biochem* **144**: 185-190.

Boer, E., Steinborn, G., Kunze, G., and Gellissen, G. (2007). Yeast expression platforms. *Applied microbiology and biotechnology* **77**: 513-523.

Bolander, F.F., Jr., Nicholas, K.R., and Topper, Y.J. (1979). Retention of glucocorticoid by isolated mammary tissue may complicate interpretation of results from *in vitro* experiments. *Biochem Biophys Res Commun* **91**: 247-252.

Bolander, F.F., Jr., Nicholas, K.R., Van Wyk, J.J., and Topper, Y.J. (1981). Insulin is essential for accumulation of casein mRNA in mouse mammary epithelial cells. *Proc Natl Acad Sci* **78**: 5682-5684.

Bole-Feysot, C., Goffin, V., Edery, M., Binart, N., and Kelly, P.A. (1998). Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. *Endocrine Reviews* **19**: 225-268.

Bonekamp, F., and Jensen, K.F. (1988). The AGG codon is translated slowly in *E. coli* even at very low expression levels. *Nucleic Acids Res* **16**: 3013-3024.

Boness, D.J., and Bowen, W.D. (1996). The evolution of maternal care in pinnipeds. *Bioscience* **46**: 645-654.

Bonnans, C., Chou, J., and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* **15**: 786-801.

Bonsing, J., and Mackinlay, A.G. (1987). Recent studies on nucleotide sequences encoding the caseins. *J Dairy Res* **54**: 447-461.

Borthwick, C.R., Young, L.J., and Old, J.M. (2014). The development of the immune tissues in marsupial pouch young. *Journal of morphology* **275**: 822-839.

Brambell, F.W. (1969). The transmission of immune globulins from the mother to the foetal and newborn young. *Proc Nutr Soc* **28**: 35-41.

Brambell, F.W.R. (1970). The transmission of passive immunity from mother to young, Vol 18 (Amsterdam: North-Holland Publishing Company).

Brawand, D., Wahli, W., and Kaessmann, H. (2008). Loss of egg yolk genes in mammals and the origin of lactation and placentation. *PLoS Biol* **6**: e63.

Brennan, A.J., Sharp, J.A., Digby, M.R., and Nicholas, K.R. (2007a). The tammar wallaby: a model to examine endocrine and local control of lactation. *IUBMB Life* **59**: 146-150.

Brennan, A.J., Sharp, J.A., Khalil, E., Digby, M.R., Mailer, S.L., Lefevre, C.M., and Nicholas, K.R. (2008a). A population of mammary epithelial cells do not require hormones or growth factors to survive. *J Endocrinol* **196**: 483-496.

Brennan, A.J., Sharp, J.A., Lefevre, C., Topcic, D., Auguste, A., Digby, M., and Nicholas, K.R. (2007b). The tammar wallaby and fur seal: models to examine local control of lactation. *J Dairy Sci* **90 Suppl 1**: E66-75.

Brennan, A.J., Sharp, J.A., Lefevre, C.M., and Nicholas, K.R. (2008b). Uncoupling the mechanisms that facilitate cell survival in hormone-deprived bovine mammary explants. *J Mol Endocrinol* **41**: 103-116.

Brew, K. (2003). α−Lactalbumin. In: Advanced Dairy Chemistry Volume 1: Proteins, P.F. Fox, and P.L.H. McSweeney, eds. (New York: Kluwer Academic/Plenum Publishers), pp. 387-419.

Brew, K. (2013). α-Lactalbumin. In: Advanced Dairy Chemistry Volume 1: Proteins, P.L.H. McSweeney, and P.F. Fox, eds. (New York, Heidelberg, Dordrecht, London: Springer), pp. 261-273.

Brew, K., Vanaman, T.C., and Hill, R.L. (1967). Comparison of the amino acid sequence of bovine alpha-lactalbumin and hens egg white lysozyme. *J Biol Chem* **242**: 3747-3749.

Brew, K., Vanaman, T.C., and Hill, R.L. (1968). The role of alpha-lactalbumin and the A protein in lactose synthetase: a unique mechanism for the control of a biological reaction. *Proc Natl Acad Sci U S A* **59**: 491-497.

Brisken, C., and Ataca, D. (2015). Endocrine hormones and local signals during the development of the mouse mammary gland. *Wiley Interdisciplinary Reviews: Developmental Biology* **4**: 181-195.

Brisken, C., Kaur, S., Chavarria, T.E., Binart, N., Sutherland, R.L., Weinberg, R.A., Kelly, P.A., and Ormandy, C.J. (1999). Prolactin controls mammary gland development via direct and indirect mechanisms. *Dev Biol* **210**: 96-106.

Brisken, C., and Rajaram, R.D. (2006). Alveolar and lactogenic differentiation. *J Mammary Gland Biol Neoplasia* **11**: 239-248.

Brizzard, B.L., Chubet, R.G., and Vizard, D.L. (1994). Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques* **16**: 730-735.

Brodbeck, U., Denton, W.L., Tanahashi, N., and Ebner, K.E. (1967). The isolation and identification of the B protein of lactose synthetase as α-lactalbumin. *J Biol Chem* **242**: 1391-1397.

Brubaker, S.W., Bonham, K.S., Zanoni, I., and Kagan, J.C. (2015). Innate immune pattern recognition: a cell biological perspective. *Annual review of immunology* **33**: 257-290.

Buck, B., Zamoon, J., Kirby, T.L., DeSilva, T.M., Karim, C., Thomas, D., and Veglia, G. (2003). Overexpression, purification, and characterization of recombinant Ca-ATPase regulators for high-resolution solution and solid-state NMR studies. *Protein Expression and Purification* **30**: 253-261.

Buczek, O., Krowarsch, D., and Otlewski, J. (2002). Thermodynamics of single peptide bond cleavage in bovine pancreatic trypsin inhibitor (BPTI). *Protein Sci* **11**: 924-932.

Bulaj, G., Koehn, R.E., and Goldenberg, D.P. (2004). Alteration of the disulfide-coupled folding pathway of BPTI by circular permutation. *Protein Sci* **13**: 1182-1196.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**: 78-94.

Burgering, M.J., Orbons, L.P., van der Doelen, A., Mulders, J., Theunissen, H.J., Grootenhuis, P.D., Bode, W., Huber, R., and Stubbs, M.T. (1997). The second Kunitz domain of human tissue factor pathway inhibitor: cloning, structure determination and interaction with factor Xa. *J Mol Biol* **269**: 395-407.

Burgoyne, R.D., and Duncan, J.S. (1998). Secretion of milk proteins. *J Mammary Gland Biol Neoplasia* **3**: 275-286.

Burgoyne, R.D., and Wilde, C.J. (1994). Control of secretory function in mammary epithelial cells. *Cellular Signalling* **6**: 607-616.

Buser, A.C., Gass-Handel, E.K., Wyszomierski, S.L., Doppler, W., Leonhardt, S.A., Schaack, J., Rosen, J.M., Watkin, H., Anderson, S.M., and Edwards, D.P. (2007). Progesterone receptor repression of prolactin/signal transducer and activator of transcription 5-mediated transcription of the β−casein gene in mammary epithelial cells. *Mol Endocrinol* **21**: 106-125.

Cabrita, L.D., Bottomley, S.P., and El-Gewely, M.R. (2004). Protein expression and refolding - a practical guide to getting the most out of inclusion bodies. In Biotechnology Annual Review (Elsevier), pp. 31-50.

Cane, K.N. (2005). The physiological and molecular regulation of lactation in fur seals. PhD Thesis, Department of Zoology (Melbourne: University of Melbourne).

Cane, K.N., Arnould, J.P.Y., and Nicholas, K.R. (2005). Characterisation of proteins in the milk of fur seals. *Comp Biochem Physiol B Biochem Mol Biol* **141**: 111-120.

Capuco, A.V., and Akers, R.M. (2009). The origin and evolution of lactation. *J Biol* **8**: 37.

Capuco, A.V., Kahl, S., Jack, L.J., Bishop, J.O., and Wallace, H. (1999). Prolactin and growth hormone stimulation of lactation in mice requires thyroid hormones. *Proc Soc Exp Biol Med* **221**: 345-351.

Carlsson, L.C., Bergelin, I.S., and Karlsson, B.W. (1974). Trypsin inhibition in urine of developing neonatal pigs and in sow's colostrum. *Enzyme* **18**: 176-188.

Case, T.J. (1978). Endothermy and parental care in the terrestrial vertebrates. *The American Naturalist* **112**: 861-874.

Casey, T.M., Crodian, J., Erickson, E., Kuropatwinski, K.K., Gleiberman, A.S., and Antoch, M.P. (2014). Tissue-specific changes in molecular clocks during the transition from pregnancy to lactation in mice. *Biol Reprod* **90**: 127.

Cassy, S., Charlier, M., Belair, L., Guillomot, M., Charron, G., Bloch, B., and Djiane, J. (1998). Developmental expression and localization of the prolactin receptor (PRL-R) gene in ewe mammary gland during pregnancy and lactation: estimation of the ratio of the two forms of PRL-R messenger ribonucleic acid. *Biol Reprod* **58**: 1290-1296.

Cattoni, D.I., Valeri, A., Le Gall, A., and Nollmann, M. (2015). A matter of scale: how emerging technologies are redefining our view of chromosome architecture. *Trends Genet* **31**: 454-464.

Cavaletto, M., Giuffrida, M.G., and Conti, A. (2008). Milk fat globule membrane components - a proteomic approach. *Adv Exp Med Biol* **606**: 129-141.

Cavanagh, M.E., Dziegielewska, K.M., Hinds, L.A., Mollgard, K., Reynolds, M.L., Saunders, N.R., and Tyndale-Biscoe, C.H. (1985). Development of the cerebral cortical plate in the tammar wallaby *Macropus-eugenii* is entirely post-natal. *Journal of Physiology (Cambridge)* **358**: 41P.

Cechova, D. (1976). Trypsin inhibitor from cow colostrum. *Methods Enzymol* **45**: 806-813.

Cechova, D., and Ber, E. (1974). Disulfide bonds of trypsin inhibitor from cow colostrum. *Collect Czech Chem Commun* **39**: 680-688.

Cechova, D., and Muszynska, G. (1970). Role of lysine 18 in active center of cow colostrum trypsin inhibitor. *FEBS Lett* **8**: 84-86.

Cechova, D., Svestkova, V.V., Keil, B., and Sorm, F. (1969). Similarities in primary structures of cow colostrum trypsin inhibitor and bovine basic pancreatic trypsin inhibitor. *FEBS Lett* **4**: 155-156.

Celik, E., and Calik, P. (2012). Production of recombinant proteins by yeast cells. *Biotechnology Advances* **30**: 1108-1118.

Cereghini, S., and Yaniv, M. (1984). Assembly of transfected DNA into chromatin: structural changes in the origin-promoter-enhancer region upon replication. *EMBO J* **3**: 1243-1253.

Ceriani, R.L. (1976). Hormone induction of specific protein synthesis in midpregnant mouse mammary cell culture. *J Exp Zool* **196**: 1-12.

Chakrabarty, A., Green, J.A., and Roberts, R.M. (2006a). Origin and evolution of the TKDP gene family. *Gene* **373**: 35-43.

Chakrabarty, A., MacLean, J.A., Hughes, A.L., Roberts, R.M., and Green, J.A. (2006b). Rapid evolution of the trophoblast Kunitz domain proteins (TKDPs) - A multigene family in ruminant ungulates. *J Mol Evol* **63**: 274-282.

Chakrabarty, A., and Roberts, M.R. (2007). Ets-2 and C/EBP-beta are important mediators of ovine trophoblast Kunitz domain protein-1 gene expression in trophoblast. *BMC molecular biology* **8**: 14.

Chand, H.S., Schmidt, A.E., Bajaj, S.P., and Kisiel, W. (2004). Structure-function analysis of the reactive site in the first Kunitz-type domain of human Tissue Factor Pathway Inhibitor-2. *J Biol Chem* **279**: 17500-17507.

Chang-Yi, C., Aragane, Y., Maeda, A., Yu-Lan, P., Takahashi, M., Lee-Hwa, K., and Tezuka, T. (1999). Bikunin, a serine protease inhibitor, is present on the cell boundary of epidermis. *J Invest Dermatol* **113**: 182-188.

Chapple, S.D., Crofts, A.M., Shadbolt, S.P., McCafferty, J., and Dyson, M.R. (2006). Multiplexed expression and screening for recombinant protein production in mammalian cells. *BMC biotechnology* **6**: 49.

Chauvet, M.T., Hurpet, D., Chauvet, J., and Acher, R. (1981). A reptilian neurohypophysial hormone, mesotocin (Ile8-oxytocin), in Australian marsupials. *FEBS Letters* **129**: 120-122.

Chen, C.-C., Stairs, D.B., Boxer, R.B., Belka, G.K., Horseman, N.D., Alvarez, J.V., and Chodosh, L.A. (2012). Autocrine prolactin induced by the Pten–Akt pathway is required for lactation initiation and provides a direct link between the Akt and Stat5 pathways. *Genes & Development* **26**: 2154-2168.

Chen, L.H., and Bissell, M.J. (1989). A novel regulatory mechanism for whey acidic protein gene expression. *Cell regulation* **1**: 45-54.

Cheng, Y.-C., Yan, F.-J., and Chang, L.-S. (2005). Taiwan cobra chymotrypsin inhibitor: cloning, functional expression and gene organization. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* **1747**: 213-220.

Choi, Y., Chakrabarti, R., Escamilla-Hernandez, R., and Sinha, S. (2009). Elf5 conditional knockout mice reveal its role as a master regulator in mammary alveolar development: failure of Stat5 activation and functional differentiation in the absence of Elf5. *Dev Biol* **329**: 227 - 241.

Chubet, R.G., and Brizzard, B.L. (1996). Vectors for expression and secretion of FLAG epitope-tagged proteins in mammalian cells. *Biotechniques* **20**: 136-141.

Clark, C.B. (1977). A preliminary report on weaning among chimpanzees of the Gombe National Park, Tanzania. In: Primate Bio-social Development, pp. 235-250.

Clevenger, C.V. (2004). Roles and regulation of Stat family transcription factors in human breast cancer. *American Journal of Pathology* **165**: 1449-1460.

Clevenger, C.V., and Plank, T.L. (1997). Prolactin as an autocrine/paracrine factor in breast tissue. *J Mammary Gland Biol Neoplasia* **2**: 59-68.

Close, Roger A., Friedman, M., Lloyd, Graeme T., and Benson, Roger B.J. (2015). Evidence for a mid-Jurassic adaptive radiation in mammals. *Current Biology* **25**: 2137-2142.

Cochrane, N.J., and Reynolds, E.C. (2012). Calcium phosphopeptides -- mechanisms of action and evidence for clinical efficacy. *Adv Dent Res* **24**: 41-47.

Coletta, R.D., Jedlicka, P., Gutierrez-Hartmann, A., and Ford, H.L. (2004). Transcriptional control of the cell cycle in mammary gland development and tumorigenesis. *J Mammary Gland Biol Neoplasia* **9**: 39-53.

Collet, C., and Joseph, R. (1993). A novel member of the lipocalin superfamily: tammar wallaby late-lactation protein. *Biochim Biophys Acta* **1167**: 219-222.

Collet, C., Joseph, R., and Nicholas, K. (1989). Molecular cloning and characterization of a novel marsupial milk protein gene. *Biochem Biophys Res Commun* **164**: 1380-1383.

Collet, C., Joseph, R., and Nicholas, K. (1991). A marsupial β-lactoglobulin gene: characterization and prolactin-dependent expression. *J Mol Endocrinol* **6**: 9-16.

Collet, C., Joseph, R., and Nicholas, K. (1992). Molecular characterization and *in-vitro* hormonal requirements for expression of two casein genes from a marsupial. *J Mol Endocrinol* **8**: 13-20.

Collier, R.J., Bauman, D.E., and Hays, R.L. (1977). Lactogenesis in explant cultures of mammary tissue from pregnant cows. *Endocrinology* **100**: 1192-1200.

Cook, H.W., Lentfer, J.W., Pearson, A.M., and Baker, B.E. (1970). Polar bear milk. IV. Gross composition, fatty acid, and mineral constitution. *Can J Zool* **48**: 217-219.

Correa, A., and Oppezzo, P. (2015). Overcoming the Solubility Problem in E. coli: Available Approaches for Recombinant Protein Production. *Insoluble Proteins: Methods and Protocols* **1258**: 27-44.

Cowan, P.E. (1989). Changes in milk composition during lactation in the common brushtail possum, *Trichosurus vulpecula* (Marsupialia: Phalangeridae). *Reproduction, Fertility and Development* **1**: 325-335.

Cowie, A.T. (1969). General Hormone factors involved in lactogenesis. In: Lactogenesis: the initiation of milk secretion at parturition, M. Reynolds, and S.J. Folley, eds. (Philadelphia: University of Pennsylvania Press), pp. 157-169.

Cowie, A.T. (1972). Lactation. In: Reproduction in mammals Book 3 : Hormonal control of reproduction, C.P. Austin, and R.V. Short, eds. (Cambridge, New York, Melbourne: Cambridge University Press), pp. 195-231.

Creamer, L.K., Plowman, J.E., Liddell, M.J., Smith, M.H., and Hill, J.P. (1998). Micelle stability: κ-casein structure and function. *J Dairy Sci* **81**: 3004-3012.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S.*, et al.* (2015). Ensembl 2015. *Nucleic Acids Res* **43**: D662-669.

Dai, D., Nanthkumar, N.N., Newburg, D.S., and Walker, W.A. (2000). Role of oligosaccharides and glycoconjugates in intestinal host defense. *J Pediatr Gastroenterol Nutr* **30 Suppl 2**: S23-33.

Dalboge, H., Jensen, E.B., Tottrup, H., Grubb, A., Abrahamson, M., Olafsson, I., and Carlsen, S. (1989). High-level expression of active human cystatin C in *Escherichia coli*. *Gene* **79**: 325-332.

Dale, T.C., Krnacik, M.J., Schmidhauser, C., Yang, C.L., Bissell, M.J., and Rosen, J.M. (1992). High-level expression of the rat whey acidic protein gene is mediated by elements in the promoter and 3' untranslated region. *Mol Cell Biol* **12**: 905-914.

Dalgleish, D.G. (1998). Casein micelles as colloids: surface structures and stabilities. *J Dairy Sci* **81**: 3013-3018.

Dall, S.R.X., and Boyd, I.L. (2004). Evolution of mammals: lactation helps mothers to cope with unreliable food supplies. *Proceedings of the Royal Society of London Series B, Biological Sciences* **271**: 2049-2057.

Dallas, D.C., Meyrand, M., and Barile, D. (2014). Production and Bioactivity of Bovine Milk Oligosaccharides. In: Food Oligosaccharides: Production, Analysis and Bioactivity, F.J. Moreno, and M.L. Sanz, eds. (Chichester, UK: John Wiley & Sons, Ltd), pp. 21-34.

Daly, K.A., Digby, M., Lefèvre, C., Mailer, S., Thomson, P., Nicholas, K., and Williamson, P. (2007). Analysis of the expression of immunoglobulins throughout lactation suggests two periods of immune transfer in the tammar wallaby (*Macropus eugenii*). *Veterinary Immunology and Immunopathology* **120**: 187-200.

Daly, K.A., Digby, M.R., Lefevre, C., Nicholas, K.R., Deane, E.M., and Williamson, P. (2008). Identification, characterization and expression of cathelicidin in the pouch young of tammar wallaby (*Macropus eugenii*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **149**: 524-533.

Daniels, D.L., ndez, J., Benink, H., ne, Niles, A., Murphy, N., Ford, M., Jones, R., Amunugama, R., Allen, D.*, et al.* (2014). Discovering Protein Interactions and Characterizing Protein Function Using HaloTag Technology. e51553.

Danielson, K.G., Oborn, C.J., Durban, E.M., Butel, J.S., and Medina, D. (1984). Epithelial mouse mammary cell line exhibiting normal morphogenesis *in vivo* and functional differentiation *in vitro*. *Proc Natl Acad Sci U S A* **81**: 3756-3760.

Darnell, J.E., Jr. (1996). The JAK-STAT pathway: summary of initial studies and recent advances. *Recent Prog Horm Res* **51**: 391-403; discussion 403-394.

Darnell, J.E., Jr., Kerr, I.M., and Stark, G.R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **264**: 1415-1421.

Darwin, C. (1872). The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, Sixth Edition edn (London: John Murray).

Das, R., and Vonderhaar, B. (1995). Transduction of prolactin's (PRL) growth signal through both long and short forms of the PRL receptor. *Mol Endocrinol* **9**: 1750-1759.

Davey, H.W., Wilkins, R.J., and Waxman, D.J. (1999). STAT5 signaling in sexually dimorphic gene expression and growth patterns. *Am J Hum Genet* **65**: 959-965.

Davies, D.T., Holt, C., and Christie, W.W. (1983). The composition of milk. In: Biochemistry of Lactation, T.B. Mepham, ed. (Amsterdam: Elsevier), pp. 71-117.

Davies, G.E., Sabatakos, G., Cryer, A., and Ramji, D.P. (2000). The ovine CCAAT-enhancer binding protein delta gene: cloning, characterization, and species-specific autoregulation. *Biochem Biophys Res Commun* **271**: 346-352.

Davis, J.A., and Linzer, D.I. (1989). Expression of multiple forms of the prolactin receptor in mouse liver. *Mol Endocrinol* **3**: 674-680.

De Leo, A.A., Lefevre, C., Topcic, D., Pharo, E., Cheng, J.F., Frappell, P., Westerman, M., Graves, J.A.M., and Nicholas, K.R. (2006). Characterization of two whey protein genes in the Australian dasyurid marsupial, the stripe-faced dunnart (*Sminthopsis macroura*). *Cytogenet Genome Res* **115**: 62-69.

de Marco, A. (2009). Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microbial Cell Factories* **8**: 26.

de Marco, A. (2013). Recombinant polypeptide production in *E. coli*: towards a rational approach to improve the yields of functional proteins. *Microbial cell factories [electronic resource]* **12**: 101.

Deane, E.M., and Cooper, D.W. (1984). Immunology of pouch young marsupials. I. Levels of immunoglobulin transferrin and albumin in the blood and milk of euros and wallaroos (hill kangaroos: marsupialia). *Developmental & Comparative Immunology* **8**: 863-876.

Deane, E.M., and Cooper, D.W. (1988). Immunological development of pouch young marsupials. In: The Developing Marsupial Models for Biomedical Research, C.H. Tyndale-Biscoe, and P.A. Janssens, eds. (Berlin, Germany: Springer-Verlag), pp. 190-199.

Deane, E.M., Cooper, D.W., and Renfree, M.B. (1990). Immunoglobulin G levels in fetal and newborn tammar wallabies (*Macropus eugenii*). *Reprod Fertil Dev* **2**: 369-375.

Deane, E.M., and Miller, R.D. (2000). Marsupial immunology: out of the pouch. *Developmental & Comparative Immunology* **24**: 443-444.

Dearth, L.R., Hutt, J., Sattler, A., Gigliotti, A., and DeWille, J. (2001). Expression and function of CCAAT/enhancer binding protein β (C/EBPβ) LAP and LIP isoforms in mouse mammary gland, tumors and cultured mammary epithelial cells. *J Cell Biochem* **82**: 357-370.

Dekker, J. (2008). Gene regulation in the third dimension. *Science* **319**: 1793-1794.

Del Tito, B.J., Jr., Ward, J.M., Hodgson, J., Gershater, C.J., Edwards, H., Wysocki, L.A., Watson, F.A., Sathe, G., and Kane, J.F. (1995). Effects of a minor isoleucyl tRNA on heterologous protein translation in Escherichia coli. *Journal of bacteriology* **177**: 7086-7091.

Delouis, C. (1978). Physiology of colostrum production. *Ann Rech Vet* **9**: 193-203.

Demain, A.L., and Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* **27**: 297-306.

Demmer, J. (1999). The prolactin receptor from the brushtail possum (*Trichosurus vulpecula*): cDNA cloning, expression and functional analysis. *Molecular and Cellular Endocrinology* **148**: 119-127.

Demmer, J., Burdon, T.G., Djiane, J., Watson, C.J., and Clark, A.J. (1995). The proximal milk protein binding factor binding site is required for the prolactin responsiveness of the sheep β-lactoglobulin promoter in Chinese hamster ovary cells. *Molecular and Cellular Endocrinology* **107**: 113-121.

Demmer, J., Ross, I.K., Ginger, M.R., Piotte, C.K., and Grigor, M.R. (1998). Differential expression of milk protein genes during lactation in the common brushtail possum (*Trichosurus vulpecula*). *J Mol Endocrinol* **20**: 37-44.

Demmer, J., Stasiuk, S.J., Grigor, M.R., Simpson, K.J., and Nicholas, K.R. (2001). Differential expression of the whey acidic protein gene during lactation in the brushtail possum (*Trichosurus vulpecula*). *Biochim Biophys Acta* **1522**: 187-194.

Derman, A.I., Prinz, W.A., Belin, D., and Beckwith, J. (1993). Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* **262**: 1744-1747.

Derrickson, E.M. (1992). Comparative reproductive strategies of altricial and precocial eutherian mammals. *Functional Ecology* **6**: 57-65.

Descombes, P., and Schibler, U. (1991). A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell* **67**: 569-579.

Devasahayam, M. (2007). Factors affecting the expression of recombinant glycoproteins. *Indian J Med Res* **126**: 22-27.

Devinoy, E., Hubert, C., Jolivet, G., Thepot, D., Clergue, N., Desaleux, M., Dion, M., Servely, J.L., and Houdebine, L.M. (1988). Recent data on the structure of rabbit milk protein genes and on the mechanism of the hormonal control of their expression. *Reprod Nutr Dev* **28**: 1145-1164.

Devinoy, E., and Rijnkels, M. (2010). Epigenetics in mammary gland biology and cancer. *J Mammary Gland Biol Neoplasia* **15**: 1-4.

Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C., and Rani, V. (2012). DNA-protein interactions: methods for detection and analysis. *Mol Cell Biochem* **365**: 279-299.

di Guan, C., Li, P., Riggs, P.D., and Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **67**: 21-30.

Diarra-Mehrpour, M., Bourguignon, J., Sesboue, R., Salier, J.P., Leveillard, T., and Martin, J.P. (1990). Structural analysis of the human inter-alpha-trypsin inhibitor light-chain gene. *Eur J Biochem* **191**: 131-139.

Diaz-Collier, J.A., Palmier, M.O., Kretzmer, K.K., Bishop, B.F., Combs, R.G., Obukowicz, M.G., Frazier, R.B., Bild, G.S., Joy, W.D., Hill, S.R.*, et al.* (1994). Refold and characterization of recombinant tissue factor pathway inhibitor expressed in *Escherichia coli*. *Thrombosis and haemostasis* **71**: 339-346.

Dickman, C.R. (2005). Marsupials of the World: An Introduction. In: Walker's Marsupials of the World (Baltimore, London: Johns Hopkins University Press), pp. 1-67.

Dils, R., and Forsyth, I.A. (1981). Preparation and culture of mammary gland explants. *Methods Enzymol* **72**: 724-742.

Dimri, G., Band, H., and Band, V. (2005). Mammary epithelial cell transformation: insights from cell culture and mouse models. *Breast Cancer Res* **7**: 171-179.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*: doi: 10.1038/nature11082.

Dong, B., and Zhao, F.Q. (2007a). Expression of the Oct-2 transcription factor in mouse mammary gland and cloning and characterization of a novel Oct-2 isoform. *Cell Tissue Res* **328**: 595-606.

Dong, B., and Zhao, F.Q. (2007b). Involvement of the ubiquitous Oct-1 transcription factor in hormonal induction of β-casein gene expression. *Biochem J* **401**: 57-64.

Donnelly, M.I., Zhou, M., Millard, C.S., Clancy, S., Stols, L., Eschenfeldt, W.H., Collart, F.R., and Joachimiak, A. (2006). An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expression and Purification* **47**: 446-454.

Doppler, W., Groner, B., and Ball, R.K. (1989). Prolactin and glucocorticoid hormones synergistically induce expression of transfected rat β-casein gene promoter constructs in a mammary epithelial cell line. *Proc Natl Acad Sci U S A* **86**: 104-108.

Doppler, W., Hock, W., Hofer, P., Groner, B., and Ball, R.K. (1990). Prolactin and glucocorticoid hormones control transcription of the b-casein gene by kinetically distinct mechanisms. *Mol Endocrinol* **4**: 912-919.

Doppler, W., Villunger, A., Jennewein, P., Brduscha, K., Groner, B., and Ball, R.K. (1991). Lactogenic hormone and cell type-specific control of the whey acidic protein gene promoter in transfected mouse cells. *Mol Endocrinol* **5**: 1624-1632.

Doppler, W., Welte, T., and Philipp, S. (1995). CCAAT/enhancer-binding protein isoforms β and δ are expressed in mammary epithelial cells and bind to multiple sites in the β-casein gene promoter. *J Biol Chem* **270**: 17962-17969.

Doppler, W., Windegger, M., Soratroi, C., Tomasi, J., Lechner, J., Rusconi, S., Cato, A.C., Almlof, T., Liden, J., Okret, S.*, et al.* (2001). Expression level-dependent contribution of glucocorticoid receptor domains for functional interaction with STAT5. *Mol Cell Biol* **21**: 3266-3279.

Dove, H., and Cork, S.J. (1989). Lactation in the tammar wallaby *Macropus eugenii* I. Milk consumption and the algebraic description of the lactation curve. *Journal of Zoology (London)* **219**: 385-398.

Dwek, R.A. (1996). Glycobiology: toward understanding the function of sugars. *Chem Rev* **96**: 683-720.

Dyson, M.R., Shadbolt, S.P., Vincent, K.J., Perera, R.L., and McCafferty, J. (2004). Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC biotechnology* **4**: 32.

Edery, M., Jolicoeur, C., Levi-Meyrueis, C., Dusanter-Fourt, I., Petridou, B., Boutin, J.M., Lesueur, L., Kelly, P.A., and Djiane, J. (1989). Identification and sequence analysis of a second form of prolactin receptor by molecular cloning of complementary DNA from rabbit mammary gland. *Proc Natl Acad Sci U S A* **86**: 2112-2116.

Edwards, M.J., Hinds, L.A., Deane, E.M., and Deakin, J.E. (2012). A review of complementary mechanisms which protect the developing marsupial pouch young. *Dev Comp Immunol* **37**: 213-220.

Eiberle, M.K., and Jungbauer, A. (2010). Technical refolding of proteins: Do we have freedom to operate? *Biotechnol J* **5**: 547-559.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *PNAS* **95**: 14863-14868.

Eisenmesser, E.Z., Kapust, R.B., Nawrocki, J.P., Mazzulla, M.J., Pannell, L.K., Waugh, D.S., and Byrd, R.A. (2000). Expression, purification, refolding, and characterization of recombinant human interleukin-13: utilization of intracellular processing. *Protein Expr Purif* **20**: 186-195.

Eiwegger, T., Stahl, B., Haidl, P., Schmitt, J., Boehm, G., Dehlink, E., Urbanek, R., and Szepfalusi, Z. (2010). Prebiotic oligosaccharides: *In vitro* evidence for gastrointestinal epithelial transfer and immunomodulatory properties. *Pediatr Allergy Immunol*.

Ekstrom, B., and Berggard, I. (1977). Human alpha1-microglobulin. Purification procedure, chemical and physiochemical properties. *J Biol Chem* **252**: 8048-8057.

Ellinger, S., Mach, M., Korn, K., and Jahn, G. (1991). Cleavage and purification of prokaryotically expressed HIV gag and env fusion proteins for detection of HIV antibodies in the ELISA. *Virology* **180**: 811-813.

Ellman, G.L. (1959). Tissue sulfhydryl groups. *Archives of Biochemistry and Biophysics* **82**: 70-77.

Emerman, J.T., Enami, J., Pitelka, D.R., and Nandi, S. (1977). Hormonal effects on intracellular and secreted casein in cultures of mouse mammary epithelial cells on floating collagen membranes. *Proc Natl Acad Sci U S A* **74**: 4466-4470.

Enami, J., and Nandi, S. (1977). Hormonal control of milk protein synthesis in cultured mouse mammary explants. *Cell Differ* **6**: 217-227.

Enders, A.C. (2009). Reasons for diversity of placental structure. *Placenta* **30**: 15-18.

Ethier, S.P., Summerfelt, R.M., Cundiff, K.C., and Asch, B.B. (1991). The influence of growth factors on the proliferative potential of normal and primary breast cancer-derived human breast epithelial cells. *Breast Cancer Res Treat* **17**: 221-230.

Evers, J.M.J.M. (2004). The milkfat globule membrane - compositional and structural changes post secretion by the mammary secretory cell. *International Dairy Journal* **14**: 661-674.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**: 186-194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175-185.

Fata, J.E., Werb, Z., and Bissell, M.J. (2004). Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. *Breast Cancer Res* **6**: 1-11.

Feeney, R.E., Means, G.E., and Bigler, J.C. (1969). Inhibition of human trypsin, plasmin, and thrombin by naturally occurring inhibitors of proteolytic enzymes. *J Biol Chem* **244**: 1957-1960.

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164-166.

Fenelon, J.C., Banerjee, A., and Murphy, B.D. (2014). Embryonic diapause: development on hold. *Int J Dev Biol* **58**: 163-174.

Fiat, A.M., and Jolles, P. (1989). Caseins of various origins and biologically active casein peptides and oligosaccharides: structural and physiological aspects. *Mol Cell Biochem* **87**: 5-30.

Fiat, A.M., Migliore-Samour, D., Jolles, P., Drouet, L., Bal dit Sollier, C., and Caen, J. (1993). Biologically active peptides from milk proteins with emphasis on two examples concerning antithrombotic and immunomodulating activities. *J Dairy Sci* **76**: 301-310.

Finch, L.M., Craig, V.A., Kind, A.J., Schnieke, A., Scott, A., Wells, M., and Wilde, C.J. (1996). Primary culture of ovine mammary epithelial cells. *Biochem Soc Trans* **24**: 369S.

Findlay, L. (1982a). Lactation in the marsupial, *Macropus eugenii*. PhD Thesis, School of Environmental and Life Sciences (Perth: Murdoch University).

Findlay, L. (1982b). The mammary glands of the tammar wallaby (*Macropus eugenii*) during pregnancy and lactation. *J Reprod Fertil* **65**: 59-66.

Findlay, L., and Renfree, M.B. (1984). Growth, development and secretion of the mammary gland of macropodid marsupials. In: Symposia of the Zoological Society of London 51 Physiological Strategies in Lactation, M. Peaker, R.G. Vernon, and C.H. Knight, eds. (London: Academic Press), pp. 403-432.

Flajnik, M.F., and Kasahara, M. (2001). Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* **15**: 351-362.

Flajnik, M.F., and Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet* **11**: 47-59.

Flo, T.H., Smith, K.D., Sato, S., Rodriguez, D.J., Holmes, M.A., Strong, R.K., Akira, S., and Aderem, A. (2004). Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. *Nature* **432**: 917-921.

Flower, D.R. (1996). The lipocalin protein family: structure and function. *Biochem J* **318 (Pt 1)**: 1-14.

Flower, D.R., North, A.C.T., and Sansom, C.E. (2000). The lipocalin protein family: structural and sequence overview. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1482**: 9-24.

Forsyth, I.A. (1971). Reviews of the progress of dairy science. Section A. Physiology. Organ culture techniques and the study of hormone effects on the mammary gland. *J Dairy Res* **38**: 419-444.

Forsyth, I.A. (1986). Variation among species in the endocrine control of mammary growth and function: the roles of prolactin, growth hormone, and placental lactogen. *J Dairy Sci* **69**: 886-903.

Forsyth, I.A. (1991). The mammary gland. *Baillieres Clin Endocrinol Metab* **5**: 809-832.

Forsyth, I.A., Gabai, G., and Morgan, G. (1999). Spatial and temporal expression of insulin-like growth factor-I, insulin-like growth factor-II and the insulin-like growth factor-I receptor in the sheep fetal mammary gland. *J Dairy Res* **66**: 35-44.

Fox, J.D., Routzahn, K.M., Bucher, M.H., and Waugh, D.S. (2003). Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS Letters* **537**: 53-57.

Fox, J.D., and Waugh, D.S. (2003). Maltose-Binding Protein as asolubility enhancer. *Methods in Molecular Biology* **205**: 99-117.

Fox, P.F., and McSweeney, P.L.H. (1998). Dairy Chemistry and Biochemistry. First Edition (London & New York: Springer - Verlag).

Fox, P.F., Uniacke-Lowe, T., McSweeney, P.L.H., and O'Mahony, J.A. (2015). Dairy Chemistry and Biochemistry. Second Edition (Switzerland: Springer International Publishing).

Freeman, M.E., Kanyicska, S., Lerant, A., and Nagy, G. (2000). Prolactin: structure, function, and regulation of secretion. *Physiological Reviews* **80**: 1523-1631.

Freyer, C., Zeller, U., and Renfree, M.B. (2003). The marsupial placenta: a phylogenetic analysis. *J Exp Zoolog A Comp Exp Biol* **299**: 59-77.

Fries, E., and Blom, A.M. (2000). Bikunin - not just a plasma proteinase inhibitor. *Int J Biochem Cell Biol* **32**: 125-137.

Frigo, L., and Woolley, P.A. (1997). Growth and Development of Pouch Young of the Stripe-faced Dunnart, *Sminthopsis macroura* (Marsupialia : Dasyuridae), in Captivity. *Australian Journal of Zoology* **45**: 157-170.

Galang, C.K., Muller, W.J., Foos, G., Oshima, R.G., and Hauser, C.A. (2004). Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors. *J Biol Chem* **279**: 11281-11292.

Galdikas, B.M.F., and Wood, J.W. (1990). Birth spacing patterns in humans and apes. *American Journal of Physical Anthropology* **83**: 185-191.

Gallego, M.I., Binart, N., Robinson, G.W., Okagaki, R., Coschigano, K.T., Perry, J., Kopchick, J.J., Oka, T., Kelly, P.A., and Hennighausen, L. (2001). Prolactin, growth

hormone, and epidermal growth factor activate Stat5 in different compartments of mammary tissue and exert different and overlapping developmental effects. *Dev Biol* **229**: 163-175.

Gamel, C.M., Davis, R.W., David, J.H., Meÿer, M.A., and Brandon, E. (2005). Reproductive energetics and female attendance patterns of Cape fur seals (*Arctocephalus pusillus pusillus*) during early lactation. *Amer Mid Nat* **153**: 152-170.

Ganguly, R., Mehta, N.M., Ganguly, N., and Banerjee, M.R. (1979). Glucocorticoid modulation of casein gene transcription in mouse mammary gland. *PNAS* **76**: 6466-6470.

Gehring, H.M. (2005). The cloning and expression of the ovarian prolactin receptor in the tammar wallaby, *Macropus eugenii*. MSc Thesis, Department of Zoology (Melbourne: University of Melbourne).

Gemmell, R.T., Peters, B., and Nelson, J. (1988). Ultrastructural identification of Merkel cells around the mouth of the newborn marsupial. *Anatomy and Embryology* **177**: 403-408.

Gemmell, R.T., and Sernia, C. (1992). The transfer of thyroxine from the mother to the young of the marsupials, the bandicoot (*Isoodon macrourus*) and the possum (*Trichosurus vulpecula*). *Comp Biochem Physiol Comp Physiol* **103**: 541-543.

Gemmell, R.T., Veitch, C., and Nelson, J. (2002). Birth in marsupials. *Comp Biochem Physiol B Biochem Mol Biol* **131**: 621-630.

Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J.*, et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**: R80.

Georges, J.-Y., and Guinet, C. (2000). Maternal care in the subantarctic fur seals on Amsterdam Island. *Ecology* **81**: 295-308.

Geymayer, S., and Doppler, W. (2000). Activation of NF-κB p50/p65 is regulated in the developing mammary gland and inhibits STAT5-mediated β-casein gene expression. *Faseb Journal* **14**: 1159-1170.

Ghrayeb, J., Kimura, H., Takahara, M., Hsiung, H., Masui, Y., and Inouye, M. (1984). Secretion cloning vectors in *Escherichia coli*. *EMBO J* **3**: 2437-2442.

Gibson, C.A., Vega, J.R., Baumrucker, C.R., Oakley, C.S., and Welsch, C.W. (1991). Establishment and characterization of bovine mammary epithelial cell lines. *In Vitro Cell Dev Biol* **27A**: 585-594.

Gigliotti, A.P., and DeWille, J.W. (1998). Lactation status influences expression of CCAAT/enhancer binding protein isoform mRNA in the mouse mammary gland. *J Cell Physiol* **174**: 232-239.

Ginger, M.R. (1998). Caseins of the common brushtailed possum (*Trichosurus vulpecula*). PhD Thesis, Department of Biochemistry (Dunedin: University of Otago, New Zealand).

Girard, T.J., Warren, L.A., Novotny, W.F., Likert, K.M., Brown, S.G., Miletich, J.P., and Broze, G.J., Jr. (1989). Functional significance of the Kunitz-type inhibitory domains of lipoprotein-associated coagulation inhibitor. *Nature* **338**: 518-520.

Gjorevski, N., and Nelson, C.M. (2011). Integrated morphodynamic signalling of the mammary gland. *Nat Rev Mol Cell Biol* **12**: 581-593.

Glass, C., Wuertzer, C., Cui, X., Bi, Y., Davuluri, R., Xiao, Y.-Y., Wilson, M., Owens, K., Zhang, Y., and Perkins, A. (2013). Global identification of EVI1 target genes in Acute Myeloid Leukemia. *PLoS ONE* **8**: e67134.

Godovac-Zimmermann, J., Conti, A., James, L., and Napolitano, L. (1988). Microanalysis of the amino-acid sequence of monomeric β-lactoglobulin I from donkey (*Equus asinus*) milk. The primary structure and its homology with a superfamily of hydrophobic molecule transporters. *Biol Chem Hoppe Seyler* **369**: 171-179.

Godovac-Zimmermann, J., Conti, A., Liberatori, J., and Braunitzer, G. (1985). Homology between the primary structures of β-lactoglobulins and human retinol-binding protein: evidence for a similar biological function? *Biol Chem Hoppe Seyler* **366**: 431-434.

Godovac-Zimmermann, J., Conti, A., Sheil, M., and Napolitano, L. (1990). Covalent structure of the minor monomeric β-lactoglobulin II component from donkey milk. *Biol Chem Hoppe Seyler* **371**: 871-879.

Goffin, V., Binart, N., Touraine, P., and Kelly, P.A. (2002). Prolactin: The new biology of an old hormone. *Annual Review of Physiology* **64**: 47-67.

Goffin, V., and Kelly, P.A. (1996). Prolactin and growth hormone receptors. *Clin Endocrinol (Oxf)* **45**: 247-255.

Goffin, V., Shiverick, K.T., Kelly, P.A., and Martial, J.A. (1996). Sequence-function relationships within the expanding family of prolactin, growth hormone, placental lactogen, and related proteins in mammals. *Endocr Rev* **17**: 385-410.

Goldenberg, D.P. (1988). Kinetic analysis of the folding and unfolding of a mutant form of bovine pancreatic trypsin inhibitor lacking the cysteine-14 and -38 thiols. *Biochemistry* **27**: 2481-2489.

Goldman, A.S. (2002). Evolution of the mammary gland defense system and the ontogeny of the immune system. *J Mammary Gland Biol Neoplasia* **7**: 277-289.

Goldman, A.S., Goldblum, R.M., and Hanson, L.A. (1990). Anti-inflammatory systems in human milk. *Adv Exp Med Biol* **262**: 69-76.

Goodman, H.S., and Rosen, J.M. (1990). Transcriptional analysis of the mouse beta-casein gene. *Mol Endocrinol* **4**: 1661-1670.

Gopal, G.J., and Kumar, A. (2013). Strategies for the Production of Recombinant Protein in Escherichia coli. *Protein J* **32**: 419-425.

Gordon, K., Fletcher, T.P., and Renfree, M.B. (1988). Reactivation of the quiescent corpus luteum and diapausing embryo after temporary removal of the sucking stimulus in the tammar wallaby (Macropus eugenii). *J Reprod Fertil* **83**: 401-406.

Gordon, K.E., Binas, B., Chapman, R.S., Kurian, K.M., Clarkson, R.W., Clark, A.J., Lane, E.B., and Watson, C.J. (2000). A novel cell culture model for studying differentiation and apoptosis in the mouse mammary gland. *Breast Cancer Res* **2**: 222-235.

Gouilleux, F., Wakao, H., Mundt, M., and Groner, B. (1994). Prolactin induces phosphorylation of Tyr694 of Stat5 (MGF), a prerequisite for DNA binding and induction of transcription. *EMBO J* **13**: 4361-4369.

Grant, T. (1984). The Platypus (Kensington, NSW: New South Wales University Press).

Gräslund S, N.P., Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schütz A, Heinemann U, Yokoyama S, Büssow K, Gunsalus KC. (2008). Protein production and purification. *Nat Meth* **5**: 135-146.

Green, B. (1984). Composition of milk and energetics of growth in marsupials. In: Symposia of the Zoological Society of London 51 Physiological Strategies in Lactation, M. Peaker, R.G. Vernon, and C.H. Knight, eds. (London: Academic Press), pp. 369–387.

Green, B., Griffiths, M., and Leckie, R.M. (1983). Qualitative and quantitative changes in milk fat during lactation in the tammar wallaby (*Macropus eugenii*). *Aust J Biol Sci* **36**: 455-461.

Green, B., and Merchant, J.C. (1988). The composition of marsupial milk. In: The Developing Marsupial Models for Biomedical Research, C.H. Tyndale-Biscoe, and P.A. Janssens, eds. (Berlin, Germany: Springer-Verlag), pp. 41-54.

Green, B., Newgrain, K., and Merchant, J. (1980). Changes in milk composition during lactation in the tammar wallaby (*Macropus eugenii*). *Aust J Biol Sci* **33**: 35-42.

Griffiths, M. (1965). Rate of growth and intake of milk in a suckling echidna. *Comp Biochem Physiol* **16**: 383-392.

Griffiths, M. (1968). Echidnas (Sydney: Pergamon Press).

Griffiths, M. (1973). Observation of the comparative anatomy and ultrastructure of mammary glands and on the fatty acids of the triglycerides in the platypus and echidna milk fats. *Journal of Zoology* **169**: 255-279.

Griffiths, M. (1978). The Biology of the Monotremes (New York, USA: Academic Press).

Griffiths, M., and Barton, A.A. (1966). The ontongeny of the stomach in the pouch young of the red kangaroo. *CSIRO Wildlife Research* **11**: 169-185.

Griffiths, M., Green, B., Leckie, R.M.C., Messer, M., and Newgrain, K.W. (1984). Constituents of platypus and echidna milk, with particular reference to the fatty acid complement of the triglycerides. *Australian Journal of Biological Sciences* **37**: 323-329.

Griffiths, M., McIntosh, D.L., and Leckie, R.M.C. (1972). The mammary glands of the Red kangaroo with observations on the fatty acid components of the milk triglycerides. *Journal of Zoology* **166**: 265-275.

Griffiths, M., and Slater, E. (1988). The significance of striated muscle in the mammary glands of marsupials. *J Anat* **156**: 141-156.

Grimley, P.M., Dong, F., and Rui, H. (1999). Stat5a and Stat5b: fraternal twins of signal transduction and transcriptional activation. *Cytokine Growth Factor Rev* **10**: 131-157.

Grimm, S.L., and Rosen, J.M. (2003). The role of C/EBPβ in mammary gland development and breast cancer. *J Mammary Gland Biol Neoplasia* **8**: 191-204.

Grodberg, J., and Dunn, J.J. (1988). ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J Bacteriol* **170**: 1245-1253.

Groenen, M.A., Dijkhof, R.J., van der Poel, J.J., van Diggelen, R., and Verstege, E. (1992). Multiple octamer binding sites in the promoter region of the bovine αs2-casein gene. *Nucleic Acids Res* **20**: 4311-4318.

Groenen, M.A.M., and van der Poel, J.J. (1994). Regulation of expression of milk protein genes: a review. *Livestock Production Science* **38**: 61-78.

Groner, B. (2002). Transcription factor regulation in mammary epithelial cells. *Domest Anim Endocrinol* **23**: 25-32.

Groner, B., Altiok, S., and Meier, V. (1994). Hormonal regulation of transcription factor activity in mammary epithelial cells. *Mol Cell Endocrinol* **100**: 109-114.

Groner, B., Fritsche, M., Stocklin, E., Berchtold, S., Merkle, C., Moriggl, R., and Pfitzner, E. (2000). Regulation of the trans-activation potential of STAT5 through its DNA-binding activity and interactions with heterologous transcription factors. *Growth Horm IGF Res* **10 Suppl B**: S15-20.

Groner, B., and Gouilleux, F. (1995). Prolactin-mediated gene activation in mammary epithelial cells. *Curr Opin Genet Dev* **5**: 587-594.

Gross, K.L., and Cidlowski, J.A. (2008). Tissue-specific glucocorticoid action: a family affair. *Trends in Endocrinology & Metabolism* **19**: 331-339.

Grzesiak, A., Helland, R., Smalas, A.O., Krowarsch, D., Dadlez, M., and Otlewski, J. (2000). Substitutions at the P1' position in BPTI strongly affect the association energy with serine proteinases. *Journal of Molecular Biology* **301**: 205-217.

Guan, H., and Kiss-Toth, E. (2008). Advanced technologies for studies on protein interactomes. *Adv Biochem Eng Biotechnol* **110**: 1-24.

Guan, M., Su, B., Ye, C., and Lu, Y. (2002). Production of extracellular domain of human tissue factor using maltose-binding protein fusion system. *Protein Expression and Purification* **26**: 229-234.

Gupta, R., Jung, E., and Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins. *In preparation*.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in Biotechnology* **22**: 346-353.

Gutierrez-Hartmann, A., Duval, D.L., and Bradford, A.P. (2007). ETS transcription factors in endocrine systems. *Trends Endocrinol Metab* **18**: 150-158.

Guyette, W.A., Matusik, R.J., and Rosen, J.M. (1979). Prolactin-mediated transcriptional and post-transcriptional control of casein gene expression. *Cell* **17**: 1013-1023.

Hajjoubi, S., Rival-Gervier, S., Hayes, H., Floriot, S., Eggen, A., Piumi, F., Chardon, P., Houdebine, L.M., and Thepot, D. (2006). Ruminants genome no longer contains Whey Acidic Protein gene but only a pseudogene. *Gene* **370**: 104-112.

Hall, L., Emery, D.C., Davies, M.S., Parker, D., and Craig, R.K. (1987). Organization and sequence of the human α-lactalbumin gene. *Biochem J* **242**: 735-742.

Halliday, J.A., Bell, K., McKenzie, H.A., and Shaw, D.C. (1990). Feline whey proteins: identification, isolation and initial characterization of alpha-lactalbumin, beta-lactoglobulin and lysozyme. *Comp Biochem Physiol B* **95**: 773-779.

Halliday, J.A., Bell, K., and Shaw, D.C. (1991). The complete amino acid sequence of feline beta-lactoglobulin II and a partial revision of the equine beta-lactoglobulin II sequence. *Biochim Biophys Acta* **1077**: 25-30.

Hambling, S.G., McAlpine, A.S., and Sawyer, L. (1992). β-Lactoglobulin. In: Advanced Dairy Chemistry Volume 1: Proteins, P.F. Fox, ed. (London: Elsevier Science Publishers Ltd), pp. 141-190.

Hambraeus, L., and Lonnerdal, B. (2003). Nutritional aspects of milk proteins. In: Advanced Dairy Chemistry Volume 1: Proteins, P.F. Fox, and P.L.H. McSweeney, eds. (New York: Kluwer Academic/Plenum Publishers), pp. 605-646.

Hanahan, D.J., and Thompson, G.A., Jr. (1963). Complex lipids. *Annu Rev Biochem* **32**: 215-240.

Handler, J.S. (1989). Overview of epithelial polarity. *Annual Review of Physiology* **51**: 729-740.

Hanley, S., and Powell, R. (1994). Sequence of a cDNA clone encoding the Atlantic salmon alpha 1-microglobulin/bikunin protein. *Gene* **147**: 297-298.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* **345**: 458-460.

Hartley, J.L. (2006). Cloning technologies for protein expression and purification. *Current opinion in biotechnology* **17**: 359-366.

Hartmann, P.E., Trevethan, P., and Shelton, J.N. (1973). Progesterone and oestrogen and the initiation of lactation in ewes. *J Endocrinol* **59**: 249-259.

Hassiotou, F., Beltran, A., Chetwynd, E., Stuebe, A.M., Twigger, A.J., Metzger, P., Trengove, N., Lai, C.T., Filgueira, L., Blancafort, P.*, et al.* (2012). Breastmilk is a novel source of stem cells with multilineage differentiation potential. *Stem Cells* **30**: 2164-2174.

Hassiotou, F., and Geddes, D.T. (2015). Immune cell–mediated protection of the mammary gland and the infant during breastfeeding. *Advances in Nutrition: An International Review Journal* **6**: 267-275.

Hassiotou, F., Geddes, D.T., and Hartmann, P.E. (2013). Cells in human milk: state of the science. *J Hum Lact* **29**: 171-182.

Hata, T., and Nakayama, M. (2007). Rapid single-tube method for small-scale affinity purification of polyclonal antibodies using HaloTag Technology. *J Biochem Biophys Methods* **70**: 679-682.

Hawdon, J.M., Datu, B., and Crowell, M. (2003). Molecular cloning of a novel multidomain Kunitz-type proteinase inhibitor from the hookworm *Ancylostoma caninum*. *Journal of Parasitology* **89**: 402-407.

Hayflick, L. (1965). The limited *in vitro* lifetime of human diploid cell strains. *Exp Cell Res* **37**: 614-636.

Hayflick, L. (1998). A brief history of the mortality and immortality of cultured cells. *Keio J Med* **47**: 174-182.

Hayflick, L., and Moorhead, P.S. (1961). The serial cultivation of human diploid cell strains. *Exp Cell Res* **25**: 585-621.

Hayssen, V. (1993). Empirical and theoretical constraints on the evolution of lactation. *J Dairy Sci* **76**: 3213-3233.

Hayssen, V., and Blackburn, D.G. (1985). α-lactalbumin and the origins of lactation. *Evolution* **39**: 1147-1149.

Hayssen, V., Lacy, R.C., and Parker, P.J. (1985). Metatherian reproduction - transitional or transcending. *American Naturalist* **126**: 617-632.

Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protocols* **2**: 1849-1861.

Hennighausen, L., and Robinson, G.W. (1998). Think globally, act locally: the making of a mouse mammary gland. *Genes Dev* **12**: 449-455.

Hennighausen, L., and Robinson, G.W. (2001). Signaling pathways in mammary gland development. *Developmental Cell* **1**: 467-475.

Hennighausen, L., and Robinson, G.W. (2008). Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B. *Genes Dev* **22**: 711-721.

Hennighausen, L., Robinson, G.W., Wagner, K.U., and Liu, W. (1997). Prolactin signaling in mammary gland development. *J Biol Chem* **272**: 7567-7569.

Hennighausen, L., and Robinson, G.W. (2005). Information networks in the mammary gland. *Nature Reviews Molecular Cell Biology* **6**: 715-725.

Hennighausen, L.G., and Sippel, A.E. (1982). Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. *Nucleic Acids Res* **10**: 2677-2684.

Hennighausen, L.G., Sippel, A.E., Hobbs, A.A., and Rosen, J.M. (1982). Comparative sequence analysis of the mRNAs coding for mouse and rat whey protein. *Nucleic Acids Res* **10**: 3733-3744.

Hernandez-Castellano, L.E., Almeida, A.M., Castro, N., and Arguello, A. (2014). The colostrum proteome, ruminant nutrition and immunity: a review. *Curr Protein Pept Sci* **15**: 64-74.

Hickford, D., Frankenberg, S., and Renfree, M.B. (2010). The tammar wallaby, *Macropus eugenii*: A model kangaroo for the study of developmental and reproductive biology. In: Emerging Model Organisms: A Laboratory Manual, Volume 2 (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 449-494.

Higdon, J.W., Bininda-Emonds, O.R., Beck, R.M., and Ferguson, S.H. (2007). Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evol Biol* **7**: 216.

Hinds, L.A. (1988). Hormonal control of lactation. In: The Developing Marsupial Models for Biomedical Research, C.H. Tyndale-Biscoe, and P.A. Janssens, eds. (Berlin, Germany: Springer-Verlag), pp. 55-67.

Hinds, L.A., and Tyndale-Biscoe, C.H. (1982). Prolactin in the marsupial *Macropus eugenii*, during the estrous cycle, pregnancy and lactation. *Biology of Reproduction* **26**: 391-398.

Hinds, L.A., and Tyndale-Biscoe, C.H. (1985). Seasonal and circadian patterns of circulating prolactin during lactation and seasonal quiescence in the tammar, *Macropus eugenii*. *J Reprod Fertil* **74**: 173-183.

Hirano, M., Das, S., Guo, P., and Cooper, M.D. (2011). The Evolution of Adaptive Immunity in Vertebrates. In: Advances in Immunology, W.A. Frederick, ed. (Academic Press), pp. 125-157.

Hirche, T.O., Benabid, R., Deslee, G., Gangloff, S., Achilefu, S., Guenounou, M., Lebargy, F., Hancock, R.E., and Belaaouaj, A. (2008). Neutrophil Elastase Mediates Innate Host Protection against Pseudomonas aeruginosa. *The Journal of Immunology* **181**: 4945-4954.

Hmiel, L.K., Brorson, K.A., and Boyne, M.T. (2015). Post-translational structural modifications of immunoglobulin G and their effect on biological activity. *Anal Bioanal Chem* **407**: 79-94.

Hoffman, B.G., and Jones, S.J. (2009). Genome-wide identification of DNA-protein interactions using chromatin immunoprecipitation coupled with flow cell sequencing. *J Endocrinol* **201**: 1-13.

Holt, A.B., Renfree, M.B., and Cheek, D.B. (1981). Comparative aspects of brain growth: a critical evaluation of mammalian species used in brain growth research with emphasis

on the Tammar wallaby. In: Fetal Brain Disorders - Recent Approaches to the Problem of Mental Deficiency, B.S. Hetzel, and R.M. Smith, eds. (Amsterdam: Elsevier/North-Holland Biomedical Press), pp. 17-43.

Holt, C. (1998). Casein micelle substructure and calcium phosphate interactions studied by sephacryl column chromatography. *J Dairy Sci* **81**: 2994-3003.

Hopp, T.P., Prickett, K.S., Price, V.L., Libby, R.T., March, C.J., Pat Cerretti, D., Urdal, D.L., and Conlon, P.J. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Nat Biotech* **6**: 1204-1210.

Horne, D.S., Anema, S., Zhu, X., Nicholas, K.R., and Singh, H. (2007). A lactational study of the composition and integrity of casein micelles from the milk of the tammar wallaby (*Macropus eugenii*). *Archives of Biochemistry and Biophysics* **467**: 107-118.

Horseman, N.D. (1999). Prolactin and mammary gland development. *J Mammary Gland Biol Neoplasia* **4**: 79-88.

Horseman, N.D., and Collier, R.J. (2014). Serotonin: a local regulator in the mammary gland epithelium. *Annual review of animal biosciences* **2**: 353-374.

Horseman, N.D., and Gregerson, K.A. (2014). Prolactin actions. *J Mol Endocrinol* **52**: R95-R106.

Houdebine, L.M. (1999). The regulation of milk protein synthesis. In: Biology of Lactation, J. Martinet, L.M. Houdebine, and H.H. Head, eds. (Paris: INRA), pp. 401-427.

Houdebine, L.M., Delouis, C., and Devinoy, E. (1978). Post-transcriptional stimulation of casein synthesis by thyroid hormone. *Biochimie* **60**: 809-812.

Hovey, R.C., McFadden, T.B., and Akers, R.M. (1999). Regulation of mammary gland growth and morphogenesis by the mammary fat pad: a species comparison. *J Mammary Gland Biol Neoplasia* **4**: 53-68.

Hovey, R.C., and Trott, J.F. (2004). Morphogenesis of mammary gland development. *Adv Exp Med Biol* **554**: 219-228.

Hovey, R.C., Trott, J.F., and Vonderhaar, B.K. (2002). Establishing a framework for the functional mammary gland: from endocrinology to morphology. *J Mammary Gland Biol Neoplasia* **7**: 17-38.

Howlett, A.R., and Bissell, M.J. (1993). The influence of tissue microenvironment (stroma and extracellular matrix) on the development and function of mammary epithelium. *Epithelial Cell Biol* **2**: 79-89.

Hu, H., Wang, J., Bu, D., Wei, H., Zhou, L., Li, F., and Loor, J.J. (2009). *In vitro* culture and characterization of a mammary epithelial cell line from Chinese Holstein dairy cow. *PLoS One* **4**: e7636.

Hu, Z.-Z., Zhuang, L., and Dufau, M.L. (1998). Prolactin receptor gene diversity: structure and regulation. *Trends in Endocrinology and Metabolism* **9**: 94-102.

Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Res* **9**: 868-877.

Huber, R., and Bode, W. (1978). Structural basis of the activation and action of trypsin. *Accounts of Chemical Research* **11**: 114-122.

Hughes, K., and Watson, C.J. (2012). The spectrum of STAT functions in mammary gland development. *JAK-STAT* **1**: 151-158.

Hughes, R.L. (1962). Reproduction in the macropod marsupial *Potorous tridactylus* (Kerr). *Aust J Zool* **10**: 193-224.

Hughes, R.L., and Hall, L.S. (1988). Structural adaptations of the newborn marsupial. In: The Developing Marsupial Models for Biomedical Research, C.H. Tyndale-Biscoe, and P.A. Janssens, eds. (Berlin, Germany: Springer-Verlag), pp. 8-27.

Hunziker, W., and Kraehenbuhl, J.P. (1998). Epithelial transcytosis of immunoglobulins. *J Mammary Gland Biol Neoplasia* **3**: 287-302.

Hurley, W.L. (2003). Immunoglobulins in mammary secretions. In: Advanced Dairy Chemistry Volume 1: Proteins, P.F. Fox, and P.L.H. McSweeney, eds. (New York: Kluwer Academic/Plenum Publishers), pp. 421-447.

Milk Composition. http://classes.aces.uiuc.edu/AnSci308/milkcomp.html.

Hurley, W.L., and Theil, P.K. (2011). Perspectives on immunoglobulins in colostrum and milk. *Nutrients* **3**: 442-474.

Hurley, W.L., and Theil, P.K. (2013). Immunoglobulins in mammary secretions. In: Advanced Dairy Chemistry Volume 1: Proteins, P.F. Fox, and P.L.H. McSweeney, eds. (New York: Kluwer Academic/Plenum Publishers), pp. 275-294.

Huynh, H.T., Robitaille, G., and Turner, J.D. (1991). Establishment of bovine mammary epithelial cells (MAC-T): an *in vitro* model for bovine lactation. *Exp Cell Res* **197**: 191-199.

Hynes, N.E., Cella, N., and Wartmann, M. (1997). Prolactin mediated intracellular signaling in mammary epithelial cells. *J Mammary Gland Biol Neoplasia* **2**: 19-27.

Ikeda, K., Nukumi, N., Iwamori, T., Osawa, M., Naito, K., and Tojo, H. (2004). Inhibitory function of whey acidic protein in the cell-cycle progression of mouse mammary epithelial cells (EpH4/K6 cells). *J Reprod Dev* **50**: 87-96.

Ikeo, K., Takahashi, K., and Gojobori, T. (1992). Evolutionary origin of a Kunitz-type trypsin inhibitor domain inserted in the amyloid beta precursor protein of Alzheimer's disease. *J Mol Evol* **34**: 536-543.

Inman, J.L., Robertson, C., Mott, J.D., and Bissell, M.J. (2015). Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* **142**: 1028-1042.

Ip, M.M., and Darcy, K.M. (1996). Three-dimensional mammary primary culture model systems. *J Mammary Gland Biol Neoplasia* **1**: 91-110.

Irwin, D.M., Biegel, J.M., and Stewart, C.B. (2011). Evolution of the mammalian lysozyme gene family. *BMC Evol Biol* **11**: 166.

IUCN (2015). The IUCN Red List of Threatened Species. *IUCN 2015 The IUCN Red List of Threatened Species* **Version 2015-4**: Downloaded on 19 November 2015.

Iwasaki, A., and Medzhitov, R. (2010). Regulation of Adaptive Immunity by the Innate Immune System. *Science* **327**: 291-295.

Iwasaki, A., and Medzhitov, R. (2015). Control of adaptive immunity by the innate immune system. *Nat Immunol* **16**: 343-353.

Jack, L.J., Kahl, S., St Germain, D.L., and Capuco, A.V. (1994). Tissue distribution and regulation of 5'-deiodinase processes in lactating rats. *J Endocrinol* **142**: 205-215.

Jacobs, P.P., and Callewaert, N. (2009). N-glycosylation engineering of biopharmaceutical expression systems. *Curr Mol Med* **9**: 774-800.

Jahn, G.A., Daniel, N., Jolivet, G., Belair, L., Bole-Feysot, C., Kelly, P.A., and Djiane, J. (1997). *In vivo* study of prolactin (PRL) intracellular signalling during lactogenesis in the rat: JAK/STAT pathway is activated by PRL in the mammary gland but not in the liver. *Biol Reprod* **57**: 894-900.

Jana, S., and Deb, J.K. (2005). Strategies for efficient production of heterologous proteins in Escherichia coli. *Applied microbiology and biotechnology* **67**: 289-298.

Janssens, P.A., and Hinds, L.A. (1981). Long-term effects of corticosteroid administration in the tammar wallaby, *Macropus eugenii*. *General and Comparative Endocrinology* **45**: 56-60.

Janssens, P.A., Hulbert, A.J., and Baudinette, R.V. (1997). Development of the pouch young from birth to pouch vacation. In: Marsupial Biology Recent Research, New Perspectives, N.R. Saunders, and L. Hinds, eds. (Sydney: University of New South Wales Press), pp. 71–89.

Janssens, P.A., and Messer, M. (1988). Changes in nutritional metabolism during weaning. In: The Developing Marsupial Models for Biomedical Research, C.H. Tyndale-Bisco, and P.A. Janssens, eds. (Berlin, Germany: Springer-Verlag), pp. 162-175.

Janssens, P.A., and Ternouth, J.H. (1987). The transition from milk to forage diets. In: The Nutrition of Herbivores, J.B. Hacker, and J.H. Ternouth, eds. (London: Academic Press), pp. 281-305.

Jarzabek, K., Koda, M., Kozlowski, L., Sulkowski, S., Kottler, M.L., and Wolczynski, S. (2009). The significance of the expression of ERRalpha as a potential biomarker in breast cancer. *J Steroid Biochem Mol Biol* **113**: 127-133.

Jenness, R. (1974). The composition of milk. In: Lactation: A Comprehensive Treatise, B.L. Larson, and V.R. Smith, eds. (New York: Academic Press, Inc.), pp. 3-107.

Jenness, R. (1979). The composition of human milk. *Semin Perinatol* **3**: 225-239.

Jenness, R. (1986). Lactational performance of various mammalian species. *J Dairy Sci* **69**: 869-885.

Jering, H., and Tschesche, H. (1976). Replacement of lysine by arginine, phenylalanine and tryptophan in the reactive site of the bovine trypsin-kallikrein inhibitor (Kunitz) and change of the inhibitory properties. *Eur J Biochem* **61**: 453-463.

Johnson, K.R., Leight, J.L., Weaver, V.M., YuLi, W., and Dennis, E.D. (2007). Demystifying the effects of a three-dimensional microenvironment in tissue morphogenesis. In: Methods in Cell Biology (Academic Press), pp. 547-583.

Johnson, T.L., Fujimoto, B.A., Jimenez-Flores, R., and Peterson, D.G. (2010). Growth hormone alters lipid composition and increases the abundance of casein and lactalbumin mRNA in the MAC-T cell line. *J Dairy Res* **77**: 199-204.

Johnston, R.F., Pickett, S.C., and Barker, D.L. (1990). Autoradiography using storage phosphor technology. *Electrophoresis* **11**: 355-360.

Jolivet, G., L'Hotte, C., Pierre, S., Tourkine, N., and Houdebine, L.M. (1996). A MGF/STAT5 binding site is necessary in the distal enhancer for high prolactin induction of transfected rabbit αs1-casein-CAT gene transcription. *FEBS Lett* **389**: 257-262.

Jolivet, G., Meusnier, C., Chaumaz, G., and Houdebine, L.M. (2001). Extracellular matrix regulates αs1-casein gene expression in rabbit primary mammary cells and CCAAT enhancer binding protein (C/EBP) binding activity. *J Cell Biochem* **82**: 371-386.

Jolles, J., Schoentgen, F., Croizier, G., Croizier, L., and Jolles, P. (1979). Insect lysozymes from three species of Lepidoptera: their structural relatedness to the C (chicken) type lysozyme. *J Mol Evol* **14**: 267-271.

Jolles, P., and Jolles, J. (1984). What's new in lysozyme research? Always a model system, today as yesterday. *Mol Cell Biochem* **63**: 165-189.

Jolles, P., Loucheux-Lefebvre, M.-H., and Henschen, A. (1978). Structural relatedness of κ-casein and fibrinogen γ-chain. *J Mol Evol* **V11**: 271-277.

Jordan, S.P., Mao, S.S., Lewis, S.D., and Shafer, J.A. (1992). Reaction pathway for inhibition of blood coagulation factor Xa by tick anticoagulant peptide. *Biochemistry* **31**: 5374-5380.

Jørgensen, A.L.W., Juul-Madsen, H.R., and Stagsted, J. (2010). Colostrum and bioactive, colostral peptides differentially modulate the innate immune response of intestinal epithelial cells. *Journal of Peptide Science* **16**: 21-30.

Joss, J., Molloy, M., Hinds, L., and Deane, E. (2007). Proteomic analysis of early lactation milk of the tammar wallaby (*Macropus eugenii*). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **2**: 150-164.

Joss, J.L., Molloy, M.P., Hinds, L., and Deane, E. (2009). A longitudinal study of the protein components of marsupial milk from birth to weaning in the tammar wallaby (*Macropus eugenii*). *Dev Comp Immunol* **33**: 152-161.

Jost, T., Lacroix, C., Braegger, C.P., Rochat, F., and Chassard, C. (2014). Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environ Microbiol* **16**: 2891-2904.

Juven-Gershon, T., Hsu, J.-Y., Theisen, J.W.M., and Kadonaga, J.T. (2008). The RNA polymerase II core promoter - the gateway to transcription. *Current Opinion in Cell Biology* **20**: 253-259.

Kahl, S., Bitman, J., Capuco, A.V., and Keys, J.E. (1991). Effect of lactational intensity on extrathyroidal 5'-deiodinase activity in rats. *J Dairy Sci* **74**: 811-818.

Kanai, A., Nonomura, N., Yoshimura, M., and Oka, T. (1993). DNA-binding proteins and their *cis*-acting sites controlling hormonal induction of a mouse β-casein: CAT fusion protein in mammary epithelial cells. *Gene* **126**: 195-201.

Kane, J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current opinion in biotechnology* **6**: 494-500.

Kassell, B., and Laskowski, M. (1956). The comparative resistance to pepsin of six naturally occurring trypsin inhibitors. *J Biol Chem* **219**: 203-210.

Kataoka, H., Itoh, H., and Koono, M. (2002). Emerging multifunctional aspects of cellular serine proteinase inhibitors in tumor progression and tissue regeneration. *Pathology International* **52**: 89-102.

Katz, E., and Streuli, C.H. (2007). The extracellular matrix as an adhesion checkpoint for mammary epithelial function. *Int J Biochem Cell Biol* **39**: 715-726.

Kawaguchi, T., Qin, L., Shimomura, T., Kondo, J., Matsumoto, K., Denda, K., and Kitamura, N. (1997). Purification and cloning of hepatocyte growth factor activator inhibitor type 2, a Kunitz-type serine protease inhibitor. *J Biol Chem* **272**: 27558-27564.

Kawasaki, K., Lafont, A.-G., and Sire, J.-Y. (2011). The Evolution of Milk Casein Genes from Tooth Genes before the Origin of Mammals. *Molecular Biology and Evolution* **28**: 2053-2061.

Kawasaki, K., and Weiss, K.M. (2003). Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc Natl Acad Sci U S A* **100**: 4060-4065.

Kelly, P.A., Bachelot, A., Kedzia, C., Hennighausen, L., Ormandy, C.J., Kopchick, J.J., and Binart, N. (2002). The role of prolactin and growth hormone in mammary gland development. *Mol Cell Endocrinol* **197**: 127-131.

Kelly, P.A., Boutin, J.M., Jolicoeur, C., Okamura, H., Shirota, M., Edery, M., Dusanter-Fourt, I., and Djiane, J. (1989). Purification, cloning, and expression of the prolactin receptor. *Biol Reprod* **40**: 27-32.

Kelly, P.A., Djiane, J., Postel-Vinay, M.C., and Edery, M. (1991). The prolactin/growth hormone receptor family. *Endocr Rev* **12**: 235-251.

Keys, J., Guidry, A., and Cifrian, E. (1997). The use of flow cytometry and fluorescein-labeled antibodies to measure specific milk proteins in bovine mammary epithelial cells. *In Vitro Cellular & Developmental Biology - Animal* **33**: 201-205.

Khaled, W.T., Read, E.K., Nicholson, S.E., Baxter, F.O., Brennan, A.J., Came, P.J., Sprigg, N., McKenzie, A.N., and Watson, C.J. (2007). The IL-4/IL-13/Stat6 signalling

pathway promotes luminal mammary epithelial cell development. *Development* **134**: 2739-2750.

Khalil, E., Digby, M.R., O'Donnell, P., and Nicholas, K.R. (2008). Changes in milk protein composition during acute involution at different phases of tammar wallaby (*Macropus eugenii*) lactation. *Comp Biochem Physiol B Biochem Mol Biol* **151**: 64-69.

Kingston, B., and Brett, R. (2005). Protein expression. In: Current Protocols in Molecular Biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. (John Wiley and Sons, Inc.), pp. 16.10.11-16.10.14.

Kingston, I.B., and Anderson, S. (1986). Sequences encoding two trypsin inhibitors occur in strikingly similar genomic environments. *Biochem J* **233**: 443-450.

Kirkman, S.P., Bester, M.N., Hofmeyr, G.J.G., Pistorius, P.A., and Makhado, A.B. (2002). Pup growth and maternal attendance patterns in Subantarctic fur seals. *Afr Zool* **37**: 13-19.

Klauser, R., Cechova, D., and Tschesche, H. (1978). The carbohydrate linkage site of cow colostrum trypsin inhibitor. *Hoppe Seylers Z Physiol Chem* **359**: 173-180.

Kleinman, H.K., Luckenbill-Edds, L., Cannon, F.W., and Sephel, G.C. (1987). Use of extracellular matrix components for cell culture. *Analytical Biochemistry* **166**: 1-13.

Kleinman, H.K., McGarvey, M.L., Hassell, J.R., Star, V.L., Cannon, F.B., Laurie, G.W., and Martin, G.R. (1986). Basement membrane complexes with biological activity. *Biochemistry* **25**: 312-318.

Kleinman, R.E., and Walker, W.A. (1979). The enteromammary immune system: an important new concept in breast milk host defense. *Digestive Diseases and Sciences* **24**: 876-882.

Knappik, A., and Pluckthun, A. (1994). An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments. *Biotechniques* **17**: 754-761.

Ko, K., Ahn, M.H., Song, M., Choo, Y.K., Kim, H.S., and Joung, H. (2008). Glyco-engineering of biotherapeutic proteins in plants. *Mol Cells* **25**: 494-503.

Koepfli, K.P., Paten, B., Genome, K.C.o.S., and O'Brien, S.J. (2015). The Genome 10K Project: a way forward. *Annu Rev Anim Biosci* **3**: 57-111.

Kolb, A.F., Albang, R., Brem, G., Erfle, V., Gunzburg, W.H., and Salmons, B. (1995). Characterization of a protein that binds a negative regulatory element in the mammary-specific whey acidic protein promoter. *Biochem Biophys Res Commun* **217**: 1045-1052.

Kolb, A.F., Gunzburg, W.H., Albang, R., Brem, G., Erfle, V., and Salmons, B. (1994). Negative regulatory elements in the mammary specific whey acidic protein promoter. *Journal of Cellular Biochemistry* **56**: 245-261.

Kontopidis, G., Holt, C., and Sawyer, L. (2004). Invited review: β-lactoglobulin: binding properties, structure, and function. *J Dairy Sci* **87**: 785-796.

Kowalczyk, K., Daiss, J., Halpern, J., and Roth, T.F. (1985). Quantitation of maternal-fetal IgG transport in the chicken. *Immunology* **54**: 755-762.

Kress, C., Ballester, M., Devinoy, E., and Rijnkels, M. (2010). Epigenetic modifications in 3D: nuclear organization of the differentiating mammary epithelial cell. *J Mammary Gland Biol Neoplasia* **15**: 73-83.

Kress, L.F., Martin, S.R., and Laskowski, M., Sr. (1971). Isolation of isoinhibitors of trypsin from porcine colostrum. *Biochim Biophys Acta* **229**: 836-844.

Krokoszynska, I., Dadlez, M., and Otlewski, J. (1998). Structure of single-disulfide variants of bovine pancreatic trypsin inhibitor (BPTI) as probed by their binding to bovine β-trypsin. *Journal of Molecular Biology* **275**: 503-513.

Kruse, P.E. (1983). The importance of colostral immunoglobulins and their absorption from the intestine of the newborn animals. *Ann Rech Vet* **14**: 349-353.

Kumagai, Y., and Akira, S. (2010). Identification and functions of pattern-recognition receptors. *Journal of Allergy and Clinical Immunology* **125**: 985-992.

Kumar, S., Dudley, J., Nei, M., and Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* **9**: 299-306.

Kunitz, M., and Northrop, J.H. (1936). Isolation from beef pancreas of crystalline trypsinogen, trypsin, a trypsin inhibitor, and an inhibitor-trypsin compound. *J Gen Physiol* **19**: 991-1007.

Kuntz, S., Kunz, C., and Rudloff, S. (2009). Oligosaccharides from human milk induce growth arrest via G2/M by influencing growth-related cell cycle genes in intestinal epithelial cells. *Br J Nutr* **101**: 1306-1315.

Kuntz, S., Rudloff, S., and Kunz, C. (2008). Oligosaccharides from human milk influence growth-related characteristics of intestinally transformed and non-transformed intestinal cells. *Br J Nutr* **99**: 462-471.

Kunz, C., Kuntz, S., and Rudloff, S. (2014). Bioactivity of human milk oligosaccharides. In: Food Oligosaccharides: Production, Analysis and Bioactivity, J.F. Moreno, and M.L. Sanz, eds. (John Wiley & Sons, Ltd), pp. 1-20.

Kunz, C., and Rudloff, S. (1993). Biological functions of oligosaccharides in human milk. *Acta Paediatr* **82**: 903-912.

Kuss, A.W., Gogol, J., and Geidermann, H. (2003). Associations of a polymorphic AP-2 binding site in the 5'-flanking region of the bovine beta-lactoglobulin gene with milk proteins. *J Dairy Sci* **86**: 2213-2218.

Kwek, J., De Iongh, R., Nicholas, K., and Familari, M. (2009a). Molecular insights into evolution of the vertebrate gut: focus on stomach and parietal cells in the marsupial, *Macropus eugenii*. *J Exp Zool B Mol Dev Evol* **312**: 613-624.

Kwek, J.H., Iongh, R.D., Digby, M.R., Renfree, M.B., Nicholas, K.R., and Familari, M. (2009b). Cross-fostering of the tammar wallaby (*Macropus eugenii*) pouch young accelerates fore-stomach maturation. *Mechanisms of Development* **126**: 449-463.

Kwek, J.H., Wijesundera, C., Digby, M.R., and Nicholas, K.R. (2007). The endocrine regulation of milk lipid synthesis and secretion in tammar wallaby (*Macropus eugenii*). *Biochim Biophys Acta* **1770**: 48-54.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Laird, J.E., Jack, L., Hall, L., Boulton, A.P., Parker, D., and Craig, R.K. (1988). Structure and expression of the guinea-pig α-lactalbumin gene. *Biochem J* **254**: 85-94.

Lamming, G.E., ed. (1994). Marshall's Physiology of Reproduction. Volume 3 Pregnancy and Lactation, Fourth edition edn (Netherlands: Springer).

Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759-1764.

Langer, P. (2008). The phases of maternal investment in eutherian mammals. *Zoology* **111**: 148-162.

Larson, B.L. (1992). Immunoglobulins of the mammary secretions. In: Advanced Dairy Chemistry Volume 1: Proteins, P.F. Fox, ed. (London: Elsevier Science Publishers Ltd), pp. 231-254.

Larson, B.L., Heary, H.L., Jr., and Devery, J.E. (1980). Immunoglobulin production and transport by the mammary gland. *J Dairy Sci* **63**: 665-671.

Lascelles, A.K. (1979). The immune system on the ruminant mammary gland and its role in the control of mastitis. *J Dairy Sci* **62**: 154-167.

Laskowski, M., Jr. (1986). Protein inhibitors of serine proteinases - mechanism and classification. *Adv Exp Med Biol* **199**: 1-17.

Laskowski, M., Jr., and Laskowski, M. (1950). *Federation Proceedings* **9**: 194.

Laskowski, M., Jr., and Laskowski, M. (1951). Crystalline trypsin inhibitor from colostrum. *J Biol Chem* **190**: 563-573.

Laskowski, M., Jr., Mars, P.H., and Laskowski, M. (1952). Comparison of trypsin inhibitor from colostrum with other crystalline trypsin inhibitors. *J Biol Chem* **198**: 745-752.

Laskowski, M., Kassell, B., and Hagerty, G. (1957). A crystalline trypsin inhibitor from swine colostrum. *Biochim Biophys Acta* **24**: 300-305.

Laskowski, M., and Kato, I. (1980). Protein inhibitors of proteinases. *Annual Review of Biochemistry* **49**: 593-626.

Laskowski, M., and Qasim, M.A. (2000). What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim Biophys Acta* **1477**: 324-337.

LaVallie, E.R., and McCoy, J.M. (1995). Gene fusion expression systems in Escherichia coli. *Current opinion in biotechnology* **6**: 501-506.

Leblond, C.P., and Inoue, S. (1989). Structure, composition, and assembly of basement membrane. *Am J Anat* **185**: 367-390.

Lebrun, J.J., Ali, S., Sofer, L., Ullrich, A., and Kelly, P.A. (1994). Prolactin-induced proliferation of Nb2 cells involves tyrosine phosphorylation of the prolactin receptor and its associated tyrosine kinase JAK2. *J Biol Chem* **269**: 14021-14026.

Lechner, J., Welte, T., and Doppler, W. (1997a). Mechanism of interaction between the glucocorticoid receptor and Stat5: role of DNA-binding. *Immunobiology* **198**: 112-123.

Lechner, J., Welte, T., Tomasi, J.K., Bruno, P., Cairns, C., Gustafsson, J., and Doppler, W. (1997b). Promoter-dependent synergy between glucocorticoid receptor and Stat5 in the activation of β-casein gene transcription. *J Biol Chem* **272**: 20954-20960.

Lee, A.K., and Cockburn, A. (1985). Evolutionary Ecology of Marsupials (New York, Cambridge: Cambridge University Press).

Lee, C.S., and Oka, T. (1992a). A pregnancy-specific mammary nuclear factor involved in the repression of the mouse β-casein gene transcription by progesterone. *J Biol Chem* **267**: 5797-5801.

Lee, C.S., and Oka, T. (1992b). Progesterone regulation of a pregnancy-specific transcription repressor to β-casein gene promoter in mouse mammary gland. *Endocrinology* **131**: 2257-2262.

Lee, E.Y., Parry, G., and Bissell, M.J. (1984). Modulation of secreted proteins of mouse mammary epithelial cells by the collagenous substrata. *J Cell Biol* **98**: 146-155.

Lee, H.J., Hinshelwood, R.A., Bouras, T., Gallego-Ortega, D., Valdes-Mora, F., Blazek, K., Visvader, J.E., Clark, S.J., and Ormandy, C.J. (2011). Lineage specific methylation of the Elf5 promoter in mammary epithelial cells. *Stem Cells* **29**: 1611-1619.

Lee, H.J., and Ormandy, C.J. (2012). Elf5, hormones and cell fate. *Trends Endocrinol Metab* **23**: 292-298.

Lefèvre, C.M., Digby, M.R., Whitley, J.C., Strahm, Y., and Nicholas, K.R. (2007). Lactation transcriptomics in the Australian marsupial, *Macropus eugenii*: transcript sequencing and quantification. *BMC Genomics* **8**: 417.

Lefevre, C.M., Sharp, J.A., and Nicholas, K.R. (2010). Evolution of lactation: ancient origin and extreme adaptations of the lactation system. *Annu Rev Genom Hum G* **11**: 219-238.

Lefèvre, C.M., Sharp, J.A., and Nicholas, K.R. (2010). Evolution of lactation: ancient origin and extreme adaptations of the lactation system. *Annu Rev Genomics Hum Genet* **11**: 219-238.

Legraverend, C., Antonson, P., Flodby, P., and Xanthopoulos, K.G. (1993). High level activity of the mouse CCAAT/enhancer binding protein (C/EBPα) gene promoter involves autoregulation and several ubiquitous transcription factors. *Nucleic Acids Res* **21**: 1735-1742.

Lemay, D.G., Neville, M.C., Rudolph, M.C., Pollard, K.S., and German, J.B. (2007). Gene regulatory networks in lactation: identification of global principles using bioinformatics. *BMC Syst Biol* **1**: 56.

Lemay, D.G., Pollard, K.S., Martin, W.F., Freeman Zadrowski, C., Hernandez, J., Korf, I., German, J.B., and Rijnkels, M. (2013). From genes to milk: genomic organization and epigenetic regulation of the mammary transcriptome. *PLoS One* **8**: e75030.

Lemon, M., and Barker, S. (1967). Changes in milk composition of the red kangaroo, *Megaleia rufa* (Desmarest), during lactation. *Aust J Exp Biol Med Sci* **45**: 213-219.

Lenton, S., Nylander, T., Teixeira, S.C., and Holt, C. (2015). A review of the biology of calcium phosphate sequestration with special reference to milk. *Dairy Sci Technol* **95**: 3-14.

Leonard, W.J., and O'Shea, J.J. (1998). JAKS and STATS: Biological Implications. *Annual Review of Immunology* **16**: 293-322.

LeRoith, D. (2000). Insulin-like growth factor I receptor signaling - overlapping or redundant pathways? *Endocrinology* **141**: 1287-1288.

Lesueur, L., Edery, M., Ali, S., Paly, J., and Kelly, P. (1991). Comparison of long and short forms of the prolactin receptor on prolactin-induced milk protein gene transcription. *Proc Natl Acad Sci U S A* **88**: 824-828.

Leung, K.-C., Johannsson, G., Leong, G.M., and Ho, K.K.Y. (2004). Estrogen Regulation of Growth Hormone Action. *Endocr Rev* **25**: 693-721.

Levine, J.F., and Stockdale, F.E. (1985). Cell-cell interactions promote mammary epithelial cell differentiation. *J Cell Biol* **100**: 1415-1422.

Levy, D.E., and Darnell, J.E., Jr. (2002). Stats: transcriptional control and biological impact. *Nat Rev Mol Cell Biol* **3**: 651-662.

Li, M.L., Aggeler, J., Farson, D.A., Hatier, C., Hassell, J., and Bissell, M.J. (1987). Influence of a reconstituted basement membrane and its components on casein gene expression and secretion in mouse mammary epithelial cells. *Proc Natl Acad Sci U S A* **84**: 136-140.

Li, R., Pei, H., and Watson, D.K. (2000). Regulation of Ets function by protein - protein interactions. *Oncogene* **19**: 6514-6523.

Li, S., and Rosen, J.M. (1994a). Distal regulatory elements required for rat whey acidic protein gene expression in transgenic mice. *J Biol Chem* **269**: 14235-14243.

Li, S., and Rosen, J.M. (1994b). Glucocorticoid regulation of rat whey acidic protein gene expression involves hormone-induced alterations of chromatin structure in the distal promoter region. *Mol Endocrinol* **8**: 1328-1335.

Li, S., and Rosen, J.M. (1995). Nuclear factor I and mammary gland factor (STAT5) play a critical role in regulating rat whey acidic protein gene expression in transgenic mice. *Mol Cell Biol* **15**: 2063-2070.

Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J., and Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein Expr Purif* **41**: 98-105.

Lin, C.Q., and Bissell, M.J. (1999). Multi-faceted regulation of cell differentiation by extracellular matrix. In: Biology of Lactation, J. Martinet, L.M. Houdebine, and H.H. Head, eds. (Paris: INRA), pp. 77-92.

Lincoln, D.W., and Renfree, M.B. (1981). Mammary gland growth and milk ejection in the agile wallaby, *Macropus agilis*, displaying concurrent asynchronous lactation. *J Reprod Fertil* **63**: 193-203.

Linskens, M.H., Harley, C.B., West, M.D., Campisi, J., and Hayflick, L. (1995). Replicative senescence and cell death. *Science* **267**: 17.

Linzell, J.L., and Peaker, M. (1971). Mechanism of milk secretion. *Physiol Rev* **51**: 564-597.

Linzell, J.L., Peaker, M., and Taylor, J.C. (1975). The effects of prolactin and oxytocin on milk secretion and on the permeability of the mammary epithelium in the rabbit. *J Physiol* **253**: 547-563.

Liu, X., Gallego, M.I., Smith, G.H., Robinson, G.W., and Hennighausen, L. (1998). Functional rescue of Stat5a-null mammary tissue through the activation of compensating signals including Stat5b. *Cell Growth Differ* **9**: 795-803.

Liu, X., Robinson, G.W., Gouilleux, F., Groner, B., and Hennighausen, L. (1995). Cloning and expression of Stat5 and an additional homologue (Stat5b) involved in prolactin signal transduction in mouse mammary tissue. *Proc Natl Acad Sci U S A* **92**: 8831-8835.

Liu, X., Robinson, G.W., and Hennighausen, L. (1996). Activation of Stat5a and Stat5b by tyrosine phosphorylation is tightly linked to mammary gland differentiation. *Mol Endocrinol* **10**: 1496-1506.

Lizcano, J.M., and Alessi, D.R. (2002). The insulin signalling pathway. *Current Biology* **12**: R236-R238.

Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M. (2012). SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial cell factories [electronic resource]* **11**: 56.

Locker, J. (1993). Transcription controls: *cis*-elements and *trans*-factors. In: Gene Transcription A Practical Approach, B.D. Hames, and S.J. Higgins, eds. (New York: Oxford University Press), pp. 2103–2121.

Logdberg, L., and Wester, L. (2000). Immunocalins: a lipocalin subfamily that modulates immune and inflammatory responses. *Biochim Biophys Acta* **1482**: 284-297.

Long, C.A. (1969). Origin and evolution of mammary glands. *Bioscience* **19**: 519-523.

Low, B.S. (1978). Environmental uncertainty and the parental strategies of marsupials and placentals. *American Naturalist* **112**: 197-213.

Lubon, H., and Hennighausen, L. (1987). Nuclear proteins from lactating mammary glands bind to the promoter of a milk protein gene. *Nucleic Acids Res* **15**: 2103-2121.

Lubon, H., and Hennighausen, L. (1988). Conserved region of the rat α-lactalbumin promoter is a target site for protein binding *in vitro*. *Biochem J* **256**: 391-396.

Luo, Z.-X., Meng, Q.-J., Ji, Q., Liu, D., Zhang, Y.-G., and Neander, A.I. (2015). Evolutionary development in basal mammaliaforms as revealed by a docodontan. *Science* **347**: 760-764.

Luo, Z.X., Yuan, C.X., Meng, Q.J., and Ji, Q. (2011). A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature* **476**: 442-445.

Macias, H., and Hinck, L. (2012). Mammary gland development. *Wiley Interdiscip Rev Dev Biol* **1**: 533-557.

MacLean, J.A., 2nd, Chakrabarty, A., Xie, S., Bixby, J.A., Roberts, R.M., and Green, J.A. (2003). Family of Kunitz proteins from trophoblast: expression of the trophoblast Kunitz domain proteins (TKDP) in cattle and sheep. *Mol Reprod Dev* **65**: 30-40.

MacLean, J.A., 2nd, Roberts, R.M., and Green, J.A. (2004). Atypical Kunitz-type serine proteinase inhibitors produced by the ruminant placenta. *Biol Reprod* **71**: 455-463.

Maher, F., and Nicholas, K.R. (1987). Pituitary-induced lactation in mammary gland explants from the pregnant tammar (*Macropus eugenii*): a negative role for cyclic AMP. *Comp Biochem Physiol A* **87**: 1107-1117.

Maina, C.V., Riggs, P.D., Grandea, A.G., 3rd, Slatko, B.E., Moran, L.S., Tagliamonte, J.A., McReynolds, L.A., and Guan, C.D. (1988). An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **74**: 365-373.

Malewski, T. (1998). Computer analysis of distribution of putative *cis*- and *trans*-regulatory elements in milk protein gene promoters. *Biosystems* **45**: 29-44.

Malewski, T., Gajewska, M., Zebrowska, T., and Zwierzchowski, L. (2002). Differential induction of transcription factors and expression of milk protein genes by prolactin and growth hormone in the mammary gland of rabbits. *Growth Horm IGF Res* **12**: 41-53.

Manhes, C., Goffin, V., Kelly, P.A., and Touraine, P. (2005). Autocrine prolactin as a promotor of mammary tumour growth. *J Dairy Res* **72 Spec No**: 58-65.

Maqsood, M.I., Matin, M.M., Bahrami, A.R., and Ghasroldasht, M.M. (2013). Immortality of cell lines: challenges and advantages of establishment. *Cell Biol Int* **37**: 1038-1045.

Marks, C.B., Vasser, M., Ng, P., Henzel, W., and Anderson, S. (1986). Production of native, correctly folded bovine pancreatic trypsin inhibitor by *Escherichia coli*. *J Biol Chem* **261**: 7115-7118.

Marlor, C.W., Delaria, K.A., Davis, G., Muller, D.K., Greve, J.M., and Tamburini, P.P. (1997). Identification and cloning of human placental bikunin, a novel serine protease inhibitor containing two Kunitz domains. *J Biol Chem* **272**: 12202-12208.

Maroulakou, I.G., and Bowe, D.B. (2000). Expression and function of Ets transcription factors in mammalian development: a regulatory network. *Oncogene* **19**: 6432-6442.

Marshall, A.M., Nommsen-Rivers, L.A., Hernandez, L.L., Dewey, K.G., Chantry, C.J., Gregerson, K.A., and Horseman, N.D. (2010). Serotonin transport and metabolism in the mammary gland modulates secretory activation and involution. *J Clin Endocrinol Metab* **95**: 837-846.

Martin, M.J., Martin-Sosa, S., and Hueso, P. (2002). Binding of milk oligosaccharides by several enterotoxigenic *Escherichia coli* strains isolated from calves. *Glycoconj J* **19**: 5-11.

Martin, P., Cebo, C., and Miranda, G. (2013). Interspecies Comparison of Milk Proteins: Quantitative Variability and Molecular Diversity. In: Advanced Dairy Chemistry Volume 1: Proteins, P.L.H. McSweeney, and P.F. Fox, eds. (New York, Heidelberg, Dordrecht, London: Springer), pp. 387-429.

Martinet, J., and Houdebine, L.M. (1999). Mammary gland, mammogenesis, growth factors, lactogenesis. In: Biology of Lactation, J. Martinet, L.M. Houdebine, and H.H. Head, eds. (Paris: INRA), pp. 1-27.

Marx, J. (1994). Chromosome ends catch fire. *Science* **265**: 1656-1658.

Mather, I.H., and Keenan, T.W. (1975). Studies on the structure of milk fat globule membrane. *J Membr Biol* **21**: 65-85.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H.*, et al.* (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* **42**: D142-147.

Mather, I.H., and Keenan, T.W. (1998). Origin and secretion of milk lipids. *J Mammary Gland Biol Neoplasia* **3**: 259-273.

Matitashvili, E., Bramley, A.J., and Zavizion, B. (1997). An *in vitro* approach to ruminant mammary gland biology. *Biotechnol Adv* **15**: 17-41.

Matsuda, M., Imaoka, T., Vomachka, A.J., Gudelsky, G.A., Hou, Z., Mistry, M., Bailey, J.P., Nieport, K.M., Walther, D.J., Bader, M.*, et al.* (2004). Serotonin regulates mammary gland development via an autocrine-paracrine loop. *Dev Cell* **6**: 193-203.

Matusik, R.J., and Rosen, J.M. (1978). Prolactin induction of casein mRNA in organ culture. A model system for studying peptide hormone regulation of gene expression. *J Biol Chem* **253**: 2343-2347.

McAllan, B. (2003). Timing of reproduction in carnivorous marsupials. In: Predators with pouches: the biology of carnivorous marsupials, M.E. Jones, C.R. Dickman, and M. Archer, eds. (Collingwood, VIC: CSIRO), pp. 147-168.

McClellan, H.L., Miller, S.J., and Hartmann, P.E. (2008). Evolution of lactation: nutrition v. protection with special reference to five mammalian species. *Nutr Res Rev* **21**: 97-116.

McConnell, S.J., and Hinds, L.A. (1985). Effect of pinealectomy on plasma melatonin, prolactin and progesterone concentrations during seasonal reproductive quiescence in the tammar, *Macropus eugenii*. *J Reprod Fertil* **75**: 433-440.

McFadden, T.B., Besser, T.E., and Barrington, G.M. (1997). Regulation of immunoglobulin transfer into mammary secretions of ruminants. In: Milk Composition, Production and Biotechnology, R.A.S. Welch, D.J.W. Burns, S.R. Davis, A.I. Popay, and C.G. Prosser, eds. (New York: CAB International), pp. 133-152.

McGrath, M.F. (1987). A novel system for mammary epithelial cell culture. *J Dairy Sci* **70**: 1967-1980.

McKnight, R.A., Burdon, T., Pursel, V.G., Shamay, A., Wall, R.J., and Hennighausen, L. (1991). The whey acidic protein. In: Genes, Oncogenes and Hormones: Advances in Cellular and Molecular Biology of Breast Cancer, R.E. Dickson, and M.E. Lippman, eds. (Norwell, MA, USA: Kluwer Academic Publishers), pp. 399-412.

McManaman, J.L., and Neville, M.C. (2003). Mammary physiology and milk secretion. *Drug Excretion in Breast Milk: Mechanisms, Models and Drug Delivery Implications for the Infant* **55**: 629-641.

McManaman, J.L., Reyland, M.E., and Thrower, E.C. (2006). Secretion and fluid transport mechanisms in the mammary gland: comparisons with the exocrine pancreas and the salivary gland. *J Mammary Gland Biol* **11**: 249-268.

McNulty, D.E., Claffee, B.A., Huddleston, M.J., Porter, M.L., Cavnar, K.M., and Kane, J.F. (2003). Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Expr Purif* **27**: 365-374.

Mead, R.A. (1993). Embryonic diapause in vertebrates. *J Exp Zool* **266**: 629-641.

Meier, V.S., and Groner, B. (1994). The nuclear factor YY1 participates in repression of the beta-casein gene promoter in mammary epithelial cells and is counteracted by mammary gland factor during lactogenic hormone induction. *Mol Cell Biol* **14**: 128-137.

Meng, J. (2014). Mesozoic mammals of China: implications for phylogeny and early evolution of mammals. *National Science Review* **1**: 521-542.

Menzies, B.R. (2008). Endocrine control of growth in the developing marsupial, *Macropus eugenii*. PhD Thesis, Department of Zoology (Melbourne: University of Melbourne).

Menzies, B.R., Shaw, G., Fletcher, T.P., Pask, A.J., and Renfree, M.B. (2012). Maturation of the growth axis in marsupials occurs gradually during post-natal life and over an equivalent developmental stage relative to eutherian species. *Mol Cell Endocrinol* **349**: 189-194.

Menzies, B.R., Shaw, G., Fletcher, T.P., and Renfree, M.B. (2007). Perturbed growth and development in marsupial young after reciprocal cross-fostering between species. *Reprod Fertil Dev* **19**: 976-983.

Menzies, K., K, Lee, H., J, Lefèvre, C., Ormandy, C., J, Macmillan, K., L, and Nicholas, K., R (2010a). Insulin, a key regulator of hormone responsive milk protein synthesis

during lactogenesis in murine mammary explants. *Functional & Integrative Genomics* **10**: 87-95.

Menzies, K.K., Lefevre, C., Macmillan, K.L., and Nicholas, K.R. (2009). Insulin regulates milk protein synthesis at multiple levels in the bovine mammary gland. *Funct Integr Genomics* **9**: 197-217.

Menzies, K.K., and Nicholas, K.R. (2007). The expression of β-1,3 galactosyltransferase and β-1,4 galactosyltransferase enzymatic activities in the mammary gland of the tammar wallaby (*Macropus eugenii*) during early lactation. *Biochimica Et Biophysica Acta-General Subjects* **1770**: 115-120.

Mepham, T.B. (1992). Biosynthesis of milk protein. In: Advanced Dairy Chemistry Volume 1: Proteins, P.F. Fox, ed. (London: Elsevier Science Publishers Ltd), pp. 491-543.

Mercier, J.C., and Grosclaude, F. (1999). The molecular genetics of milk proteins and their genes. In: Biology of Lactation, J. Martinet, L.M. Houdebine, and H.H. Head, eds. (Paris: INRA), pp. 369-400.

Mercier, J.C., and Vilotte, J.L. (1993). Structure and function of milk protein genes. *J Dairy Sci* **76**: 3079-3098.

Merril, C.R., Dunau, M.L., and Goldman, D. (1981). A rapid sensitive silver stain for polypeptides in polyacrylamide gels. *Anal Biochem* **110**: 201-207.

Messer, M., and Green, B. (1979). Milk carbohydrates of marsupials. II. Quantitative and qualitative changes in milk carbohydrates during lactation in the tammar wallaby (Macropus eugenii). *Aust J Biol Sci* **32**: 519-531.

Messer, M., Griffiths, M., and Green, B. (1984). Changes in milk carbohydrates and electrolytes during early lactation in the tammar wallaby, *Macropus eugenii*. *Australian Journal of Biological Sciences* **37**: 1-6.

Messer, M., and Mossop, G.S. (1977). Milk carbohydrates of marsupials. I. Partial separation and characterization of neutral milk oligosaccharides of the eastern grey kangaroo. *Australian Journal of Biological Sciences* **30**: 379-388.

Messer, M., Weiss, A.S., Shaw, D.C., and Westerman, M. (1998). Evolution of the monotremes: phylogenetic relationship to marsupials and eutherians, and estimation of divergence dates based on α-lactalbumin amino acid sequences. *Journal of Mammalian Evolution* **5**: 95-105.

Miller, R.D., and Hansen, V.L. (2014). Comparative Immunoglobulin Genetics. In: Marsupial and Monotreme Immunoglobulin Genetics, A.K. Kaushik, and Y. Pasman, eds. (Oakville: Apple Academic Press Inc.), pp. 85-105.

Miner, J.N., and Yamamoto, K.R. (1991). Regulatory crosstalk at composite response elements. *Trends in biochemical sciences* **16**: 423-426.

Mink, S., Hartig, E., Jennewein, P., Doppler, W., and Cato, A.C. (1992). A mammary cell-specific enhancer in mouse mammary tumor virus DNA is composed of multiple regulatory elements including binding sites for CTF/NFI and a novel transcription factor, mammary cell-activating factor. *Mol Cell Biol* **12**: 4906-4918.

Modepalli, V., Kumar, A., Hinds, L.A., Sharp, J.A., Nicholas, K.R., and Lefevre, C. (2014). Differential temporal expression of milk miRNA during the lactation cycle of the marsupial tammar wallaby (*Macropus eugenii*). *BMC Genomics* **15**: 1012.

Molina, J.R., and Adjei, A.A. (2006). The Ras/Raf/MAPK Pathway. *Journal of Thoracic Oncology* **1**: 7-9.

Monks, A., Scudiero, D., Skehan, P., Shoemaker, R., Paull, K., Vistica, D., Hose, C., Langley, J., Cronise, P., Vaigro-Wolff, A.*, et al.* (1991). Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. *J Natl Cancer Inst* **83**: 757-766.

Morrow, A.L., Ruiz-Palacios, G.M., Jiang, X., and Newburg, D.S. (2005). Human-milk glycans that inhibit pathogen binding protect breast-feeding infants against infectious diarrhea. *J Nutr* **135**: 1304-1307.

Mossman, H.W. (1987). Vertebrate Fetal Membranes. (Brunswick, NJ: Rutgers University Press).

Motejadded, H., Kranz, B., Berensmeier, S., Franzreb, M., and Altenbuchner, J. (2010). Expression, one-step purification, and immobilization of HaloTag(TM) fusion proteins on chloroalkane-functionalized magnetic beads. *Appl Biochem Biotechnol* **162**: 2098-2110.

Mukhopadhyay, S.S., Wyszomierski, S.L., Gronostajski, R.M., and Rosen, J.M. (2001). Differential interactions of specific nuclear factor I isoforms with the glucocorticoid receptor and STAT5 in the cooperative regulation of WAP gene transcription. *Molecular and Cellular Biology* **21**: 6859-6869.

Murray, P.J. (2007). The JAK-STAT signaling pathway: input and output integration. *J Immunol* **178**: 2623-2629.

Murray, S.C., Haenni, S., Howe, F.S., Fischl, H., Chocian, K., Nair, A., and Mellor, J. (2015). Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Research*. epub: June 29th 2016, doi: 10.1093/nar/gkv666.

Muschler, J., and Streuli, C.H. (2010). Cell–Matrix Interactions in Mammary Gland Development and Breast Cancer. *Cold Spring Harbor Perspectives in Biology* **2**.

Muszynska, G., and Cechova, D. (1970). Effect of modification of lysine residues of cow colostrum trypsin inhibitor on its antitryptic and antichymotryptic activity. *FEBS Lett* **8**: 274-276.

Nagaiah, K., Bolander, F.F., Jr., Nicholas, K.R., Takemoto, T., and Topper, Y.J. (1981). Prolactin-induced accumulation of casein mRNA in mouse mammary explants: a selective role of glucocorticoid. *Biochem Biophys Res Commun* **98**: 380-387.

Nakai, K., and Kanehisa, M. (1988). Prediction of *in-vivo* modification sites of proteins from their primary structures. *J Biochem (Tokyo)* **104**: 693-699.

Nakamura, T., Kawase, H., Kimura, K., Watanabe, Y., Ohtani, M., Arai, I., and Urashima, T. (2003). Concentrations of sialyloligosaccharides in bovine colostrum and milk during the prepartum and early lactation. *J Dairy Sci* **86**: 1315-1320.

Naqvi, M.A., Irani, K.A., Katanishooshtari, M., and Rousseau, D. (2015). Disorder in Milk Proteins: Formation, Structure, Function, Isolation and Applications of Casein Phosphopeptides. *Curr Protein Pept Sci*.

Nathanielsz, P.W. (1978). Endocrine mechanisms of parturition. *Annual Review of Physiology* **40**: 411-445.

Naylor, M.J., Ginsburg, E., Iismaa, T.P., Vonderhaar, B.K., Wynick, D., and Ormandy, C.J. (2003). The neuropeptide galanin augments lobuloalveolar development. *J Biol Chem* **278**: 29145-29152.

NEB (2006). pMAL™ Protein and Fusion Purification System Instruction Manual, #E8000S. **Version 5.1 (01/06)**.

Nelson, W.J., Hammerton, R.W., Wang, A.Z., and Shore, E.M. (1990). Involvement of the membrane-cytoskeleton in development of epithelial cell polarity. *Semin Cell Biol* **1**: 359-371.

Neville, M.C., Allen, J.C., and Watters, C. (1983). The mechanisms of milk secretion. In: Lactation: Physiology, Nutrition, and Breast-Feeding, M.C. Neville, and M.R. Neifert, eds. (New York: Plenum Publishing Co), pp. 49-92.

Neville, M.C., McFadden, T.B., and Forsyth, I. (2002). Hormonal regulation of mammary differentiation and milk secretion. *J Mammary Gland Biol Neoplasia* **7**: 49-66.

Neville, M.C., Medina, D., Monks, J., and Hovey, R.C. (1998). The mammary fat pad. *Journal of Mammary Gland Biology and Neoplasia* **3**: 109-116.

Neville, M.C., and Morton, J. (2001). Physiology and endocrine changes underlying human lactogenesis II. *J Nutr* **131**: 3005S-3008S.

Newburg, D.S. (1999). Human milk glycoconjugates that inhibit pathogens. *Curr Med Chem* **6**: 117-127.

Newburg, D.S. (2009). Neonatal protection by an innate immune system of human milk consisting of oligosaccharides and glycans. *J Anim Sci* **87**: 26-34.

Newburg, D.S. (2013). Glycobiology of human milk. *Biochemistry Biokhimiia* **78**: 771-785.

Newburg, D.S., and Walker, W.A. (2007). Protection of the neonate by the innate immune system of developing gut and of human milk. *Pediatr Res* **61**: 2-8.

Nguyen, D.A., Parlow, A.F., and Neville, M.C. (2001). Hormonal regulation of tight junction closure in the mouse mammary epithelium during the transition from pregnancy to lactation. *J Endocrinol* **170**: 347-356.

NH&MRC (National Health & Medical Research Council, Australia) (2013) Australian code for the care and use of animals for scientific purposes, 8th edition. Canberra: National Health and Medical Research Council.

Niall, H.D., Hogan, M.L., Sauer, R., Rosenblum, I.Y., and Greenwood, F.C. (1971). Sequences of pituitary and placental lactogenic and growth hormones: evolution from a primordial peptide by gene reduplication. *Proc Natl Acad Sci U S A* **68**: 866-870.

Nicholas, K., Loughnan, M., Messer, M., Munks, S., Griffiths, M., and Shaw, D. (1989). Isolation, partial sequence and asynchronous appearance during lactation of lysozyme and α-lactalbumin in the milk of a marsupial, the common ringtail possum (*Pseudocheirus peregrinus*). *Comp Biochem Physiol B* **94**: 775-778.

Nicholas, K.R. (1988a). Asynchronous dual lactation in a marsupial, the tammar wallaby (*Macropus eugenii*). *Biochem Biophys Res Commun* **154**: 529-536.

Nicholas, K.R. (1988b). Control of milk protein synthesis in the marsupial *Macropus eugenii*: a model system to study prolactin-dependent development. In: The Developing Marsupial Models for Biomedical Research, C.H. Tyndale-Biscoe, and P.A. Janssens, eds. (Berlin, Germany: Springer-Verlag), pp. 68-85.

Nicholas, K.R., Fisher, J.A., Muths, E., Trott, J., Janssens, P.A., Reich, C., and Shaw, D.C. (2001). Secretion of whey acidic protein and cystatin is down regulated at mid-lactation in the red kangaroo (*Macropus rufus*). *Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology* **129**: 851-858.

Nicholas, K.R., Messer, M., Elliott, C., Maher, F., and Shaw, D.C. (1987). A novel whey protein synthesized only in late lactation by the mammary gland from the tammar (*Macropus eugenii*). *Biochem J* **241**: 899-904.

Nicholas, K.R., Sankaran, L., and Topper, Y.J. (1981). The induction of α-lactalbumin in rat mammary explants in the absence of exogenous prolactin: effects of progesterone and estrogen. *Endocrinology* **109**: 978-980.

Nicholas, K.R., Sankaran, L., and Topper, Y.J. (1983). A unique and essential role for insulin in the phenotypic expression of rat mammary epithelial cells unrelated to its function in cell maintenance. *Biochim Biophys Acta* **763**: 309-314.

Nicholas, K.R., Simpson, K.J., Wilson, M., Trott, J., and Shaw, D. (1997). The tammar wallaby: a model to study putative autocrine-induced changes in milk composition. *J Mammary Gland Biol Neoplasia* **2**: 299-310.

Nicholas, K.R., and Topper, Y.J. (1980). Enhancement of α-lactabumin-like activity in mammary explants from pregnant rats in the absence of exogenous prolactin. *Biochem Biophys Res Commun* **94**: 1424-1431.

Nicholas, K.R., and Tyndale-Biscoe, C.H. (1985). Prolactin-dependent accumulation of α-lactalbumin in mammary gland explants from the pregnant tammar wallaby (*Macropus eugenii*). *J Endocrinol* **106**: 337-342.

Nicholas, K.R., Wilde, C.J., Bird, P.H., and Hendry, K.A.K. (1994). Asynchronous expression of milk protein genes during lactation in the tammar wallaby (*Macropus eugenii*). *Trends in Comparative Biochemistry and Physiology* **1**: 959-972.

Nicholas, K.R., Wilde, C.J., Bird, P.H., Hendry, K.A.K., Tregenza, K., and Warner, B. (1995). Asynchronous concurrent secretion of milk proteins in the tammar wallaby (*Macropus eugenii*). In: Intercellular Signalling in the Mammary Gland, C.J. Wilde, M. Peaker, and C.H. Knight, eds. (New York, USA: Plenum Press), pp. 153-170.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**: 1-6.

Niiranen, L., Espelid, S., Karlsen, C.R., Mustonen, M., Paulsen, S.M., Heikinheimo, P., and Willassen, N.P. (2007). Comparative expression study to increase the solubility of cold adapted *Vibrio* proteins in *Escherichia coli*. *Protein Expression and Purification* **52**: 210-218.

Noguchi, H., and Kobayashi, N. (2013). Controlled expansion of mammalian cell populations by reversible immortalization. *J Biotechnol Biomater* **3**.

Nothaft, H., and Szymanski, C.M. (2010). Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **8**: 765-778.

Nothaft, H., and Szymanski, C.M. (2013). Bacterial protein N-Glycosylation: new perspectives and applications. *Journal of Biological Chemistry* **288**: 6912-6920.

Nozach, H., Fruchart-Gaillard, C., Fenaille, F., Beau, F., Ramos, O.H., Douzi, B., Saez, N.J., Moutiez, M., Servent, D., Gondry, M.*, et al.* (2013). High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in *E. coli*. *Microbial cell factories [electronic resource]* **12**: 37.

Nukumi, N., Iwamori, T., Kano, K., Naito, K., and Tojo, H. (2007a). Reduction of tumorigenesis and invasion of human breast cancer cells by whey acidic protein (WAP). *Cancer Lett*.

Nukumi, N., Iwamori, T., Kano, K., Naito, K., and Tojo, H. (2007b). Whey acidic protein (WAP) regulates the proliferation of mammary epithelial cells by preventing serine protease from degrading laminin. *J Cell Physiol* **213**: 793-800.

O'Neal, K.D., and Yu-Lee, L.Y. (1993). The proline-rich motif (PRM): a novel feature of the cytokine/hematopoietin receptor superfamily. *Lymphokine Cytokine Res* **12**: 309-312.

O'Rourke, J., Yuan, R., and DeWille, J. (1997). CCAAT/enhancer-binding protein-delta (C/EBP-δ) is induced in growth-arrested mouse mammary epithelial cells. *J Biol Chem* **272**: 6291-6296.

O'Shea, J.J., Schwartz, D.M., Villarino, A.V., Gadina, M., McInnes, I.B., and Laurence, A. (2015). The JAK-STAT pathway: impact on human disease and therapeutic intervention. *Annu Rev Med* **66**: 311-328.

Oakes, S., Rogers, R., Naylor, M., and Ormandy, C. (2008a). Prolactin Regulation of Mammary Gland Development. *Journal of Mammary Gland Biology and Neoplasia* **13**: 13-28.

Oakes, S.R., Naylor, M.J., Asselin-Labat, M.L., Blazek, K.D., Gardiner-Garden, M., Hilton, H.N., Kazlauskas, M., Pritchard, M.A., Chodosh, L.A., Pfeffer, P.L.*, et al.* (2008b). The Ets transcription factor Elf5 specifies mammary alveolar cell fate. *Genes Dev* **22**: 581-586.

Oakley, R.H., and Cidlowski, J.A. (2011). Cellular Processing of the Glucocorticoid Receptor Gene and Protein: New Mechanisms for Generating Tissue-specific Actions of Glucocorticoids. *Journal of Biological Chemistry* **286**: 3177-3184.

Oftedal, O. (2013). Origin and evolution of the major constituents of milk. In: Advanced Dairy Chemistry Volume 1: Proteins, P.F. Fox, and P.L.H. McSweeney, eds. (New York: Springer), pp. 1-42.

Oftedal, O.T. (1984). Milk composition, milk yield and energy output at peak lactation: A comparative review. In Symposia of the Zoological Society of London **51** Physiological Strategies in Lactation, M. Peaker, R.G. Vernon, and C.H. Knight, eds. (London: Academic Press), pp. 33–85.

Oftedal, O.T. (1997). Lactation in whales and dolphins: evidence of divergence between baleen- and toothed-species. *Journal of Mammary Gland Biology and Neoplasia* **V2**: 205-230.

Oftedal, O.T. (2002a). The mammary gland and its origin during synapsid evolution. *Journal of Mammary Gland Biology and Neoplasia* **7**: 225-252.

Oftedal, O.T. (2002b). The origin of lactation as a water source for parchment-shelled eggs. *Journal of Mammary Gland Biology and Neoplasia* **7**: 253-266.

Oftedal, O.T. (2004). Lactation, marine mammals, species comparisons. *Encyclopedia of Animal Science*: 565-567.

Oftedal, O.T. (2012). The evolution of milk secretion and its ancient origins. *Animal* **6**: 355-368.

Oftedal, O.T., Boness, D.J., and Bowen, W.D. (1988). The composition of Hooded Seal (*Cystophora cristata*) milk - an adaption for postnatal fattening. *Canadian Journal of Zoology - Revue Canadienne De Zoologie* **66**: 318-322.

Oftedal, O.T., Boness, D.J., and Tedman, R.A. (1987). The behavior, physiology, and anatomy of lactation in the pinnipedia. *Curr Mamm* **1**: 175–245.

Oftedal, O.T., and Dhouailly, D. (2013). Evo-devo of the mammary gland. *J Mammary Gland Biol Neoplasia* **18**: 105-120.

Oftedal, O.T., Hintz, H.F., and Schryver, H.F. (1983). Lactation in the horse: milk composition and intake by foals. *J Nutr* **113**: 2096-2106.

Oftedal, O.T., and Iverson, S.J. (1995). Phylogenetic variation in the gross composition of milks. In: Handbook of Milk Composition, R.G. Jensen, ed. (San Diego, CA: Academic Press), pp. 749 - 780.

Old, J.M., and Deane, E.M. (2000). Development of the immune system and immunological protection in marsupial pouch young. *Developmental & Comparative Immunology* **24**: 445-454.

Oliver, C.H., and Watson, C.J. (2013). Making milk: A new link between STAT5 and Akt1. *JAKSTAT* **2**: e23228.

Ormandy, C., Camus, A., Barra, J., Damotte, D., Lucas, B., Buteau, H., Edery, M., Brousse, N., Babinet, C., Binart, N.*, et al.* (1997a). Null mutation of the prolactin receptor gene produces multiple reproductive defects in the mouse. *Genes Dev* **11**: 167-178.

Ormandy, C.J., Binart, N., and Kelly, P.A. (1997b). Mammary gland development in prolactin receptor knockout mice. *J Mammary Gland Biol Neoplasia* **V2**: 355-364.

Osada, S., Yamamoto, H., Nishihara, T., and Imagawa, M. (1996). DNA binding specificity of the CCAAT/enhancer-binding protein transcription factor family. *J Biol Chem* **271**: 3891-3896.

Overton, T.W. (2014). Recombinant protein production in bacterial hosts. *Drug Discov Today* **19**: 590-601.

Oyama, F., Hashino, K., Oyama, R., Kato, I., and Titani, K. (1993). Improved method for expression of Kunitz-type serine proteinase inhibitor domain of β-amyloid protein precursor in *Escherichia coli* and characterization of disulfide bonds of the product. *J Biochem (Tokyo)* **114**: 813-819.

Pacheco, A.R., Barile, D., Underwood, M.A., and Mills, D.A. (2015). The impact of the milk glycobiome on the neonate gut microbiota. *Annu Rev Anim Biosci* **3**: 419-445.

Pai, A.A., Pritchard, J.K., and Gilad, Y. (2015). The Genetic and Mechanistic Basis for Variation in Gene Regulation. *PLoS Genetics* **11**: e1004857.

Palmiter, R.D. (1969). Hormonal induction and regulation of lactose synthetase in mouse mammary gland. *Biochem J* **113**: 409-417.

Parry, G., Beck, J.C., Moss, L., Bartley, J., and Ojakian, G.K. (1990). Determination of apical membrane polarity in mammary epithelial cell cultures: the role of cell-cell, cell-substratum, and membrane-cytoskeleton interactions. *Exp Cell Res* **188**: 302-311.

Parry, L.J., Guymer, F.J., Fletcher, T.P., and Renfree, M.B. (1996). Release of an oxytocic peptide at parturition in the marsupial, *Macropus eugenii*. *J Reprod Fertil* **107**: 191-198.

Patton, S., and Neville, M.C. (1997). Maternal adaptation to lactation. *J Mammary Gland Biol Neoplasia* **2**: 201-203.

Peaker, M. (2002). The mammary gland in mammalian evolution: a brief commentary on some of the concepts. *J Mammary Gland Biol Neoplasia* **7**: 347-353.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J.-C., Ferrier, P., and Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO Journal* **30**: 4198-4210.

Pena, R.N., Sanchez, A., Coll, A., and Folch, J.M. (1999). Isolation, sequencing and relative quantitation by fluorescent-ratio PCR of feline β-lactoglobulin I, II, and III cDNAs. *Mamm Genome* **10**: 560-564.

Perrot-Applanat, M., Gualillo, O., Pezet, A., Vincent, V., Edery, M., and Kelly, P.A. (1997). Dominant negative and cooperative effects of mutant forms of prolactin receptor. *Mol Endocrinol* **11**: 1020-1032.

Pervaiz, S., and Brew, K. (1987). Homology and structure-function correlations between α1-acid glycoprotein and serum retinol-binding protein and its relatives. *Faseb J* **1**: 209-214.

Pesce, M.A., and Strande, C.S. (1973). A new micromethod for determination of protein in cerebrospinal fluid and urine. *Clin Chem* **19**: 1265-1267.

Pharo, E.A. (2014). Expression of the mammary gland-specific tammar wallaby *Early Lactation Protein* gene is maintained *in vitro* in the absence of prolactin. *Mol Cell Endocrinol* **382**: 871-880.

Pharo, E.A., Cane, K.N., McCoey, J., Buckle, A.M., Oosthuizen, W.H., Guinet, C., and Arnould, J.P.Y. (2016). A colostrum trypsin inhibitor gene expressed in the cape fur seal mammary gland during lactation. *Gene* **578**: 7-16.

Pharo, E.A., De Leo, A.A., Renfree, M.B., Thomson, P.C., Lefevre, C.M., and Nicholas, K.R. (2012). The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene. *BMC Evol Biol* **12**: 8.

Phillips, T.A., VanBogelen, R.A., and Neidhardt, F.C. (1984). *lon* gene product of *Escherichia coli* is a heat-shock protein. *J Bacteriol* **159**: 283-287.

Picciano, M.F. (2001). Nutrient composition of human milk. *Pediatr Clin North Am* **48**: 53-67.

Pierre, S., Jolivet, G., Devinoy, E., and Houdebine, L.M. (1994). A combination of distal and proximal regions is required for efficient prolactin regulation of transfected rabbit αs1-casein chloramphenicol acetyltransferase constructs. *Mol Endocrinol* **8**: 1720-1730.

Piletz, J.E., Heinlen, M., and Ganschow, R.E. (1981). Biochemical characterization of a novel whey protein from murine milk. *J Biol Chem* **256**: 11509-11516.

Pineiro, A., Brock, J.H., and Esparza, I. (1978). Isolation and properties of bovine colostral trypsin inhibitor. *Ann Rech Vet* **9**: 281-286.

Pineiro, A., Ortega, F., and Uriel, J. (1975). Trypsin inhibitor from cow colostrum. Isolation, electrophoretic characterization and immunologic properties. *Biochim Biophys Acta* **379**: 201-206.

Piotte, C.P., and Grigor, M.R. (1996). A novel marsupial protein expressed by the mammary gland only during the early lactation and related to the Kunitz proteinase inhibitors. *Archives of Biochemistry and Biophysics* **330**: 59-64.

Piotte, C.P., Hunter, A.K., Marshall, C.J., and Grigor, M.R. (1998). Phylogenetic analysis of three lipocalin-like proteins present in the milk of *Trichosurus vulpecula* (Phalangeridae, Marsupialia). *J Mol Evol* **46**: 361-369.

Piotte, C.P., Marshall, C.J., Hubbard, M.J., Collet, C., and Grigor, M.R. (1997). Lysozyme and α-lactalbumin from the milk of a marsupial, the common brush-tailed possum (*Trichosurus vulpecula*). *Biochim Biophys Acta* **1336**.

Plachot, C., Chaboub, L.S., Adissu, H.A., Wang, L., Urazaev, A., Sturgis, J., Asem, E.K., and Lelievre, S.A. (2009). Factors necessary to produce basoapical polarity in human glandular epithelium formed in conventional and high-throughput three-dimensional culture: example of the breast epithelium. *BMC Biol* **7**: 77.

Plath-Gabler, A., Gabler, C., Sinowatz, F., Berisha, B., and Schams, D. (2001). The expression of the IGF family and GH receptor in the bovine mammary gland. *J Endocrinol* **168**: 39-48.

Plath, A., Einspanier, R., Gabler, C., Peters, F., Sinowatz, F., Gospodarowicz, D., and Schams, D. (1998). Expression and localization of members of the fibroblast growth factor family in the bovine mammary gland. *J Dairy Sci* **81**: 2604-2613.

Pomraning, K.R., Smith, K.M., and Freitag, M. (2009). Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* **47**: 142-150.

Pond, C.M. (1977). Significance of lactation in evolution of mammals. *Evolution* **31**: 177-199.

Poole, W.E., Simms, N.G., Wood, J.T., and Lubulwa, M. (1991). Tables for age determination of the Kangaroo Island wallaby (tammar), *Macropus eugenii*, from body measurements. In: Technical Memorandum No32 CSIRO Division of Wildlife and Ecology, Canberra.

Poorten, T.J., and Kuhn, R.E. (2009). Maternal transfer of antibodies to eggs in *Xenopus laevis*. *Developmental & Comparative Immunology* **33**: 171-175.

Potvin, G., and Zhang, Z.S. (2010). Strategies for high-level recombinant protein expression in transgenic microalgae: A review. *Biotechnology Advances* **28**: 910-918.

Prager, E.M., and Wilson, A.C. (1988). Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. *J Mol Evol* **27**: 326-335.

Prinz, W.A., Aslund, F., Holmgren, A., and Beckwith, J. (1997). The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J Biol Chem* **272**: 15661-15667.

Puissant, C., and Houdebine, L.M. (1991). Cortisol induces rapid accumulation of whey acid protein mRNA but not of α_s1 and β-casein mRNA in rabbit mammary explants. *Cell Biol Int Rep* **15**: 121-129.

Qasba, P.K., and Kumar, S. (1997). Molecular divergence of lysozymes and α-lactalbumin. *Crit Rev Biochem Mol Biol* **32**: 255-306.

Qian, X., and Zhao, F.Q. (2014). Current major advances in the regulation of milk protein gene expression. *Crit Rev Eukaryot Gene Expr* **24**: 357-378.

Ramji, D.P., and Foka, P. (2002). CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem J* **365**: 561-575.

Ranganathan, S., Simpson, K.J., Shaw, D.C., and Nicholas, K.R. (1999). The whey acidic protein family: a new signature motif and three-dimensional structure by comparative modeling. *J Mol Graph Model* **17**: 106-113, 134-106.

Raran-Kurussi, S., and Waugh, D.S. (2012). The Ability to Enhance the Solubility of Its Fusion Partners Is an Intrinsic Property of Maltose-Binding Protein but Their Folding Is Either Spontaneous or Chaperone-Mediated. *PLoS ONE* **7**: e49589.

Ratner, D. (1974). The interaction bacterial and phage proteins with immobilized *Escherichia coli* RNA polymerase. *J Mol Biol* **88**: 373-383.

Raught, B., Khursheed, B., Kazansky, A., and Rosen, J. (1994). YY1 represses β-casein gene expression by preventing the formation of a lactation-associated complex. *Mol Cell Biol* **14**: 1752-1763.

Raught, B., Liao, W.S., and Rosen, J.M. (1995). Developmentally and hormonally regulated CCAAT/enhancer-binding protein isoforms influence β-casein gene expression. *Mol Endocrinol* **9**: 1223-1232.

Rawlings, N.D., Tolle, D.P., and Barrett, A.J. (2004). Evolutionary families of peptidase inhibitors. *Biochemical Journal* **378**: 705-716.

Rawlings, N.D., Waller, M., Barrett, A.J., and Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **42**: D503-509.

Reddy, A., Grimwood, B.G., Plummer, T.H., and Tarentino, A.L. (1998). High-level expression of the endo-β-N-acetylglucosaminidase F2 gene in *E.coli*: one step purification to homogeneity. *Glycobiology* **8**: 633-636.

Reich, C.M. (2009). Functional and evolutionary studies of α-lactalbumin. PhD Thesis, Department of Zoology (Melbourne: University of Melbourne).

Reich, C.M., and Arnould, J.P. (2007). Evolution of Pinnipedia lactation strategies: a potential role for α-lactalbumin? *Biology letters* **3**: 546-549.

Reimers, M., and Carey, V.J. (2006). Bioconductor: an open source framework for bioinformatics and computational biology. *Methods Enzymol* **411**: 119-134.

Reinhardt, T.A., and Lippolis, J.D. (2006). Bovine milk fat globule membrane proteome. *Journal of Dairy Research* **73**: 406-416.

Renfree, M.B. (1973). The composition of fetal fluids of the marsupial *Macropus eugenii*. *Developmental Biology* **33**: 62-79.

Renfree, M.B. (1980). Embryonic diapause in the honey possum *Tarsipes-spencerae*. *Search* **11**: 81-81.

Renfree, M.B. (1982). Implantation and placentation. In: Reproduction in Mammals, C.R. Austin, and R.V. Short, eds. (Cambridge: Cambridge University Press), pp. 26–69.

Renfree, M.B. (1983). Marsupial reproduction the choice between placentation and lactation. *Oxford Reviews of Reproductive Biology* **5**: 1-29.

Renfree, M.B. (1994). Endocrinology of pregnancy, parturition and lactation in marsupials. In: Marshall's Physiology of Reproduction Volume 3 Pregnancy and Lactation, G.E. Lamming, ed. (London, UK: Chapman & Hall), pp. 677-766.

Renfree, M.B., Fletcher, T.P., Blanden, D.R., Lewis, D.R., Shaw, G., Gordon, K., Short, R.V., Parer-Cook, E., and Parer, D. (1989). Physiological and behavioural events around the time of birth in macropodid marsupials. In: Kangaroos, Wallabies and Rat-Kangaroos,

G. Grigg, P.J. Jarman, and I.D. Hume, eds. (Sydney: Surrey Beatty and Sons Pty. Ltd.), pp. 323-337.

Renfree, M.B., Holt, A.B., Green, S.W., Carr, J.P., and Cheek, D.B. (1982). Ontogeny of the brain in a marsupial (*Macropus eugenii*) throughout pouch life. I. Brain growth. *Brain Behav Evol* **20**: 57-71.

Renfree, M.B., Papenfuss, A.T., Deakin, J.E., Lindsay, J., Heider, T., Belov, K., Rens, W., Waters, P.D., Pharo, E.A., Shaw, G.*, et al.* (2011). Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol* **12**: R81.

Renfree, M.B., Pask, A.J., and Shaw, G. (2001). Sex down under: the differentiation of sexual dimorphisms during marsupial development. *Reprod Fertil Dev* **13**: 679-690.

Renfree, M.B., Shaw, G., and Fletcher, T.P. (1994). Evidence for the essential role of prostaglandins for parturition in a marsupial, *Macropus eugenii*. *J Reprod Fertil* **102**: 433-446.

Renfree, M.B., and Tyndale-Biscoe, C.H. (1978). Manipulation of marsupial embryos and pouch young. In: Methods in Mammalian Reproduction, J.C. Daniel, ed. (New York: Academic Press), pp. 307-331.

Resau, J.H., Sakamoto, K., Cottrell, J.R., Hudson, E.A., and Meltzer, S.J. (1991). Explant organ culture: a review. *Cytotechnology* **7**: 137-149.

Reynolds, M.L., Cavanagh, M.E., Dziegielewska, K.M., Hinds, L.A., Saunders, N.R., and Tyndale-Biscoe, C.H. (1985). Postnatal development of the telencephalon of the tammar wallaby (*Macropus eugenii*). An accessible model of neocortical differentiation. *Anat Embryol (Berl)* **173**: 81-94.

Reynolds, M.L., and Saunders, N.R. (1988). Development of the neocortex. In: The Developing Marsupial Models for Biomedical Research, C.H. Tyndale-Bisco, and P.A. Janssens, eds. (Berlin, Germany: Springer-Verlag), pp. 101-116.

Richardson, S.J., Aldred, A.R., Leng, S.L., Renfree, M.B., Hulbert, A.J., and Schreiber, G. (2002). Developmental profile of thyroid hormone distributor proteins in a marsupial, the tammar wallaby Macropus eugenii. *General and Comparative Endocrinology* **125**: 92-103.

Richert, M.M., Schwertfeger, K.L., Ryder, J.W., and Anderson, S.M. (2000). An atlas of mouse mammary gland development. *J Mammary Gland Biol Neoplasia* **5**: 227-241.

Rijnkels, M. (2002). Multispecies comparison of the casein gene loci and evolution of casein gene family. *J Mammary Gland Biol Neoplasia* **7**: 327-345.

Rijnkels, M., Freeman-Zadrowski, C., Hernandez, J., Potluri, V., Wang, L., Li, W., and Lemay, D.G. (2013). Epigenetic modifications unlock the milk protein gene loci during mouse mammary gland development and differentiation. *PLoS One* **8**: e53270.

Rijnkels, M., Kabotyanski, E., Montazer-Torbati, M.B., Beauvais, C.H., Vassetzky, Y., Rosen, J.M., and Devinoy, E. (2010). The epigenetic landscape of mammary gland development and functional differentiation. *J Mammary Gland Biol Neoplasia* **15**: 85-100.

Riley, L.G., Gardiner-Garden, M., Thomson, P.C., Wynn, P.C., Williamson, P., Raadsma, H.W., and Sheehy, P.A. (2010). The influence of extracellular matrix and prolactin on global gene expression profiles of primary bovine mammary epithelial cells *in vitro*. *Anim Genet* **41**: 55-63.

Rival-Gervier, S., Thepot, D., Jolivet, G., and Houdebine, L.M. (2003). Pig whey acidic protein gene is surrounded by two ubiquitously expressed genes. *Biochim Biophys Acta* **1627**: 7-14.

Robinson, G.W., Johnson, P.F., Hennighausen, L., and Sterneck, E. (1998). The C/EBPβ transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland. *Genes Dev* **12**: 1907-1916.

Robinson, R. (2008). For mammals, loss of yolk and gain of milk went hand in hand. *PLoS Biol* **6**: e77.

Rodriguez-Boulan, E., and Salas, P.J. (1989). External and internal signals for epithelial cell surface polarization. *Annu Rev Physiol* **51**: 741-754.

Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* **5**.

Rosen, J.M., Matusik, R.J., Richards, D.A., Gupta, P., and Rodgers, J.R. (1980). Multihormonal regulation of casein gene expression at the transcriptional and posttransciptional levels in the mammary gland. *Recent Prog Horm Res* **36**: 157-193.

Rosen, J.M., Poyet, P., Goodman, H., and Lee, K.F. (1989). Mechanisms by which prolactin and glucocorticoids regulate casein gene expression. *Biochem Soc Symp* **55**: 115-123.

Rosen, J.M., Wyszomierski, S.L., and Hadsell, D. (1999). Regulation of milk protein gene expression. *Annual Review of Nutrition* **19**: 407-436.

Rosen, J.M., Zahnow, C., Kazansky, A., and Raught, B. (1998). Composite response elements mediate hormonal and developmental regulation of milk protein gene expression. *Biochem Soc Symp* **63**: 101-113.

Roth, J., Zuber, C., Park, S., Jang, I., Lee, Y., Kysela, K.G., Le Fourn, V., Santimaria, R., Guhl, B., and Cho, J.W. (2010). Protein N-glycosylation, protein folding, and protein quality control. *Mol Cells* **30**: 497-506.

Roy, L., Cates, S., Schifferli, K., Pichet, J., Ciccarone, V., Bennett, S., and Hawley-Newson, P. (1999). High transfection efficiency of cloned cell lines. In Focus, pp. 62-62.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**.

Rubinstein, L.V., Shoemaker, R.H., Paull, K.D., Simon, R.M., Tosini, S., Skehan, P., Scudiero, D.A., Monks, A., and Boyd, M.R. (1990). Comparison of *in vitro* anticancer-drug-screening data generated with a tetrazolium assay versus a protein assay against a diverse panel of human tumor cell lines. *J Natl Cancer Inst* **82**: 1113-1118.

Rudolph, M.C., McManaman, J.L., Hunter, L., Phang, T., and Neville, M.C. (2003). Functional development of the mammary gland: use of expression profiling and trajectory

clustering to reveal changes in gene expression during pregnancy, lactation, and involution. *J Mammary Gland Biol Neoplasia* **8**: 287-307.

Rudolph, M.C., Neville, M.C., and Anderson, S.M. (2007). Lipid synthesis in lactation: diet and the fatty acid switch. *J Mammary Gland Biol Neoplasia* **12**: 269-281.

Rui, H., Djeu, J.Y., Evans, G.A., Kelly, P.A., and Farrar, W.L. (1992). Prolactin receptor triggering. Evidence for rapid tyrosine kinase activation. *J Biol Chem* **267**: 24076-24081.

Rui, H., Kirken, R., and Farrar, W. (1994). Activation of receptor-associated tyrosine kinase JAK2 by prolactin. *J Biol Chem* **269**: 5364-5368.

Ruta, M., Botha-Brink, J., Mitchell, S.A., and Benton, M.J. (2013). The radiation of cynodonts and the ground plan of mammalian morphological diversity. *Proceedings of the Royal Society of London B: Biological Sciences* **280**.

Sachdev, D., and Chirgwin, J.M. (1998). Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli*. *Biochem Biophys Res Commun* **244**: 933-937.

Sachdev, D., and Chirgwin, J.M. (2000). Fusions to maltose-binding protein: Control of folding and solubility in protein purification. In: Methods in Enzymology, pp. 312-321.

Salier, J.P., Rouet, P., Raguenez, G., and Daveau, M. (1996). The inter-alpha-inhibitor family: from structure to regulation. *Biochem J* **315 (Pt 1)**: 1-9.

Salmon, H., Berri, M., Gerdts, V., and Meurens, F. (2009). Humoral and cellular factors of maternal immunity in swine. *Dev Comp Immunol* **33**: 384-393.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual, Third edn (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* **8**: 424-436.

Sangawa, T., Tabata, S., Suzuki, K., Saheki, Y., Tanaka, K., and Takagi, J. (2013). A multipurpose fusion tag derived from an unstructured and hyperacidic region of the amyloid precursor protein. *Protein Sci* **22**: 840-850.

Santoro, C., Mermod, N., Andrews, P.C., and Tjian, R. (1988). A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. *Nature* **334**: 218-224.

Saunier, E., Dif, F., Kelly, P.A., and Edery, M. (2003). Targeted expression of the dominant-negative prolactin receptor in the mammary gland of transgenic mice results in impaired lactation. *Endocrinology* **144**: 2669-2675.

Sawyer, L. (2013). β-Lactoglobulin. In: Advanced Dairy Chemistry Volume 1: Proteins, P.L.H. McSweeney, and P.F. Fox, eds. (Springer), pp. 211-259.

Sawyer, L., and Holt, C. (1993). The secondary structure of milk proteins and their biological function. *J Dairy Sci* **76**: 3062-3078.

Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S.*, et al.* (2009). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **37**: D5-15.

Schindler, C., Levy, D.E., and Decker, T. (2007). JAK-STAT signaling: from interferons to cytokines. *J Biol Chem* **282**: 20059-20063.

Schmidhauser, C., Casperson, G.F., Myers, C.A., Sanzo, K.T., Bolten, S., and Bissell, M.J. (1992). A novel transcriptional enhancer is involved in the prolactin- and extracellular matrix-dependent regulation of β-casein gene expression. *Mol Biol Cell* **3**: 699-709.

Schmitt-Ney, M., Doppler, W., Ball, R.K., and Groner, B. (1991). β-Casein gene promoter activity is regulated by the hormone-mediated relief of transcriptional repression and a mammary-gland-specific nuclear factor. *Mol Cell Biol* **11**: 3745-3755.

Schmitt-Ney, M., Happ, B., Ball, R.K., and Groner, B. (1992). Developmental and environmental regulation of a mammary gland-specific nuclear factor essential for transcription of the gene encoding b-casein. *Proc Natl Acad Sci U S A* **89**: 3130-3134.

Schneider, N.Y., Fletcher, T.P., Shaw, G., and Renfree, M.B. (2009). The olfactory system of the tammar wallaby is developed at birth and directs the neonate to its mother's pouch odours. *Reproduction* **138**: 849-857.

Schoneveld, O.J., Gaemers, I.C., and Lamers, W.H. (2004). Mechanisms of glucocorticoid signalling. *Biochim Biophys Acta* **1680**: 114-128.

Schulz, T.M., and Bowen, W.D. (2005). The evolution of lactation strategies in pinnipeds: A phylogenetic analysis. *Ecological Monographs* **75**: 159-177.

Schumacher, I.M., Rostal, D.C., Yates, R.A., Brown, D.R., Jacobson, E.R., and Klein, P.A. (1999). Persistence of maternal antibodies against *Mycoplasma agassizii* in desert tortoise hatchlings. *American Journal of Veterinary Research* **60**: 826-831.

Schwalm, J.W., and Tucker, A. (1978). Glucocorticoids in mammary secretions and blood serum during reproduction and lactation and distributions of glucocorticoids, progesterone, and estrogens in fractions of milk. *Journal of Dairy Science* **61**: 550-560.

Schweisguth, F. (2004). Formation and remodeling of epithelial polarity. *Dev Cell* **6**: 749-755.

Scott, C.J., and Taggart, C.C. (2010). Biologic protease inhibitors as novel therapeutic agents. *Biochimie*. **92**: 1681-1688.

Seagroves, T.N., Krnacik, S., Raught, B., Gay, J., Burgess-Beusse, B., Darlington, G.J., and Rosen, J.M. (1998). C/EBPβ, but not C/EBPα, is essential for ductal morphogenesis, lobuloalveolar proliferation, and functional differentiation in the mouse mammary gland. *Genes Dev* **12**: 1917-1928.

Sebastian, L.T., De Matteo, L., Shaw, G., and Renfree, M.B. (1998). Mesotocin receptors during pregnancy, parturition and lactation in the tammar wallaby. *Anim Reprod Sci* **51**: 57-74.

Sehl, L.C., Nguyen, H.V., Berleau, L.T., Arcila, P., Bennett, W.F., and Keyt, B.A. (1996). Locating the unpaired cysteine of tissue-type plasminogen activator. *Protein Eng* **9**: 283-290.

Sejrsen, K., Purup, S., Vestergaard, M., Weber, M.S., and Knight, C.H. (1999). Growth hormone and mammary development. *Domest Anim Endocrinol* **17**: 117-129.

Selwood, L., and Woolley, P.A. (1991). A timetable of embryonic development, and ovarian and uterine changes during pregnancy, in the stripe-faced dunnart, *Sminthopsis macroura* (Marsupialia: Dasyuridae). *J Reprod Fertil* **91**: 213-227.

Sementchenko, V.I., and Watson, D.K. (2000). Ets target genes: past, present and future. *Oncogene* **19**: 6533-6548.

Sernia, C., and Tyndale-Biscoe, C.H. (1979). Prolactin receptors in the mammary gland, corpus luteum and other tissues of the tammar wallaby, *Macropus eugenii*. *J Endocrinol* **83**: 79-89.

Setchell, P.J. (1974). The development of thermoregulation and thyroid function in the marsupial *Macropus eugenii* (Desmarest). *Comparative Biochemistry and Physiology Part A: Physiology* **47**: 1115-1116.

Sethuraman, N., and Stadheim, T.A. (2006). Challenges in therapeutic glycoprotein production. *Current opinion in biotechnology* **17**: 341-346.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. *Nature* **439**: 84-88.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annual Review of Biochemistry* **76**: 75-100.

Shamay, A., Zeelon, E., Ghez, Z., Cohen, N., Mackinlay, A.G., and Gertler, A. (1987). Inhibition of casein and fat synthesis and α-lactalbumin secretion by progesterone in explants from bovine lactating mammary glands. *J Endocrinol* **113**: 81-88.

Shannon, J.M., and Pitelka, D.R. (1981). The influence of cell shape on the induction of functional differentiation in mouse mammary cells *in vitro*. *In Vitro* **17**: 1016-1028.

Sharman, G.B. (1962). Initiation and maintenance of lactation in marsupial, *Trichosurus vulpecula*. *Journal of Endocrinology* **25**: 375-385.

Sharman, G.B., and Calaby, J.H. (1964). Reproductive behaviour in the red Kangaroo, Megaleia rufa, in captivity. *CSIRO Wildlife Research* **9**: 58-85.

Sharman, G.B., and Pilton, P.E. (1964). The life history and reproduction of the red kangaroo, *Megaleia rufa*. *Proc Zool Soc Lond* **142**: 29-48.

Sharp, J.A., Cane, K., Arnould, J.P., and Nicholas, K.R. (2005). The lactation cycle of the fur seal. *J Dairy Res* **72**: 81-89.

Sharp, J.A., Cane, K.N., Mailer, S.L., Oosthuizen, W.H., Arnould, J.P., and Nicholas, K.R. (2006). Species-specific cell-matrix interactions are essential for differentiation of

alveoli like structures and milk gene expression in primary mammary cells of the Cape fur seal (*Arctocephalus pusillus pusillus*). *Matrix Biol* **25**: 430-442.

Sharp, J.A., Digby, M., Lefevre, C., Mailer, S., Khalil, E., Topcic, D., Auguste, A., Kwek, J.H., Brennan, A.J., Familari, M.*, et al.* (2009). The comparative genomics of tammar wallaby and Cape fur seal lacation models to examine functions of milk proteins. In: Milk Proteins: from Expression to Food, A. Thompson, M. Boland, and H. Singh, eds. (San Diego, CA: Academic Press/Elsevier), pp. 55-79.

Sharp, J.A., Lefevre, C., and Nicholas, K.R. (2007). Molecular evolution of monotreme and marsupial whey acidic protein genes. *Evolution & Development* **9**: 378-392.

Sharp, J.A., Lefevre, C., and Nicholas, K.R. (2008). Lack of functional α-lactalbumin prevents involution in Cape fur seals and identifies the protein as an apoptotic milk factor in mammary gland involution. *BMC Biol* **6**: 48.

Sharp, J.A., Modepalli, V., Enjapoori, A.K., Bisana, S., Abud, H.E., Lefevre, C., and Nicholas, K.R. (2014). Bioactive Functions of Milk Proteins: a Comparative Genomics Approach. *J Mammary Gland Biol Neoplasia* **19**: 289-302.

Sharp, J.A., Watt, A., Bisana, S., Modepalli, V., Wanyonyi, S., Kumar, A., Kwek, J., Collins, R., Lefèvre, C., and Nicholas, K.R. (2014). The comparative genomics of monotremes, marsupials, and pinnipeds: models to examine functions of milk proteins. In Milk Proteins : From Expression to Food (Second Edition), H. Singh, M. Boland, and A. Thompson, eds. (San Diego: Academic Press), pp. 75-112.

Sharrocks, A.D. (2001). The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol* **2**: 827-837.

Shaw, D.C., Messer, M., Scrivener, A.M., Nicholas, K.R., and Griffiths, M. (1993). Isolation, partial characterisation, and amino acid sequence of α-lactalbumin from platypus (*Ornithorhynchus anatinus*) milk. *Biochimica Et Biophysica Acta* **1161**: 177-186.

Shaw, G., and Renfree, M.B. (2006). Parturition and perfect prematurity: birth in marsupials. *Australian Journal of Zoology* **54**: 139-149.

Shaw, G., Renfree, M.B., and Fletcher, T.P. (1996). A role for glucocorticoids in parturition in a marsupial, *Macropus eugenii*. *Biol Reprod* **54**: 728-733.

Shay, J.W., and Wright, W.E. (2000). Hayflick, his limit, and cellular ageing. *Nat Rev Mol Cell Biol* **1**: 72-76.

Sheehy, P.A., Della-Vedova, J.J., Nicholas, K.R., and Wynn, P.C. (2004). Hormone-dependent milk protein gene expression in bovine mammary explants from biopsies at different stages of pregnancy. *Journal of Dairy Research* **71**: 135-140.

Sheehy, P.A., Nicholas, K.R., and Wynn, P.C. (2000). An investigation of the role of insulin in bovine milk protein gene expression in mammary explant culture. *Asian-Australasian Journal of Animal Sciences* **13**: 272-275.

Shennan, D.B., and Peaker, M. (2000). Transport of milk constituents by the mammary gland. *Physiological Reviews* **80**: 925-951.

Shepherd, T., and Hassell, J.A. (2001). Role of Ets transcription factors in mammary gland development and oncogenesis. *J Mammary Gland Biol Neoplasia* **6**: 129-140.

Shewale, J.G., Sinha, S.K., and Brew, K. (1984). Evolution of α-lactalbumins. The complete amino acid sequence of the α-lactalbumin from a marsupial (*Macropus rufogriseus*) and corrections to regions of sequence in bovine and goat α-lactalbumins. *J Biol Chem* **259**: 4947-4956.

Shia, S., Stamos, J., Kirchhofer, D., Fan, B., Wu, J., Corpuz, R.T., Santell, L., Lazarus, R.A., and Eigenbrot, C. (2005). Conformational lability in serine protease active sites: structures of hepatocyte growth factor activator (HGFA) alone and with the inhibitory domain from HGFA inhibitor-1B. *J Mol Biol* **346**: 1335-1349.

Shimomura, T., Denda, K., Kawaguchi, T., Matsumoto, K., Miyazawa, K., and Kitamura, N. (1999). Multiple sites of proteolytic cleavage to release soluble forms of hepatocyte growth factor activator inhibitor type 1 from a transmembrane form. *J Biochem (Tokyo)* **126**: 821-828.

Shiu, R.P.C., and Friesen, H.G. (1980). Mechanism of action of prolactin in the control of mammary gland function. *Annual Review of Physiology* **42**: 83-96.

Shuai, K., Ziemiecki, A., Wilks, A.F., Harpur, A.G., Sadowski, H.B., Gilman, M.Z., and Darnell, J.E. (1993). Polypeptide signalling to the nucleus through tyrosine phosphorylation of Jak and Stat proteins. *Nature* **366**: 580-583.

Siddle, H.V., Sanderson, C.E., Deakin, J.E., and Belov, K. (2010). Genetic architecture of the macropodid immune system. In: Macropods: The Biology of Kangaroos, Wallabies and Rat-kangaroos, G. Coulson, and M.D.B. Eldridge, eds. (Collingwood: CSIRO Publishing), pp. 13-24.

Simpson, K. (1998). The tammar wallaby, *Macropus eugenii*, a model to study autocrine and endocrine regulation of lactation. PhD Thesis, Department of Agriculture (Melbourne: Latrobe University).

Simpson, K., Shaw, D., and Nicholas, K. (1998a). Developmentally-regulated expression of a putative protease inhibitor gene in the lactating mammary gland of the tammar wallaby, *Macropus eugenii*. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **120**: 535-541.

Simpson, K.J., Bird, P., Shaw, D., and Nicholas, K. (1998b). Molecular characterisation and hormone-dependent expression of the porcine whey acidic protein gene. *J Mol Endocrinol* **20**: 27-35.

Simpson, K.J., and Nicholas, K.R. (2002). The comparative biology of whey proteins. *J Mammary Gland Biol Neoplasia* **V7**: 313-326.

Simpson, K.J., Ranganathan, S., Fisher, J.A., Janssens, P.A., Shaw, D.C., and Nicholas, K.R. (2000). The gene for a novel member of the whey acidic protein family encodes three four-disulfide core domains and is asynchronously expressed during lactation. *J Biol Chem* **275**: 23074-23081.

Singh, K., Erdman, R.A., Swanson, K.M., Molenaar, A.J., Maqbool, N.J., Wheeler, T.T., Arias, J.A., Quinn-Walsh, E.C., and Stelwagen, K. (2010). Epigenetic regulation of milk production in dairy cows. *J Mammary Gland Biol Neoplasia* **15**: 101-112.

Skehan, P., Storeng, R., Scudiero, D., Monks, A., McMahon, J., Vistica, D., Warren, J.T., Bokesch, H., Kenney, S., and Boyd, M.R. (1990). New colorimetric cytotoxicity assay for anticancer-drug screening. *J Natl Cancer Inst* **82**: 1107-1112.

Smilowitz, J.T., Lebrilla, C.B., Mills, D.A., German, J.B., and Freeman, S.L. (2014). Breast milk oligosaccharides: structure-function relationships in the neonate. *Annu Rev Nutr* **34**: 143-169.

Smith-Kirwin, S.M., O'Connor, D.M., De Johnston, J., Lancey, E.D., Hassink, S.G., and Funanage, V.L. (1998). Leptin expression in human mammary epithelial cells and breast milk. *J Clin Endocrinol Metab* **83**: 1810-1813.

Smith, L.A., Olson, M.A., Lafaye, P.J., and Dolly, J.O. (1995). Cloning and expression of mamba toxins. *Toxicon* **33**: 459-474.

Smith, R.F., Wiese, B.A., Wojzynski, M.K., Davison, D.B., and Worley, K.C. (1996). BCM Search Launcher - an integrated interface to molecular biology data base search and analysis services available on the World Wide Web. *Genome Research* **6**: 454-462.

Smith, T., and Little, R.B. (1922). The significance of colostrum to the new-born calf. *J Exp Med* **36**: 181-198.

Smyth, E., Clegg, R.A., and Holt, C. (2004). A biological perspective on the structure and function of caseins and casein micelles. *International Journal of Dairy Technology* **57**: 121-126.

Song, G., Ouyang, G., and Bao, S. (2005). The activation of Akt/PKB signaling pathway and cell survival. *Journal of Cellular and Molecular Medicine* **9**: 59-71.

Song, S., and Oka, T. (2003). Regulation of type II deiodinase expression by EGF and glucocorticoid in HC11 mouse mammary epithelium. *Am J Physiol Endocrinol Metab* **284**: E1119-1124.

Sordillo, L.M., and Streicher, K.L. (2002). Mammary gland immunity and mastitis susceptibility. *J Mammary Gland Biol Neoplasia* **7**: 135-146.

Sorensen, H.P., and Mortensen, K.K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial cell factories [electronic resource]* **4**: 1.

Sorensen, M.A., Kurland, C.G., and Pedersen, S. (1989). Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol* **207**: 365-377.

Soule, H.D., Vazguez, J., Long, A., Albert, S., and Brennan, M. (1973). A human cell line from a pleural effusion derived from a breast carcinoma. *J Natl Cancer Inst* **51**: 1409-1416.

Spanjaard, R.A., and van Duin, J. (1988). Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. *Proc Natl Acad Sci U S A* **85**: 7967-7971.

Spivakov, M. (2014). Spurious transcription factor binding: Non-functional or genetically redundant? *Bioessays* **36**: 798-806.

Stampfer, M.R., and Bartley, J.C. (1988). Human mammary epithelial cells in culture: differentiation and transformation. *Cancer Treat Res* **40**: 1-24.

Stampfer, M.R., and Yaswen, P. (1993). Culture systems for study of human mammary epithelial cell proliferation, differentiation and transformation. *Cancer Surv* **18**: 7-34.

Stein, T., Salomonis, N., and Gusterson, B.A. (2007). Mammary gland involution as a multi-step process. *J Mammary Gland Biol Neoplasia* **12**: 25-35.

Steinmetz, R.W., Grant, A.L., and Malven, P.V. (1993). Transcription of prolactin gene in milk secretory cells of the rat mammary gland. *J Endocrinol* **136**: 271-276.

Stelwagen, K., Carpenter, E., Haigh, B., Hodgkinson, A., and Wheeler, T.T. (2009). Immune components of bovine colostrum and milk. *J Anim Sci* **87**: 3-9.

Sternlicht, M.D. (2006). Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. *Breast Cancer Res* **8**: 201.

Stewart, E.J., Aslund, F., and Beckwith, J. (1998). Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO J* **17**: 5543-5550.

Stewart, F. (1984). Mammogenesis and changing prolactin receptor concentrations in the mammary glands of the tammar wallaby (*Macropus eugenii*). *J Reprod Fertil* **71**: 141-148.

Stocklin, E., Wissler, M., Gouilleux, F., and Groner, B. (1996). Functional interactions between Stat5 and the glucocorticoid receptor. *Nature* **383**: 726-728.

Strange, R., Li, F., Friis, R.R., Reichmann, E., Haenni, B., and Burri, P.H. (1991). Mammary epithelial differentiation *in vitro*: minimum requirements for a functional response to hormonal stimulation. *Cell Growth Differ* **2**: 549-559.

Streuli, C.H. (1993). Extracellular matrix and gene expression in mammary epithelium. *Semin Cell Biol* **4**: 203-212.

Streuli, C.H., Bailey, N., and Bissell, M.J. (1991). Control of mammary epithelial differentiation: basement membrane induces tissue-specific gene expression in the absence of cell-cell interaction and morphological polarity. *J Cell Biol* **115**: 1383-1395.

Streuli, C.H., and Bissell, M.J. (1991). Mammary epithelial cells, extracellular matrix, and gene expression. *Cancer Treat Res* **53**: 365-381.

Streuli, C.H., Schmidhauser, C., Bailey, N., Yurchenco, P., Skubitz, A.P., Roskelley, C., and Bissell, M.J. (1995). Laminin mediates tissue-specific gene expression in mammary epithelia. *J Cell Biol* **129**: 591-603.

Stuart, D.I., Acharya, K.R., Walker, N.P., Smith, S.G., Lewis, M., and Phillips, D.C. (1986). α-lactalbumin possesses a novel calcium binding loop. *Nature* **324**: 84-87.

Sun, Y.L., Lin, C.S., and Chou, Y.C. (2006). Establishment and characterization of a spontaneously immortalized porcine mammary epithelial cell line. *Cell Biol Int* **30**: 970-976.

Swain, P., and Nayak, S.K. (2009). Role of maternally derived immunity in fish. *Fish & Shellfish Immunology* **27**: 89-99.

Swaisgood, H.E. (1993). Review and update of casein chemistry. *J Dairy Sci* **76**: 3054-3061.

Swietnicki, W. (2006). Folding aggregated proteins into functionally active forms. *Current opinion in biotechnology* **17**: 367-372.

Szdzuy, K., Zeller, U., Renfree, M., Tzschentke, B., and Janke, O. (2008). Postnatal lung and metabolic development in two marsupial and four eutherian species. *J Anat* **212**: 164-179.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular biology and evolution* **24**: 1596-1599.

Tang, S., Han, H., and Bajic, V.B. (2004). ERGDB: Estrogen Responsive Genes Database. *Nucleic Acids Research* **32**: D533-D536.

Teahan, C.G., McKenzie, H.A., Shaw, D.C., and Griffiths, M. (1991). The isolation and amino acid sequences of echidna (*Tachyglossus aculeatus*) milk lysozyme I and II. *Biochem Int* **24**: 85-95.

Teglund, S., McKay, C., Schuetz, E., van Deursen, J.M., Stravopodis, D., Wang, D., Brown, M., Bodner, S., Grosveld, G., and Ihle, J.N. (1998). Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. *Cell* **93**: 841-850.

Telemo, E., and Hanson, L.A. (1996). Antibodies in milk. *Journal of Mammary Gland Biology and Neoplasia* **1**: 243-249.

Thompson, M.D., Dave, J.R., and Nakhasi, H.L. (1985). Molecular cloning of mouse mammary gland kappa-casein: comparison with rat kappa-casein and rat and human gamma-fibrinogen. *DNA* **4**: 263-271.

Thomson, P.C. (2006). Analysis of microarray data: A mixed model finite-mixture approach. Paper presented at: Proceedings of the XXIIIrd International Biometric Conference (Montreal, Canada.).

Topcic, D. (2010). Identifying the function of milk protein genes and factors regulating their expression by exploiting the lactation strategy of the Tammar Wallaby (*Macropus eugenii*). PhD Thesis, Department of Zoology (Melbourne: University of Melbourne).

Topcic, D., Auguste, A., De Leo, A.A., Lefevre, C., Digby, M.R., and Nicholas, K.R. (2009). Characterization of the tammar wallaby (*Macropus eugenii)* whey acidic protein gene: new insights into the function of the protein. *Evol Dev* **11**: 363-375.

Topper, Y.J., and Freeman, C.S. (1980). Multiple hormone interactions in the developmental biology of the mammary gland. *Physiol Rev* **60**: 1049-1106.

Travis, J. (2009). Origins. On the origin of the immune system. *Science* **324**: 580-582.

Trejo, R., Dokland, T., Jurat-Fuentes, J., and Harte, F. (2011). Cryo-transmission electron tomography of native casein micelles from bovine milk. *J Dairy Sci* **94**: 5770-5775.

Trott, J.F. (1999). Endocrine and local control of milk protein gene expression in the tammar wallaby, *Macropus eugenii*. PhD Thesis, Department of Molecular Biology, Genetics & Zoology (Melbourne: University of Melbourne).

Trott, J.F., Adams, T.E., Wilson, M., and Nicholas, K.R. (2005). Positive and negative regulatory elements in the late lactation protein-A gene promoter from the tammar wallaby (*Macropus eugenii*). *Biochimica Et Biophysica Acta-Gene Structure and Expression* **1728**: 65-76.

Trott, J.F., Hovey, R.C., Koduri, S., and Vonderhaar, B.K. (2004). Multiple new isoforms of the human prolactin receptor gene. *Adv Exp Med Biol* **554**: 495-499.

Trott, J.F., Simpson, K.J., Moyle, R.L., Hearn, C.M., Shaw, G., Nicholas, K.R., and Renfree, M.B. (2003). Maternal regulation of milk composition, milk production, and pouch young development during lactation in the tammar wallaby (*Macropus eugenii*). *Biol Reprod* **68**: 929-936.

Trott, J.F., Wilson, M.J., Hovey, R.C., Shaw, D.C., and Nicholas, K.R. (2002). Expression of novel lipocalin-like milk protein gene is developmentally-regulated during lactation in the tammar wallaby, *Macropus eugenii*. *Gene* **283**: 287-297.

Tschesche, H., Klauser, R., Cechova, D., and Jonakova, V. (1975). On the carbohydrate composition of bovine colostrum trypsin inhibitor. *Hoppe Seylers Z Physiol Chem* **356**: 1759-1764.

Tucker, A.H. (1985). Endocrine and neural control of the mammary gland. In: Lactation, B.L. Larson, ed. (Ames, Iowa, USA: The Iowa State University Press), pp. 39-79.

Turk, V., Stoka, V., and Turk, D. (2008). Cystatins: biochemical and structural properties, and medical relevance. *Frontiers in bioscience : a journal and virtual library* **13**: 5406-5420.

Turvey, S.E., and Broide, D.H. (2010). Innate immunity. *J Allergy Clin Immunol* **125**: S24-32.

Tyndale-Biscoe, C.H. (1973). Life of Marsupials (London, UK: Edward Arnold).

Tyndale-Biscoe, C.H. (2005). Life of Marsupials (Collingwood, VIC, Australia: CSIRO Publishing).

Tyndale-Biscoe, C.H., Hearn, J.P., and Renfree, M.B. (1974). Control of reproduction in macropodid marsupials. *J Endocrinol* **63**.

Tyndale-Biscoe, C.H., and Hinds, L.A. (1984). Seasonal patterns of circulating progesterone and prolactin and response to bromocriptine in the female tammar *Macropus eugenii*. *Gen Comp Endocrinol* **53**.

Tyndale-Biscoe, C.H., and Janssens, P.A., eds. (1988a). The Developing Marsupial. Models for Biomedical Research (Berlin, Germany: Springer-Verlag).

Tyndale-Biscoe, C.H., and Janssens, P.A. (1988b). Introduction. In: The Developing Marsupial Models for Biomedical Research, C.H. Tyndale-Bisco, and P.A. Janssens, eds. (Berlin, Germany: Springer-Verlag), pp. 1-7.

Tyndale-Biscoe, C.H., and Renfree, M.B. (1987). Reproductive Physiology of Marsupials (Cambridge, UK: Cambridge University Press).

Tyndale-Biscoe, C.H., Stewart, F., and Hinds, L.A. (1984). Some factors in the initiation and control of lactation in the tammar wallaby. *Symposia of the Zoological Society of London 51 Physiological Strategies in Lactation*: 389-401.

Urashima, T., Arita, M., Yoshida, M., Nakamura, T., Arai, I., Saito, T., Arnould, J.P., Kovacs, K.M., and Lydersen, C. (2001). Chemical characterisation of the oligosaccharides in hooded seal (*Cystophora cristata*) and Australian fur seal (*Arctocephalus pusillus doriferus*) milk. *Comp Biochem Physiol B Biochem Mol Biol* **128**: 307-323.

Urashima, T., Saito, T., Tsuji, Y., Taneda, Y., Takasawa, T., and Messer, M. (1994). Chemical characterization of sialyl oligosaccharides isolated from tammar wallaby (*Macropus eugenii*) milk. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1200**: 64-72.

Urh, M., Hartzell, D., Mendez, J., Klaubert, D.H., and Wood, K. (2008). Methods for detection of protein-protein and protein-DNA interactions using HaloTag. *Methods Mol Biol* **421**: 191-209.

van Oers, M.M., Pijlman, G.P., and Vlak, J.M. (2015). Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *J Gen Virol* **96**: 6-23.

Verger, A., and Duterque-Coquillaud, M. (2002). When Ets transcription factors meet their partners. *Bioessays* **24**: 362-370.

Verrier, D., Groscolas, R., Guinet, C., and Arnould, J.P.Y. (2011). Development of fasting abilities in subantarctic fur seal pups: balancing the demands of growth under extreme nutritional restrictions. *Functional Ecology* **25**: 704-717.

Veselsky, L., Cechova, D., and Jonakova, V. (1978). Secretion and immunochemical properties of the trypsin inhibitor from bovine colostrum. *Hoppe Seylers Z Physiol Chem* **359**: 873-878.

Veselsky, L., Jonakova, V., and Cechova, D. (1985). A Kunitz type of proteinase inhibitor isolated from boar seminal vesicle fluid. *Andrologia* **17**: 352-358.

Veselsky, L., Jonakova, V., and Cechova, D. (1986). Detection of antigen immunologically related to basic pancreatic inhibitor (Kunitz) in porcine blood plasma by immunoaffinity chromatography. *J Chromatogr* **376**: 395-400.

Vetr, H., and Gebhard, W. (1990). Structure of the human alpha 1-microglobulin-bikunin gene. *Biol Chem Hoppe Seyler* **371**: 1185-1196.

Villarino, A.V., Kanno, Y., Ferdinand, J.R., and O'Shea, J.J. (2015). Mechanisms of Jak/STAT signaling in immunity and disease. *J Immunol* **194**: 21-27.

Vilotte, J.L., Chanat, E., Le Provost, F., Whitelaw, C.B.A., Kolb, A., and Shennan, D.B. (2013). Genetics and Biosynthesis of Milk Proteins. In: Advanced Dairy Chemistry Volume 1: Proteins, P.L.H. McSweeney, and P.F. Fox, eds. (New York, Heidelberg, Dordrecht, London: Springer), pp. 431-461.

Vilotte, J.L., Whitelaw, C.B.A., Ollivier-Bousquet, M., and Shennan, D.B. (2003). Biosynthesis of milk proteins. In: Advanced Dairy Chemistry Volume 1: Proteins, P.F. Fox, and P.L.H. McSweeney, eds. (New York: Kluwer Academic/Plenum Publishers), pp. 699-738.

Voigt, W. (2005). Sulforhodamine B assay and chemosensitivity. *Methods Mol Med* **110**: 39-48.

Vonderhaar, B.K., and Greco, A.E. (1979). Lobulo-alveolar development of mouse mammary glands is regulated by thyroid hormones. *Endocrinology* **104**: 409-418.

Vonderhaar, B.K., and Ziska, S.E. (1989). Hormonal regulation of milk protein gene expression. *Annu Rev Physiol* **51**: 641-652.

Vorbach, C., Capecchi, M.R., and Penninger, J.M. (2006). Evolution of the mammary gland from the innate immune system? *BioEssays* **28**: 606-616.

Voytovich, A.E., and Topper, Y.J. (1967). Hormone-dependent differentiation of immature mouse mammary gland *in vitro*. *Science* **158**: 1326-1327.

Wagner, G., Wutherich, K., and Tschesche, H. (1978). A 1H nuclear-magnetic-resonance study of the conformation and the molecular dynamics of the glycoprotein cow-colostrum trypsin inhibitor. *Eur J Biochem* **86**: 67-76.

Waite, R., Giraud, A., Old, J., Howlett, M., Shaw, G., Nicholas, K., and Familari, M. (2005). Cross-fostering in *Macropus eugenii* leads to increased weight but not accelerated gastrointestinal maturation. *Journal of Experimental Zoology Part A: Comparative Experimental Biology* **303A**: 331-344.

Wakao, H., Schmitt-Ney, M., and Groner, B. (1992). Mammary gland-specific nuclear factor is present in lactating rodent and bovine mammary tissue and composed of a single polypeptide of 89 kDa. *J Biol Chem* **267**: 16365-16370.

Walker, W.A. (2004). The dynamic effects of breastfeeding on intestinal development and host defense. *Adv Exp Med Biol* **554**: 155-170.

Walker, W.A. (2013). Initial intestinal colonization in the human infant and immune homeostasis. *Annals of nutrition & metabolism* **63 Suppl 2**: 8-15.

Wall, J.G., and Pluckthun, A. (1995). Effects of overexpressing folding modulators on the *in vivo* folding of heterologous proteins in *Escherichia coli*. *Current opinion in biotechnology* **6**: 507-516.

Waller, K.P. (2000). Mammary gland immunology around parturition. Influence of stress, nutrition and genetics. *Adv Exp Med Biol* **480**: 231-245.

Wang, J., Wong, E.S., Whitley, J.C., Li, J., Stringer, J.M., Short, K.R., Renfree, M.B., Belov, K., and Cocks, B.G. (2011). Ancient antimicrobial peptides kill antibiotic-resistant pathogens: Australian mammals provide new options. *PLoS One* **6**: e24030.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57-63.

Wang, Y., Li, X., and Hu, H. (2014). H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics* **103**: 222-228.

Wanyonyi, S.S., Lefevre, C., Sharp, J.A., and Nicholas, K.R. (2013a). The extracellular matrix locally regulates asynchronous concurrent lactation in tammar wallaby (Macropus eugenii). *Matrix Biology* **32**: 342-351.

Wanyonyi, S.S., Lefevre, C., Sharp, J.A., and Nicholas, K.R. (2013b). The extracellular matrix regulates *MaeuCath1a* gene expression. *Dev Comp Immunol* **40**: 289-299.

Ward, K.L., and Renfree, M.B. (1984). Effects of progesterone on parturition in the tammar, *Macropus eugenii*. *J Reprod Fertil* **72**: 21-28.

Ward, R.E., and German, J.B. (2004). Understanding milk's bioactive components: a goal for the genomics toolbox. *Journal of Nutrition* **134**: 962S-967S.

Warren, W.C., Hillier, L.W., Marshall Graves, J.A., Birney, E., Ponting, C.P., Grutzner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T.*, et al.* (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **455**: 256-256.

Wasylyk, B., Hahn, S.L., and Giovane, A. (1993). The Ets family of transcription factors. *Eur J Biochem* **211**: 7-18.

Watson, C. (2006). Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. *Breast Cancer Research* **8**: 203.

Watson, C.J. (2009). Immune cell regulators in mouse mammary development and involution. *J Anim Sci* **87**: 35-42.

Watson, C., and Burdon, T. (1996). Prolactin signal transduction mechanisms in the mammary gland: the role of the Jak/Stat pathway. *Rev Reprod* **1**: 1-5.

Watson, C.J., Gordon, K.E., Robertson, M., and Clark, A.J. (1991). Interaction of DNA-binding proteins with a milk protein gene promoter *in vitro*: identification of a mammary gland-specific factor. *Nucl Acids Res* **19**: 6603-6610.

Watson, C.J., and Neoh, K. (2008). The Stat family of transcription factors have diverse roles in mammary gland development. *Semin Cell Dev Biol* **19**: 401-406.

Watson, D.L. (1980). Immunological functions of the mammary gland and its secretion - comparative review. *Australian Journal of Biological Sciences* **33**: 403-422.

Wearne, S.J. (1990). Factor Xa cleavage of fusion proteins: Elimination of non-specific cleavage by reversible acylation. *FEBS Letters* **263**: 23-26.

Weaver, L.T. (1997). Significance of bioactive substances in milk to the human neonate. *Livestock Production Science*
Third International Workshop on the Biology of Lactation in Farm Animals **50**: 139-146.

Weaver, V.M., and Bissell, M.J. (1999). Functional culture models to study mechanisms governing apoptosis in normal and malignant mammary epithelial cells. *J Mammary Gland Biol Neoplasia* **4**: 193-201.

Wei, G.H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R., *et al.* (2010). Genome-wide analysis of ETS-family DNA-binding *in vitro* and *in vivo*. *EMBO J* **29**: 2147-2160.

Weinrauch, Y., Drujan, D., Shapiro, S.D., Weiss, J., and Zychlinsky, A. (2002). Neutrophil elastase targets virulence factors of enterobacteria. *Nature* **417**: 91-94.

Welte, T., Garimorth, K., Philipp, S., and Doppler, W. (1994a). Prolactin-dependent activation of a tyrosine phosphorylated DNA binding factor in mouse mammary epithelial cells. *Mol Endocrinol* **8**: 1091-1102.

Welte, T., Garimorth, K., Philipp, S., Jennewein, P., Huck, C., Cato, A.C., and Doppler, W. (1994b). Involvement of Ets-related proteins in hormone-independent mammary cell-specific gene expression. *Eur J Biochem* **223**: 997-1006.

Welte, T., Philipp, S., Cairns, C., Gustafsson, J.A., and Doppler, W. (1993). Glucocorticoid receptor binding sites in the promoter region of milk protein genes. *J Steroid Biochem Mol Biol* **47**: 75-81.

Wenzel, H.R., and Tschesche, H. (1981). "Chemical mutation" by amino acid exchange in the reactive site of a proteinase inhibitor and alteration or its inhibitor specifity. *Angewandte Chemie International Edition in English* **20**: 295-296.

West, K.L., Oftedal, O.T., Carpenter, J.R., Krames, B.J., Campbell, M., and Sweeney, J.C. (2007). Effect of lactation stage and concurrent pregnancy on milk composition in the bottlenose dolphin. *Journal of Zoology* **273**: 148-160.

Wheelan, S.J., Church, D.M., and Ostell, J.M. (2001). Spidey: a tool for mRNA-to-genomic alignments. *Genome Res* **11**: 1952-1957.

Wheeler, T.T., Hodgkinson, A.J., Prosser, C.G., and Davis, S.R. (2007). Immune components of colostrum and milk--a historical perspective. *J Mammary Gland Biol Neoplasia* **12**: 237-247.

Whitelaw, C.B., Archibald, A.L., Harris, S., McClenaghan, M., Simons, J.P., and Clark, A.J. (1991). Targeting expression to the mammary gland: intronic sequences can enhance the efficiency of gene expression in transgenic mice. *Transgenic Res* **1**: 3-13.

Wicha, M.S., Lowrie, G., Kohn, E., Bagavandoss, P., and Mahn, T. (1982). Extracellular matrix promotes mammary epithelial growth and differentiation *in vitro*. *PNAS* **79**: 3213-3217.

Wilde, C.J., Knight, C.H., and Flint, D.J. (1999). Control of milk secretion and apoptosis during mammary involution. *J Mammary Gland Biol Neoplasia* **4**: 129-136.

Winklehner-Jennewein, P., Geymayer, S., Lechner, J., Welte, T., Hansson, L., Geley, S., and Doppler, W. (1998). A distal enhancer region in the human β-casein gene mediates the response to prolactin and glucocorticoid hormones. *Gene* **217**: 127-139.

Wiseman, B.S., and Werb, Z. (2002). Stromal effects on mammary gland development and breast cancer. *Science* **296**: 1046-1049.

Withers, D.J., and White, M. (2000). Perspective: The insulin signaling system - a common link in the pathogenesis of type 2 diabetes. *Endocrinology* **141**: 1917-1921.

Wobus, M., Wandel, E., Prohaska, S., Findei, S., Tschöp, K., and Aust, G. (2008). Transcriptional regulation of the human CD97 promoter by Sp1/Sp3 in smooth muscle cells. *Gene* **413**: 67-75.

Woelfle, J., Chia, D.J., and Rotwein, P. (2003). Mechanisms of growth hormone (GH) action. Identification of conserved Stat5 binding sites that mediate GH-induced insulin-like growth factor-I gene activation. *J Biol Chem* **278**: 51261-51266.

Wojcik, S.F., Capen, C.C., and Rosol, T.J. (1999). Expression of PTHrP and the PTH/PTHrP receptor in purified alveolar epithelial cells, myoepithelial cells, and stromal fibroblasts derived from the lactating rat mammary gland. *Exp Cell Res* **248**: 415-422.

Wongtrakul-Kish, K., Kolarich, D., Pascovici, D., Joss, J.L., Deane, E., and Packer, N.H. (2013). Characterization of N- and O-linked glycosylation changes in milk of the tammar wallaby (Macropus eugenii) over lactation. *Glycoconj J* **30**: 523-536.

Wood, D.W. (2014). New trends and affinity tag designs for recombinant protein purification. *Curr Opin Struct Biol* **26**: 54-61.

Woodburne, M.O., Rich, T.H., and Springer, M.S. (2003). The evolution of tribospheny and the antiquity of mammalian clades. *Molecular phylogenetics and evolution* **28**: 360-385.

Wooding, P., and Burton, G. (2008). Comparative Placentation: Structures, Functions and Evolution (Berlin, Heidelberg: Springer Verlag).

Woodlee, G.L., Gooley, A.A., Collet, C., and Cooper, D.W. (1993). Origin of late lactation protein from β-lactoglobulin in the tammar wallaby. *J Hered* **84**: 460-465.

Wurm, F.M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotech* **22**: 1393-1398.

Wyszomierski, S.L., and Rosen, J.M. (2001). Cooperative effects of STAT5 (signal transducer and activator of transcription 5) and C/EBPβ (CCAAT/enhancer-binding protein-β) on β-casein gene transcription are mediated by the glucocorticoid receptor. *Molecular Endocrinology* **15**: 228-240.

Wyszomierski, S.L., Yeh, J., and Rosen, J.M. (1999). Glucocorticoid receptor/signal transducer and activator of transcription 5 (STAT5) interactions enhance STAT5 activation by prolonging STAT5 DNA binding and tyrosine phosphorylation. *Mol Endocrinol* **13**: 330-343.

Xiao, S., Shiloach, J., and Betenbaugh, M.J. (2014). Engineering cells to improve protein expression. *Current Opinion in Structural Biology* **26**: 32-38.

Xu, Y., Carr, P.D., Guss, J.M., and Ollis, D.L. (1998). The crystal structure of bikunin from the inter-α-inhibitor complex: A serine protease inhibitor with two kunitz domains. *J Mol Biol* **276**: 955-966.

Yadav, M. (1971). The transmission of antibodies across the gut of pouch young marsupials. *Immunology* **21**: 839-851.

Yamashita, H., Nevalainen, M.T., Xu, J., LeBaron, M.J., Wagner, K.U., Erwin, R.A., Harmon, J.M., Hennighausen, L., Kirken, R.A., and Rui, H. (2001). Role of serine

phosphorylation of Stat5a in prolactin-stimulated β-casein gene expression. *Molecular and Cellular Endocrinology* **183**: 151-163.

Yang, J., Richards, J., Bowman, P., Guzman, R., Enami, J., McCormick, K., Hamamoto, S., Pitelka, D., and Nandi, S. (1979). Sustained growth and three-dimensional organization of primary mammary tumor epithelial cells embedded in collagen gels. *PNAS* **76**: 3401-3405.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**: e15.

Yaswen, P., and Stampfer, M.R. (2002). Molecular changes accompanying senescence and immortalization of cultured human mammary epithelial cells. *Int J Biochem Cell Biol* **34**: 1382-1394.

Yordy, J.S., and Muise-Helmericks, R.C. (2000). Signal transduction and the Ets family of transcription factors. *Oncogene* **19**: 6503-6513.

Yoshimoto, M., and Laskowski, M., Sr. (1982). Large scale purification of isoinhibitors of trypsin from swine colostrum using zinc chelate chromatography and chromatofocusing. *Prep Biochem* **12**: 235-254.

Young, L., Basden, K., Cooper, D.W., and Deane, E.M. (1997). Cellular components of the milk of the tammar wallaby (*Macropus eugenii*). *Australian Journal of Zoology* **45**: 423-433.

Yu-Lee, L.Y. (1997). Molecular actions of prolactin in the immune system. *Proc Soc Exp Biol Med* **215**: 35-52.

Yu-Lee, L.Y., Luo, G., Book, M.L., and Morris, S.M. (1998). Lactogenic hormone signal transduction. *Biol Reprod* **58**: 295-301.

Yu-Lee, L.Y., Richter-Mann, L., Couch, C.H., Stewart, A.F., Mackinlay, A.G., and Rosen, J.M. (1986). Evolution of the casein multigene family: conserved sequences in the 5' flanking and exon regions. *Nucleic Acids Res* **14**: 1883-1902.

Yu, W., Xu, J., Wu, Y., and Yang, G. (2012). A comparative study of mammalian diversification pattern. *International Journal of Biological Sciences* **8**: 486.

Zahn, K. (1996). Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. *Journal of bacteriology* **178**: 2926-2933.

Zahnow, C.A., Cardiff, R.D., Laucirica, R., Medina, D., and Rosen, J.M. (2001). A role for CCAAT/enhancer binding protein β-liver-enriched inhibitory protein in mammary epithelial cell proliferation. *Cancer Res* **61**: 261-269.

Zakharova, E., Horvath, M.P., and Goldenberg, D.P. (2009). Structure of a serine protease poised to resynthesize a peptide bond. *Proc Natl Acad Sci U S A* **106**: 11034-11039.

Zhang, J.X., and Goldenberg, D.P. (1993). Amino acid replacement that eliminates kinetic traps in the folding pathway of pancreatic trypsin inhibitor. *Biochemistry* **32**: 14075-14081.

Zhang, N., and Chen, J.-L. (1998). Purification of recombinant proteins and study of protein interaction by epitope tagging. In: Current Protocols in Molecular Biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. (John Wiley and Sons, Inc.), pp. 10.15.11-10.15.19.

Zhang, H., Lee, L., and Ip, M.M. (2004). The liver-enriched inhibitory protein isoform of CCAAT/enhancer-binding protein β, but not nuclear factor-κB, mediates the transcriptional inhibition of beta-casein by tumor necrosis factor-alpha. *Endocrinology* **145**: 2833-2844.

Zhao, F.Q., and Keating, A.F. (2007). Expression and regulation of glucose transporters in the bovine mammary gland. *J Dairy Sci* **90 Suppl 1**: E76-86.

Zhou, J., and Cidlowski, J.A. (2005). The human glucocorticoid receptor: one gene, multiple proteins and diverse responses. *Steroids* **70**: 407-417.

Zhou, J., Chehab, R., Tkalcevic, J., Naylor, M.J., Harris, J., Wilson, T.J., Tsao, S., Tellis, I., Zavarsek, S., Xu, D.K.*, et al.* (2005). Elf5 is essential for early embryogenesis and mammary gland development during pregnancy and lactation. *EMBO J* **24**: 635-644.

Zhou, Y., Akers, R.M., and Jiang, H. (2008). Growth hormone can induce expression of four major milk protein genes in transfected MAC-T cells. *Journal of Dairy Science* **91**: 100-108.

Zhu, J.W. (2012). Mammalian cell protein expression for biopharmaceutical production. *Biotechnology Advances* **30**: 1158-1170.

Zhuang, S. (2013). Regulation of STAT signaling by acetylation. *Cellular Signalling* **25**: 1924-1931.

Zimmerman, L.M., Vogel, L.A., and Bowden, R.M. (2010). Understanding the vertebrate immune system: insights from the reptilian perspective. *Journal of Experimental Biology* **213**: 661-671.

Ziska, S.E., Bhattacharjee, M., Herber, R.L., Qasba, P.K., and Vonderhaar, B.K. (1988). Thyroid hormone regulation of alpha-lactalbumin: differential glycosylation and messenger ribonucleic acid synthesis in mouse mammary glands. *Endocrinology* **123**: 2242-2248.

Zopf, D., and Roth, S. (1996). Oligosaccharide anti-infective agents. *Lancet* **347**: 1017-1021.

Appendices

Appendix I:	Animal List	376
Appendix II:	Buffers and Solutions	377-386
Appendix III:	Preparation of Competent Cells	387
Appendix IV:	Sequencing Reaction Setup, PCR Conditions and Clean-Up and Primers	388-392
Appendix V:	Optimisation of MBP-tELP Fusion protein production	393-395
Appendix VI:	Northern Analysis of the tammar *ELP* and *CST3* and bovine *CTI* transcripts produced by mammalian expression vectors *in vitro*	396
Appendix VII:	Other publications arising from this thesis	399-454

Appendix I

Animal List

Animals used for tammar wallaby mammary gland explant culture experiments

*Phase 1**

4 & 8 day cultures		12 day culture	
Animal #1: 0833	(d 24P)	Animal #1: 1295	(d 24P)
Animal #2: 1230	(d 24P)	Animal #2: no tag	(d 24P)
Animal #3: 0414	(d 24P)	Animal #3: 0845	(d 24P)

*Phase 2A**

4 & 8 day cultures		14 day culture	
Animal #1: 00-05FE-73F9	(d 76L)	Animal #1: 1583	(d 80L)
Animal #2: 00-05FE-8C6F	(d 76L)	Animal #2: 0855	(d 84L)
		Animal #2: 1116	(d 80L)

*Phase 2B**

4 & 8 day cultures	
Animal #1: 00-05FD-35BC	(d 135L)
Animal #2: 00-05FD-3096	(d 171L)
Animal #3: 00-05FS-25FC	(d 150L)
Animal #4: 00-05FE-911F	(d 127L)

*Phase 3**

18 day culture	
Animal #1: 1768	(d 260L)

* Stage of lactation was based upon the head length measurement of the pouch young (Poole *et al.*, 1991) or on the stage of development *in utero*.

Appendix II

Buffers and Solutions

Protein Solutions

Lower gel buffer (500 mL)

3 mL MQ H_2O
1 mL 0.5 M Tris-HCl pH 6.8 (62.5 mM)
1.6 mL glycerol (20% (v/v))
1.6 mL 10% SDS (2% (w/v))
0.4 mL β-mercaptoethanol (5% (v/v))
0.4 mL 0.5% (w/v) bromophenol blue (in dd H_2O)
Dilute sample at least 1:4 with protein sample buffer and heat at 95°C for 4 mins

Lower gel buffer (500 mL)

91 g Tris base
2 g SDS
400 mL MQ H_2O
Adjust pH to 8.8 with conc. HCl
Make up to 500 mL with MQ H_2O

Upper gel buffer (250 mL)

15 g Tris base
1 g SDS
200 mL MQ H_2O
Adjust pH to 6.8 with conc. HCl
Make up to 250 mL with MQ H_2O

10 X Tris-glycine SDS electrophoresis buffer

30.3 g Tris base (final concentration 250 mM)
144 g glycine (1.92 M)
10 g SDS (1% (w/v))

Coomassie Blue Stain

2 g Coomassie Brilliant Blue powder
1 L methanol
200 mL acetic acid
800 mL MQ H_2O
Stir for a minimum 2 hours and filter through Whatman filter discs

Coomassie blue destain (1 L)

50 mL methanol (5% (v/v))
70 mL acetic acid (7% (v/v))
Make up to 1 L with MQ H_2O

Silver Stain - Fixative

50% methanol
10% acetic acid (v/v) in MQ H_2O

Silver Stain – Fixative rinse

5% methanol
7% acetic acid (v/v) in MQ H_2O

Silver Stain – Pre-treatment (20 mL)

8 mL 25% glutaraldehyde solution (final concentration of 10% (v/v))
12 mL MQ H_2O

Silver Stain (20 mL)

0.02 g $AgNO_3$ (silver nitrate, final concentration of 0.1% (w/v))
20 mL MQ H_2O (make fresh)

Silver Stain – Developer (500 mL)

15 g Na_2CO_3 (final concentration of 3% w/v))
625 µL 40% formaldehyde (final concentration of 0.05% v/v)

Silver Stain – Stop Solution (20 mL)

1 mL of 2.3 M citric acid
(to stop the reaction add 1 mL stop solution per 20 mL developer)

Silver Stain – Gel storage Solution

0.3% Na_2CO_3

10 X TBS (Tris buffered saline) (1 L)

12.1 g Tris base (20 mM)
40 g sodium chloride (1.4 M)
Make up to 400 mL with MQ H_2O and pH to 7.6 with conc. HCl
Add MQ H_2O to 500 mL

TBS-T (Tris buffered saline-Tween) – Wash Buffer(1 L)

100 mL 10 X TBS
1 mL Tween-20 (final concentration 0.1% v/v)
Add MQ H_2O to 1 L and store at 2-8°C

10 X Protein transfer buffer (1 L)

30.3 g Tris (0.25 M)
144 g glycine (1.92 M)
Make up to 1 L with MQ H_2O

1 X Protein transfer buffer (1 L)

100 mL 10 X transfer buffer
700 mL MQ H_2O
200 mL methanol

Blocking buffer (1 L)

50 mL TBS-T
550 µL fish gelatin (Sigma G-7765, 0.5% v/v)
Make fresh each time

Stripping Solution (protein membranes)

103 µL β-mercaptoethanol
2% SDS
62.5 mM Tris-HCl pH 6.7
Incubate at 50°C for 30 min with occasional shaking
Wash 2 X 10 min in TBS-T at RT using a large volume of buffer

RNA Solutions

DEPC-treated H_2O (1 L)

1 mL diethylpyrocarbonate (DEPC) to a final concentration of 0.1%
1 L MQ H_2O
Mix and leave at 37°C for 1 hour or overnight at RT and then autoclave

RNA loading dye (1.5 mL)

187.5 µL MQ H_2O
150 µL 10X MOPS
261 µL formaldehyde
750 µL deoinised formamide
150 µL 10X bromophenol blue
7.5 µL ethidium bromide
Makes 1.5 mL. Dispense into aliquots and store at -20°C

1x RNA Markers (20 µL)

2.5 µL RNA marker (1 X 2.5 µg Promega RNA markers)
7.5 µL MQ H_2O
10 µL RNA loading dye

10 X MOPS (1 L)

41.8 g MOPS (3(N-Morpholino)Propane Sulfonic Acid)
700 mL DEPC H_2O
20 mL 0.5 M EDTA (pH 8)
20 mL sodium acetate (pH 5.2)
Adjust pH to 7.0, make up to 1 L with DEPC H_2O and filter sterlise

20 X SSC (RNA Transfer buffer) (2 L)

350 g NaCl
176 g sodium citrate
Dissolve in 1.5 L MQ H_2O, adjust to pH = 7 with 1.5 M HCl.
Add MQ H_2O to 2 L and autoclave.
7.5 µL MQ H_2O
10 µL RNA loading dye

Prehybridisation Buffer (500 mL)

150 mL deionised formamide
125 mL 20 X SSC
1.95 g $NaH_2PO_4.2H_2O$ (sodium dihydrogen phosphate dihydrate)
1.78 g Na_2HPO_4 (disodium hydrogen phosphate)
0.25 g sodium pyrophosphate decahydrate (final concentration 1 mM)
50 mL 50 X Denhardt's (final = 5X Denhardt's)
Make up to 500 mL and filter sterilize

Before use add salmon sperm DNA (10 mg/mL) to give a final concentration of 100 µg/µL and add SDS to 0.1%. Prehybridise membranes at 42°C

50 X Denhardt's (1 L)

10 g Ficoll (Type 400)
10 g polyvinylpyrrolidone
10 g BSA
Store at -20°C

Wash Solution (post hybridization)

High stringency wash
0.1 X SCC
0.1% SDS
Heat to 60°C and wash 2 X 15 min. Perform a final rinse with cold wash solution

Stripping Solution (RNA membranes)

0.1 X SCC
0.1% SDS
Boil solution and apply to membranes for 2 X 15 min.

DNA Solutions

6 X DNA Loading Buffer (10 mL)

25 mg Bromophenol Blue (final concentration 0.25% (w/v))
4 g sucrose (final concentration 40% (w/v))
25 mg xylene cyanol (final concentration 0.25% (w/v))
Add dd H_2O to 10 mL

6 X Orange G DNA Loading Buffer (50 mL)

0.0125 g Orange G (final concentration of 2%)
25 mL 30% Ficol (final concentration of 15%)
Make up to 50 mL with warm 1 X TAE
Aliquot and store at -20°C

RIPA (Radioimmunoprecitation) buffer (100 mL)

790 mg Tris base
75 mL MQ H_2O
Add 900 mg sodium chloride and stir until dissolved
Adjust the pH to 7.4 with HCl
Add 10 mL NP-40 (10% solution)
2.5 mL sodium deoxycholate (10% solution)
1 mL 100 mM EDTA
Adjust the volume to 100 mL
Store at 2-8°C

Before use add protease and phoshatase inhibitors:
100 µL aprotinin (final concentration 1 µg/mL)
100 µL leupeptin (final concentration 1 µg/mL)
100 µL pepstatin (final concentration 1 µg/mL)
500 µL PMSF (final concentration 1 mM)
500 µL Na_3VO_4 (activated sodium orthovanadate; final concentration 1 mM)
500 µL NaF (sodium fluoride; final concentration 1 mM))

50 X TAE (1 L)

242 g Tris
57.1 mL glacial acetic acid
100 mL 0.5M EDTA, pH 8.0

5 X TBE (1 L)

54 g Tris base
27.5 g boric acid
20 mL 0.5 M EDTA (pH 8.0)

TE buffer

10 mM Tris-HCl (pH 8.0)
1 mM EDTA (pH 8.0)

pMAL Column buffer (1 L)

20 mL 1.0 M Tris-HCl, pH 7.4 (final concentration 20 mM Tris-HCl)
11.7 g NaCl (final concentration 200 mM)
2.0 mL 0.5 M EDTA (final concentration 1 mM)
1 mL 1 M sodium azide (final concentration 1 mM)
1 mL β-ME (β-mercaptoethanol, final concentration 10 mM)

pMAL Rich broth (1 L)

10 g tryptone (1% (w/v))
5 g yeast extract
5 g sodium chloride (0.5% (w/v))
2 g glucose (0.2% (w/v))
Autoclave

2YT broth (1 L)

16 g tryptone
10 g yeast
5 g sodium chloride
800 mL MQ H_2O
pH to 7 with 5 M sodium hydroxide, add MQ H_2O to 1 L and autoclave

LB broth (400 mL)

4 g tryptone
4 g sodium chloride
2 g yeast
Add MQ H_2O to 400 mL and autoclave

LB agar (400 mL)

400 mL LB broth
Add 6g bacteriological agar No. 1 (1.5% w/v) and autoclave

Preparation of glycerol stocks

Remove a 500 µL aliquot from an overnight bacterial culture, spin down (~2,000 *g* for 5 min) and then remove excess media
Add 1 mL of 50% glycerol/50% LB broth
Resuspend by pipetting and store at -80°C

Calcium chloride solution (100 mL) for preparation of chemically competent cells

6 mL calcium chloride
15 mL 100% glycerol
10 mL 0.1 M Pipes, pH 7.0
69 mL MQ H_2O

Tissue and Cell Culture Solutions

Tammar wallaby mammary gland explant culture medium (1 L)

Medium 199 (dry powder)
575 µL BSA (35% Sigma Fraction V) for a final concentration 200 µg/mL
2.2 g sodium bicarbonate ($NaHCO_3$)
4.76 g Hepes
3.4 mL L-glutamine (for a final concentration of 200 mM)
10 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
Add MQ H_2O to 900 mL, pH to 7.55 with 10 M sodium hydroxide (will be pH 7.4 at 37°C) and then add MQ H_2O to 1 L
Filter sterilse (0.2 µm)

Preparation of tammar wallaby primary mammary epithelial cells

Digest medium (150 mL)

15 mL 10 X HBSS (Hanks balanced salt solution)
1.5 mL Fungizone
1.5 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
1.5 mL BSA
150 mg Collagenase 2
150 µL of 500 mM glucose
Add 10 mL digest medium per 1g tissue

Wash medium (500 mL)

50 mL 10 X HBSS (Hanks balanced salt solution)
10 mg DNase I
50 mg trypsin inhibitor

Wallaby proliferation media (500 mL)

225 mL DMEM
225 mL Hams F12 nutrient rich mixture
50 mL FCS
5 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
500 µL 1 mg/mL insulin (for a final concentration of 1 µg/mL)
500 µL 1 mg/mL cortisol (for a final concentration of 1 µg/mL)
500 µL 10 µg/mL EGF (for a final concentration of 10 ng/mL)

Wallaby differation media (500 mL)

225 mL DMEM
225 mL Hams F12 nutrient rich mixture
50 mL FCS
5 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
500 µL 1 mg/mL insulin (for a final concentration of 1 µg/mL)
500 µL 1 mg/mL cortisol (for a final concentration of 1 µg/mL)
2.5 mL 1 mg/mL prolactin (for a final concentration of 5 µg/mL)

HEK 293T Cell culture (500 mL)

440 mL DMEM
50 mL FCS
5 mL L-glutamine
5 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)

HC11 proliferation medium (500 mL)

365 mL MQ H_2O
50 mL FCS
41.6 mL 10 X RPMI
19 mL 7.5% $NaHCO_3$
10 mL 1 M Hepes
4.83 mL 1 X PBS
4.2 mL glutamine
5 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
2.5 mL 1 mg/mL insulin (final concentration of 5 µg/mL)
50 µL EGF (final concentration of 10 ng/mL)

HC11 differentiation medium (500 mL)

365 mL MQ H_2O
50 mL FCS
41.6 mL 10 X RPMI
19 mL 7.5% $NaHCO_3$
10 mL 1 M Hepes
4.83 mL 1 X PBS
4.2 mL glutamine
5 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
2.5 mL 1 mg/mL insulin (final concentration of 5 µg/mL)
2.5 mL 1 mg/mL prolactinn (final concentration of 5 µg/mL)
25 µL 1 mg/mL dexamethasone (to a final concentration of 400 µg/mL; 1 mM). NB: dilute dexamethasone 1:10 with 1 X PBS before use.

Kato-III Cell culture (500 mL)

50 mL RPMI 1640
50 mL fetal calf serum
6.65 mL 7.5% sodium bicarbonate ($NaHCO_3$)
5 mL L-glutamine
5 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
Add 383.35 mL MQ H_2O to a final volume of 500 mL and then filter sterilse (0.2 µm)

KIM-2 proliferation medium (500 mL)

225 mL DMEM
225 mL Hams F12 nutrient rich mixture
50 mL FCS
5 mL L-glutamine
5 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
500 µL 1 mg/mL insulin (to a final concentration of 5 µg/mL)
500 µL 10 µg/mL EGF (to a final concentration of 10 ng/mL)
500 µL linoleic acid (to a final concentration of 5 µg/mL)
Filter sterilse (0.2 µm)

KIM-2 differentiation medium (500 mL)

220 mL DMEM
220 mL Hams F12 nutrient rich mixture
50 mL FCS
5 mL L-glutamine
5 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
500 µL 1 mg/mL insulin (to a final concentration of 5 µg/mL)
500 µL 1 mg/mL prolactin (to a final concentration of 5 µg/mL)
25 µL 1 mg/mL dexamethasone (to a final concentration of 400 µg/mL; 1 mM). NB: dilute dexamethasone 1:10 with 1 X PBS before use.
Filter sterilse (0.2 µm)

KIM-2 Trypsin-TEG– for trypsinising/passaging of cells (200 mL)

1.26 g sodium chloride
0.024 g Na_2HPO_4
0.0432 g KH_2PO_4
0.0666 g KCl
0.18 g D-glucose
0.54 g Tris
360 µL Phenol red (0.05% solution)
Make up to 160 mL with MQ H_2O

Add 20 mL Trypsin (at 10X or 2.5%)
0.08 g EGTA
0.02 g PVA
Adjust pH to 7.6 and make up to 200 mL
Filter sterlise, aliquot and store at -20°C

KIM-2 Collagen-coating solution for flasks (50 mL)

2.8 mL 10% acetic acid (filtered)
5 mL collagen solution
42.2 mL MQ H_2O

Add 10 mL collagen coating solution to 75 cm^2 flask and leave at 4°C overnight. Remove excess solution and then wash 3X with 1 X PBS. Add 1 X PBS and leave to soak for 30 min. Discard and add 20 mL proliferation media. If media turns pink repeat washes until excess acetic acid removed.

Minimal media (100 mL)

10 mL 10 X RPMI
2 mL 1M Hepes
3.8 mL sodium bicarbonate
1 mL vitamins
1 ml non-essential amino acids
1 mL sodium pyruvate
1 mL ITS (insulin-transferrin-selenium)
1 mL 10% BSA (268 µL Sigma fraction V)
1 mL Penicillin/Streptomycin (10,000 U/mL Penicillin G sodium, 10,000 µg/mL Streptomycin sulphate in 0.85% saline)
840 µL L-glutamine
Add MQ H2O to 100 mL and filter sterilse

10 X PBS (phosphate buffered saline) (1 L)

8.00 g of NaCl
0.20 g of KCl
1.44 g of Na_2HPO_4
0.24 g of KH_2PO_4
Dissolve in 800 ml of distilled H_2O
Adjust the pH to 7.4 with HCl or NaOH
Add distilled H_2O to 1 liter.
(10 X PBS pH = 6.8 but when diluted 1 X PBS pH should be 7.4)

Appendix III

Preparation of competent cells

Chemically competent cells

1. Streak out the required cells from a glycerol stock on an LB agar plate with antibiotics (if required) and grow overnight at 37°C
2. Pick a colony and grow overnight in LB broth with antibiotics (if required)
3. Pre-cool centrifuge rotors, tips, solutions, etc.
4. Innoculate 400 mL of LB broth with 1 mL fresh overnight culture growth
5. Grow at 37°C with shaking (~180 rpm) until OD_{550nm} = 0.4 (measure at 2, 3 amd 4 hr). OD_{550nm} 0.2 to 0.5 is acceptable, i.e. when cells are in early to mid log phase of growth
6. Pre-cool centrifuge rotors, tips, solutions, etc.
7. Centrifuge at ~3000 *g* for 15 min at 4°C
8. Resuspend in 40 mL cold (4°C) $CaCl_2$ (calcium chloride)
9. Pellet at ~700 *g* for 10 min at 4°C
10. Resuspend in 40 mL cold (4°C) $CaCl_2$ and incubate on ice for 30 min
11. Pellet at ~700 *g* for 10 min at 4°C
12. Resuspend in 40 mL cold (4°C) $CaCl_2$
13. Aliquot 100 µL to tubes, keep on ice and then snap freeze in liquid N_2
14. Store at -80°C

Electrocompetent cells

1. Streak out the required cells from a glycerol stock on an LB agar plate with antibiotics (if required) and grow overnight at 37°C
2. Pick a colony and grow overnight in LB broth with antibiotics (if required)
3. Pre-cool centrifuge rotors, tips, solutions, etc.
4. Innoculate 1 L of LB broth with 20 mL fresh overnight culture growth
5. Grow at 37°C with vigorous shaking for ~3 hr
6. Chill culture on ice for 15-30 min
7. Centrifuge at ~3000 *g* for 15 min at 4°C
8. Resuspend in 1 L (1:1 vol) of cold (4°C) autoclaved MQ H_2O
9. Pellet at ~3000 *g* for 15 min at 4°C
10. Resuspend in 500 mL of cold (4°C) autoclaved MQ H_2O
11. Pellet at ~3000 *g* for 15 min at 4°C
12. Resuspend in 20 mL of cold (4°C) 10% glycerol (put on ice at step 3)
13. Pellet at ~3000 *g* for 15 min at 4°C
14. Resuspend in 2-5 mL of cold (4°C) 10% glycerol to give an approximate concentration of 3 x 10^{10} cells/mL
15. Aliquot 50 µL to tubes, keep on ice and then snap freeze in liquid N_2
16. Store at -80°C

Appendix IV

Sequencing Reaction Setup, PCR Conditions and Clean-Up and Primers

Table 1. Sequencing reaction set-up for sequencing of plasmid DNA by the Department of Pathology, The University of Melbourne

Reagents	X1
Big Dye v3.1 sequencing mix	1.5 µL
Dilution buffer	4.5 µL
Primer (5 µM)	1.0 µL
MQ H_2O	x µL
DNA (100 ng/kb)*	y µL
TOTAL	**20 µL**

*Plasmid included in calculations

Table 2. PCR conditions for sequencing plasmid DNA

Initial Denaturation	Melting	Annealing	Extension	Cycles
94°C, 5 min	94°C, 10 sec	50°C, 5 sec	60°C, 3min	x30

Sequencing Reaction Clean-Up Protocol

1. Transfer 20 µL sequencing reaction to a 1.5 mL tube
2. Add 80 µL 75% (v/v) isopropanol and vortex for 10 sec
3. Leave at RT for 20 min
4. Spin at high speed (~18,000 *g*) for 20 min
5. Gently pipette off supernatant and pulse spin
6. Pipette off residual supernatant
7. Add 200 µL 75% (v/v) isopropanol and vortex for 5 sec
8. Spin at high speed (~18,000 *g*) for 10 min
9. Gently pipette off supernatant and pulse spin
10. Pipette off residual supernatant
11. Dry either by placing sample at 90°C for 1 min or in a Speedy-Vac for 10 min.

*Sequencing Primers**

*Primer length (bp; base pairs), melting temperature ™ and percentage GC (GC%) are indicated for each primer
^DNA fragment cloned into pGEMT-Easy and so T7 and SP6 used for sequencing

Table 3. Primers used to sequence the tammar (*Macropus eugenii*) *ELP* distal promoter^
A genomic *ELP* fragment (14.7 kb) was originally cloned into lambda EMBL3 SP6/T7 (Simpson, 1998) and then sub-cloned into a modified pBeloBACII vector. The 6.2 kb distal region of the tammar ELP promoter was amplified from the pBeloBAC II-14.7kb tELP plasmid and cloned into the pGEM-T Easy vector.

Gene (region)	Primer F (5' … 3')
Tammar wallaby *ELP* (promoter)	**T7** (20bp,Tm48°C,GC40%)^ TAATACGACTCACTATAGGG
	ELP_265F (30bp,Tm668°C,GC57%) GAGCCTTGCAGAAGGTATGGCTGGAGTCTG
	ELP_581F (30bp,Tm62°C,GC47%) CCTGTGGCTGGCCAATAATGTTCAGATTTC
	ELP_1250F (30bp,Tm59°C,GC40%) GAAGTAGGACACTATGTGATTATAAAGGCC
	ELP_1924F (28bp,Tm61°C,GC50%) GCTCAACTCTCCCAAATGTCAGTCTCAG
	ELP_-2198R (27bp,Tm63°C,GC56%) GCAGAGTTGCACGAAGTCATCAGCCTC
	ELP_-2757R (29bp,Tm63°C,GC52%) CTCCGTTCTCCACCGTGAAATTCCTTCAG
	ELP_-3206R (30bp,Tm63°C,GC50%) GTCTTGGGGCAAGTCACTTAACCTGAGTTC
	ELP_-3738R (30bp,Tm55°C,GC30%) GCTATAGATACATGTATGTGTCATAATTTC
	ELP_-4393R (31bp,Tm59°C,GC39%) CTATTTTCCATCTCCTTGGACTATATTGTGG
	ELP_-5112R (30bp,Tm56°C,GC33%) GGTGCCATTATTATATCCAATTTACAGTTG
	SP6 (20bp,Tm44°C,GC30%)^ ATTTAGGTGACACTATAGAA

Table 4. Primers used to sequence the Victorian Koala (*Phascolarctos cinereus victor*) *ELP* gene^

Gene (region)	Primer F (5' … 3')
Koala *ELP* (gene)	**T7** (20bp,Tm48°C,GC40%)^ TAATACGACTCACTATAGGG
	SP6 (20bp,Tm44°C,GC30%)^ ATTTAGGTGACACTATAGAA
	ELP_KF_348F (28bp,Tm64°C,GC57%) CCAGTGCTCTATCCACTGCGCCATCTAG
	ELP_KR_1109R (29bp,Tm59°C,GC41%) CGTCTGGCTACCATTATCATCATCATTAG
	ELP_KR (28bp,Tm60°C,GC46%) GAAACTCACAATGGCTGTCATCATCACC

Table 5. Primers used to sequence the Brushtail possum (*Trichosurus vulpecula*) *ELP* gene^

Gene (region)	Primer F (5' … 3')
Brushtail possum *ELP* (gene)	**T7** (20bp,Tm48°C,GC40%)^ TAATACGACTCACTATAGGG
	SP6 (20bp,Tm44°C,GC30%)^ ATTTAGGTGACACTATAGAA
	pELP_519F (30bp,Tm60°C,GC43%) CTGAGGACCATTGAGTGTATTTTGCCTTTC
	pELP_993F (27bp,Tm58°C,GC44%) GAGGTTGAACTAGATGGTCTTTAAGGC
	pELP_1455F (29bp,Tm62°C,GC48%) GGAAGTGATGAGGGGCTGTAATAGGATAG
	pELP_3R (30bp,Tm60°C,GC43%) GACAGACCCACGCATATCATTTGAATATGC
	pELP_2R (28bp,Tm63°C,GC54%) AGCCTGTCTGCCTCAGTTTCCTCATCTG
	pELP_R (30bp,Tm62°C,GC47%)^ CAGGCATAAAGATGGAACCAATCGGGATAG

Table 6. Primers used to sequence the Fat-tailed dunnart (*Sminthopsis crassicaudata*) *ELP* gene^

Gene (region)	Primer F (5' … 3')
Fat-tailed dunnart *ELP* (gene)	**T7** (20bp,Tm48°C,GC40%)^ TAATACGACTCACTATAGGG
	SP6 (20bp,Tm44°C,GC30%)^ ATTTAGGTGACACTATAGAA
	scELP_F (29bp,Tm62°C,GC48%) GTCAAGTGTTATCTACTGGCAGCACCATG
	scELP_340F (29bp,Tm64°C,GC55%) GGTTAAGTGACTTGCCCAGGGTCACACAG
	scELP_1021F (30bp,Tm59°C,GC40%) GATTTTTCTTGGGGAGTATGACAAATGTGG
	scELP_1688F (30bp,Tm60°C,GC43%) GGAGATAATCTGATGCCAAGCACCTTATTC
	scELP_4R (30bp,Tm60°C,GC43%) GAGGGTTCCCTTCACCAATAAAATCAGAAG
	scELP_3R (30bp,Tm60°C,GC43%) CTGAGTTCAAATTCAGCCTCACTTCTTAGC
	scELP_2R (27bp,Tm58°C,GC44%) GACTCAGTAGAATGCTGTTTGAAGACC
	scELP_R (30bp,Tm59°C,GC40%) CCCAAAGTGCTGTTAATGCTTTATTGTAGC

Table 7. Primers used to sequence the Cape fur seal (*Arctocephalus pusillus pusillus*) *CTI* gene^

Gene (region)	Primer F (5' … 3')
Cape fur seal *CTI* (gene)	**T7** (20bp,Tm48°C,GC40%)^ TAATACGACTCACTATAGGG
	SP6 (20bp,Tm44°C,GC30%)^ ATTTAGGTGACACTATAGAA
	sCTI_539F (29bp,Tm59°C,GC41%) CTCCCAGATTCTGATTTAATTAGCCTTGG
	sCTI_1099F (30bp,Tm55°C,GC30%) CATTACTCATGGATACTGAAGATTATTTAG
	sCTI_1354R (29bp,Tm62°C,GC48%) GACAAGATCTAGAATGGTCCCCGTTGATG
	sCTI_Ex3R (30bp,Tm64°C,GC53%) CTATAAGAGGACTCACCAGGGTGCTTGCAG

Table 8. Primers used to sequence the Victorian Koala (*Phascolarctos cinereus victor*) *ELP* promoter^

Gene (region)	Primer F (5' … 3')
Koala *ELP* (promoter)	**T7** (20bp,Tm48°C,GC40%)^ TAATACGACTCACTATAGGG
	SP6 (20bp,Tm44°C,GC30%)^ ATTTAGGTGACACTATAGAA

Table 9. Primers used to sequence the Fat-tailed dunnart (*Sminthopsis crassicaudata*) *ELP* promoter^

Gene (region)	Primer F (5' … 3')
Fat-tailed dunnart *ELP* (promoter)	**T7** (20bp,Tm48°C,GC40%)^ TAATACGACTCACTATAGGG
	SP6 (20bp,Tm44°C,GC30%)^ ATTTAGGTGACACTATAGAA
	Scrass_Ex1_R (28bp,Tm63°C,GC54%) GTTTTTCTGAGCTGGTCATGCCAGCCAG

Table 10. Primers used to sequence the Cape fur seal (*Arctocephalus pusillus pusillus*) *CTI* promoter^

Gene (region)	Primer F (5' … 3')
Cape fur seal *CTI* (promoter)	**T7** (20bp,Tm48°C,GC40%)^ TAATACGACTCACTATAGGG
	SP6 (20bp,Tm44°C,GC30%)^ ATTTAGGTGACACTATAGAA
	sELP_SSP (29bp,Tm63°C,GC52%) CTGCAATGTTGAGACAACTAGCTGGGAGG
	sELP_SSP2 (26bp,Tm64°C,GC62%) TTGCTCAGCAGGGAGCCTGCTTCTCC
	sELP_SSP3 (25bp,Tm61°C,GC56%) AGACCTGGCTCCACTGAATTCTGCC
	sELP_PF1 (32bp,Tm59°C,GC38%) CTTAAGTATAACAACTTCTGAGTGCCATGAG T

Table 11. Primers used to sequence the mammalian *ELP* and *CTI* and bacterial[#] *ELP* and *CST3* expression constructs

Primer F (5' … 3')
T7 (20bp,Tm48°C,GC40%) TAATACGACTCACTATAGGG
SP6 (20bp,Tm44°C,GC30%) ATTTAGGTGACACTATAGAA
***malE*[#]** (24bp,Tm61°C,GC58%) GGTCGTCAGACTGTCGATGAAGCC

Appendix V

Optimisation of MBP-tELP fusion protein production

Briefly, 80 mL of rich medium [tryptone (1.0% w/v), yeast (0.5% w/v), NaCl (0.5% w/v) and glucose (0.2% w/v)] supplemented with ampicillin (100 μg/mL), (Appendix II) was inoculated with 800 μL of a 5 mL overnight culture from a single colony and grown at 37°C on a shaking platform (300 rpm) to an OD_{600nm} ~0.5-0.6. A 1 mL aliquot (Sample1: Uninduced cells) was removed, spun for 2 min, the supernatant discarded and the cell pellet resuspended with SDS-PAGE sample buffer (50 μL) and stored at -20°C. Fusion protein expression was induced with IPTG at four different concentrations (0.3 mM, 0.8 mM, 1 mM and 2 mM) for 2 h, 150 rpm at 37°C, 30°C, and 28°C. A 500 μL sample (Sample 2: Induced cells) was removed and processed as for sample 1, but the cell pellet resuspended in 100 μL of SDS-PAGE sample buffer. The remainder of the culture was divided in two. The first half was used to determine whether the pMAL-c2-tELP and pMAL-p2-tELP/pMAL-c2-tCST3 and pMAL-p2-tCST3 vectors produced a soluble or insoluble fusion protein. The second half was used to assess whether the fusion protein was produced in the periplasm (pMAL-p2-tELP/ pMAL-p2-tCST3 constructs only). Both sets of cultures were centrifuged at 4000 *g* for 10 min and the supernatants discarded.

In order determine whether the fusion proteins were soluble, the first cell pellet was resuspended in 5 mL Column Buffer, pH 8.0 [20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM sodium azide, 10 mM β-ME (β-mercaptoethanol; Appendix II)] and frozen overnight (-20°C). The next day, cells were thawed in cold water and sonicated in an ice-water bath for 2 min (15s sonication and 15s pause, x4) with an output power of 4W (Virsonic Ultrasonic Cell Disruptor 100, Virtis, USA). The lysate was centrifuged at 9,000 *g*, 4°C for 20 min, the supernatant (soluble crude extract fraction) decanted and placed on ice. The pellet was resuspended in 5 mL Column Buffer (insoluble fraction). 2X SDS-PAGE sample buffer (5 μL) was added to 5 μL of each of the crude (soluble) and insoluble fractions (Samples 3 and 4) respectively.

The second cell pellet (pMAL-p2-tELP/pMAL-p2-tCST3 constructs only, periplasmic fraction) was resuspended in 10 mL 30 mM Tris-HCL, 20% sucrose (w/v) and frozen overnight. Cells were thawed and the osmotic-shock sample prepared. After the addition of 0.5 M EDTA (20 μL), the cells were incubated at RT for 10 min on a shaking platform, centrifuged at 8,000 *g*, 4°C for 20 min and the supernatant discarded. The cell pellet was resuspended in ice-cold 5 mM $MgSO_4$ (10 mL), shaken for an additional 10

min in an ice-cold water bath and centrifuged at 8,000 *g*, 4°C for 20 min. Supernatant (cold osmotic shock fluid, 10 µL) was added to 10 µL of 2X SDS-PAGE sample buffer (Sample 6: periplasmic extract, osmotic shock fraction sample) and frozen (-º20C).

Additional steps to increase the solubility of the MBP-tELP MBP-tCST3 fusion proteins were tested. These included: the addition of 100 µg/mL lysosyme (#L6876, Sigma) to bacterial cells and incubation on ice for 40 min prior to sonication, and the addition of Triton X-100 detergent (#T8787, Sigma; 0.05%, 0.1% or 0.2% v/v) just before sonication. Optimisation of MBP-tCST3 was stopped at this stage.

The binding affinity of MBP-tELP (soluble crude extract) for the amylose resin (#E8021S, NEB) and hence the ability to purify the fusion protein was tested. Resin (200 µL) was spun briefly and the buffer discarded. The resin was spun again, the remaining buffer removed and the resin resuspended in column buffer (200 µL). The crude extract (50 µL) was mixed with amylose resin slurry (50 µL), incubated on ice for 15 min, centrifuged for 1 min and the supernatant discarded. The cell pellet was washed with column buffer (1 mL), centrifuged for 1 min and resuspended in 50 µL SDS-PAGE sample buffer (Sample 5: fusion protein bound to amylose resin). After optimisation, MBP-tELP production was scaled up to 1 L culture volumes.

Purification of soluble MBP-tELP from the E. coli cytoplasm

Affinity purification of MBP-tELP with an amylose resin column (2.5 diameter x 10 cm high, #7372512, Bio-Rad) was done in a 4°C coolroom. Briefly, resin (15 mL) was poured into the column, allowed to settle and then washed with column buffer (8 vol). MBP-tELP was bound to the amylase resin by applying the crude cell extract (diluted with column buffer to ~2.5 mg/mL) at a flow rate of ~1 mL/min. Non-specific *E. coli* proteins were washed from the resin with column buffer (12 vol). MBP-tELP was eluted from the resin with column buffer containing 10 mM maltose (#M9171, Sigma). Sequential 3 mL fractions were collected and analysed by spectrophotometer (A_{280nm}) and Western analysis. Fractions containing MBP-tELP were pooled and concentrated to ~1 mg/mL under N_2 with an Amicon Ultrafiltration Cell (#8050, Millipore) and a 30 kDa regenerated cellulose NMWL (nominal molecular weight limit) ultrafiltration membrane (#PLTK04310, Amicon, Millipore). The amylose resin column (which was re-used up to ten times) was regenerated promptly after each use by washing sequentially with MQ water (3 vol), 0.1% SDS (3 vol), MQ water (3 vol) and then a final rinse with column buffer (3 vol).

Optimisation of Factor Xa cleavage of MBP-tELP

Cleavage of tELP from its MBP fusion partner was initially carried out at RT for 24 h, using 1 mg Factor Xa (#P8010S, NEB) per 100 mg fusion protein, as per the manufacturer's instructions. The addition of SDS (0.005%, 0.01%, 0.025% and 0.05% w/v), increased incubation times of up to five days and the denaturation of MBP-tELP were also tested to improve Factor Xa cleavage.

MBP-tELP was denatured by dialysis [6,000-8,000 MWCO dialysis tubing (#132-650, Spectrum Labs)] against at least 10 vol of 20 mM Tris-HCl, 6 M guanidine hydrochloride, pH 7.4 for 4 h and then 100 vol of column buffer for 2X 4 h and then cleaved with Factor Xa. The sample was concentrated to ~1 mL with Microsep 1K Omega spin columns (1,000 Da MWCO, #OD001C40, Pall Corporation).

Appendix VI

Northern Analysis of the tammar *ELP* and *CST3* and bovine *CTI* transcripts produced by mammalian expression vectors transfected into HEK293Ts

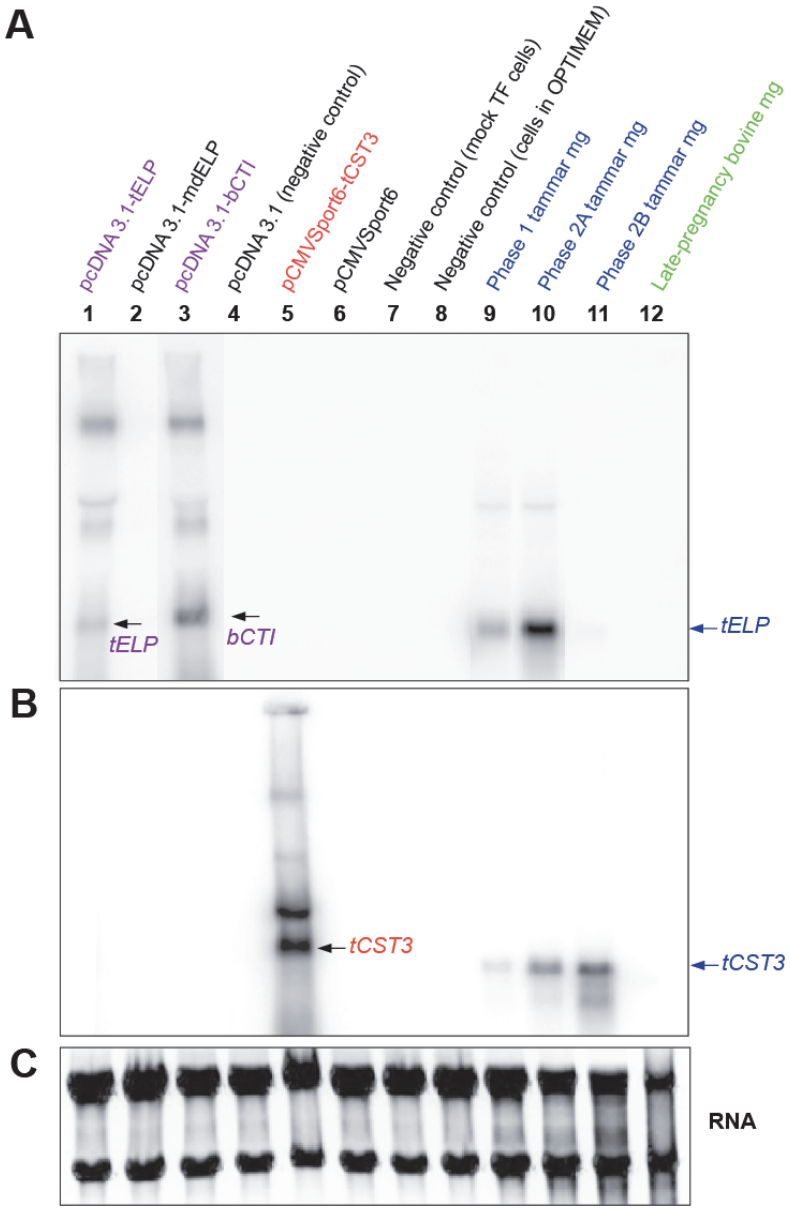


Figure 1. Northern analysis of tammar *ELP* and *CST3* and bovine *CTI* transcripts produced by the pcDNA3.1-tELP, pcDNA3.1-bCTI and pCMVSport6-CST3 mammalian expression vectors.

A. Tammar *ELP*, bovine *CTI* and **B.** tammar *CST3* transcripts at the expected sizes were detected in RNA islolated from HEK293T cells transfected with pcDNA3.1-tELP, pcDNA3.1-bCTI and pCMVSport6-CST3 respectively. However, while the pCMVSport6-tCST3 resulted in high levels of tammar *CST3* transcripts and a high level of CST3 secretion into conditioned media, this is not so for *ELP* and *CTI* (Chapter 6). Control total RNA from the Phase 1, 2A and 2B tammar mammry gland and the late-pregnant bovine mammary gland are shown. The tELP and bCTI results have been superimposed on the same membrane (**A**). **C.** Ribosomal RNA bands show RNA integrity and loading.

Appendix VII:

Other publications arising from this thesis

De Leo, A.A., Lefevre, C., Topcic, D., **Pharo, E.**, Cheng, J.F., Frappell, P., Westerman, M., Graves, J.A.M., and Nicholas, K.R. (2006). Characterization of two whey protein genes in the Australian dasyurid marsupial, the stripe-faced dunnart (*Sminthopsis macroura*). *Cytogenet Genome Res* **115**: 62-69.

Renfree, M.B., Papenfuss, A.T., Deakin, J.E., Lindsay, J., Heider, T., Belov, K., Rens, W., Waters, P.D., **Pharo, E.A.**, Shaw, G., Wong, E.S., Lefevre, C.M., Nicholas, K.R., Kuroki, Y., Wakefield, M.J., Zenger, K.R., Wang, C., Ferguson-Smith, M., Nicholas, F.W., Hickford, D., Yu, H., Short, K.R., Siddle, H.V., Frankenberg, S.R., Chew, K.Y., Menzies, B.R., Stringer, J.M., Suzuki, S., Hore, T.A., Delbridge, M.L., Patel, H.R., Mohammadi, A., Schneider, N.Y., Hu, Y., O'Hara, W., Al Nadaf, S., Wu, C., Feng, Z.P., Cocks, B.G., Wang, J., Flicek, P., Searle, S.M., Fairley, S., Beal, K., Herrero, J., Carone, D.M., Suzuki, Y., Sugano, S., Toyoda, A., Sakaki, Y., Kondo, S., Nishida, Y., Tatsumoto, S., Mandiou, I., Hsu, A., McColl, K.A., Lansdell, B., Weinstock, G., Kuczek, E., McGrath, A., Wilson, P., Men, A., Hazar-Rethinam, M., Hall, A., Davis, J., Wood, D., Williams, S., Sundaravadanam, Y., Muzny, D.M., Jhangiani, S.N., Lewis, L.R., Morgan, M.B., Okwuonu, G.O., Ruiz, S.J., Santibanez, J., Nazareth, L., Cree, A., Fowler, G., Kovar, C.L., Dinh, H.H., Joshi, V., Jing, C., Lara, F., Thornton, R., Chen, L., Deng, J., Liu, Y., Shen, J.Y., Song, X.Z., Edson, J., Troon, C., Thomas, D., Stephens, A., Yapa, L., Levchenko, T., Gibbs, R.A., Cooper, D.W., Speed, T.P., Fujiyama, A., Graves, J.A., O'Neill, R.J., Pask, A.J., Forrest, S.M., and Worley, K.C. (2011). Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol* **12**: R81.

Original Article

Cytogenet Genome Res 115:62–69 (2006)
DOI: 10.1159/000094802

Cytogenetic and Genome Research

Characterization of two whey protein genes in the Australian dasyurid marsupial, the stripe-faced dunnart *(Sminthopsis macroura)*

A.A. De Leo[a] C. Lefevre[a, b] D. Topcic[a] E. Pharo[a] J.F. Cheng[c] P. Frappell[d] M. Westerman[e] J.A.M. Graves[a, f] K.R. Nicholas[a]

[a]CRC for Innovative Dairy Products, Department of Zoology, The University of Melbourne, Melbourne
[b]Victorian Bioinformatics Consortium, Monash University, Clayton (Australia)
[c]U.S. Department of Energy, Joint Genome Institute, Walnut Creek, CA (USA)
[d]Department of Zoology; [e]Department of Genetics, LaTrobe University, Bundoora
[f]Comparative Genomics Research Group, Research School of Biological Sciences, Australian National University, Canberra (Australia)

Manuscript received 18 October 2005; accepted in revised form for publication by M. Schmid, 25 January 2006.

Abstract. We report the first isolation and sequencing of genomic BAC clones containing the marsupial milk protein genes Whey Acidic Protein *(WAP)* and Early Lactation Protein *(ELP)*. The stripe-faced dunnart *WAP* gene sequence contained five exons, the middle three of which code for the *WAP* motifs and four disulphide core domains which characterize WAP. The dunnart *ELP* gene sequence contained three exons encoding a protein with a Kunitz motif common to serine protease inhibitors. Fluorescence in situ hybridization located the *WAP* gene to chromosome 1p in the stripe-faced dunnart, and the *ELP* gene to 2q. Northern blot analysis of lactating mammary tissue of the closely related fat-tailed dunnart has shown asynchronous expression of these milk protein genes. *ELP* was expressed at only the earlier phase of lactation and *WAP* only at the later phase of lactation, in contrast to β-lactoglobulin *(BLG)* and α-lactalbumin *(ALA)* genes, which were expressed in both phases of lactation. This asynchronous expression during the lactation cycle in the fat-tailed dunnart is similar to other marsupials and it probably represents a pattern that is ancestral to Australian marsupials. Copyright © 2006 S. Karger AG, Basel

Marsupials have evolved a unique reproductive strategy of short gestation, birth of immature young and a relatively long lactation period, which contrasts with the relatively long gestation period of eutherian mammals (Tyndale-Biscoe and Janssens, 1988). Whereas eutherians usually produce milk of a constant composition after the expression of the initial colostrum (Jenness, 1986), marsupials progressively change the composition of milk during lactation (Green, 1984; Nicholas, 1988). The components of marsupial milk provide important nutrients and putative growth factors for the development of the young (Ballard et al., 1995; Trott et al., 2003), and may regulate the function of the mammary gland by autocrine/paracrine mechanisms (Nicholas et al., 1997; Hendry et al., 1998). Macropod marsupials, such as the tammar wallaby *(Macropus eugenii)* and red kangaroo *(Macropus rufus)*, are able to feed a pouch young and an older animal at heel with milk of different constituents from adjacent mammary glands (Bailey and Lemon, 1966; Griffiths et al., 1972; Nicholas, 1988; Nicholas et al., 1997). Therefore, milk secretion is likely to be controlled by endocrine stimuli, and other factors intrinsic to the mammary gland (Nicholas et al., 1997, 2001). The molecular mechanisms which regulate marsupial milk composition are not well understood.

The cDNA for the Early Lactation Protein *(ELP)* gene has been isolated from the lactating mammary gland of two Australian marsupials from the order Diprotodontia, the tammar wallaby *(Macropus eugenii)* (Simpson et al., 1998b)

Funding for this project was provided by the CRC for Innovative Dairy Products, Australia.

Request reprints from A.A. De Leo
CRC for Innovative Dairy Products, Department of Zoology
The University of Melbourne, Melbourne, Victoria, 3010 (Australia)
telephone: +61 3 9344 6256; fax: +61 3 8344 7909
e-mail: a.deleo@pgrad.unimelb.edu.au

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2006 S. Karger AG, Basel
1424–8581/06/1151–0062$23.50/0

Accessible online at:
www.karger.com/cgr

and brush-tailed possum *(Trichosurus vulpecula)* (Piotte and Grigor, 1996). Marsupial *ELP* is expressed only in early lactation (Piotte and Grigor, 1996; Simpson et al., 1998b), suggesting that it has a specific temporal role in the development of either the pouch young or the mammary gland. The marsupial *ELP* gene belongs to the Kunitz family of serine protease inhibitors and has significant homology to the sequence of bovine colostrum inhibitor (Simpson et al., 1998b). A potential role of the *ELP* protein as a protease inhibitor to protect immunoglobulins in the milk from proteolysis has been suggested (Piotte and Grigor, 1996; Simpson et al., 1998b). However, the function and potential physiological substrate of *ELP* has yet to be established.

Whey Acidic Protein *(WAP)* is a major whey protein gene, expressed throughout lactation in many eutherian species, including rat, mouse (Henninghausen et al., 1982), rabbit (Devinoy et al., 1988), camel (Beg et al., 1986) and pig (Simpson et al., 1998a). However, it is not found in ruminants or primates (including human) milk (Clauss et al., 2002; Rival-Gervier et al., 2003; NCBI Accession AY267466) suggesting that the gene either became inactive due to DNA mutations or was lost after the divergence of these lineages. A recent study using mouse *Wap* gene knockouts has shown that the expression of WAP is necessary for adequate nourishment of the offspring but not functional differentiation of mammary epithelial cells (Triplett et al., 2005). WAP is asynchronously secreted in the milk of diprotodontid marsupials including the tammar (Simpson et al., 2000), red kangaroo (Nicholas et al., 2001) and possum (Demmer et al., 2001). Partial WAP protein sequences have also been determined from the milk of the two monotremes, platypus and echidna (Simpson et al., 2000) but the pattern of secretion in these species is yet to be established. WAP protein secretion has not been characterized in dasyurid marsupials.

Alignment of the amino acid sequence of WAP proteins from many marsupial and eutherian species shows limited sequence identity of amino acids (Simpson et al., 2000). However, these genes are all recognized by the presence of the WAP motif (KXGXCP) and a two-domain structure known as the four disulphide core (4-DSC) domain (Ranganathan et al., 1999), which consists of eight cysteine residues (Hennighausen et al., 1982). The 4-DSC may be a required conformation for the action and stable structure of a specific class of protease inhibitors (Ranganathan et al., 1999).

The aim of the present study was to isolate and sequence large insert Bacterial Artificial Chromosome (BAC) clones containing the *ELP* and *WAP* genes from the dasyurid marsupial species *Sminthopsis macroura* (the stripe-faced dunnart). The presence of the *ELP* and *WAP* genes in dasyurid marsupials (which are only distantly related to the previously studied tammar wallaby and possum) would imply that both genes are ubiquitous at least in Australian marsupials. Furthermore, the long and homologous DNA segments of the BAC clones are ideal for fluorescence in situ hybridization (FISH) mapping to provide the first reliable map position of marsupial milk protein genes.

Materials and methods

Library screening

The BAC library RZPD 688 stripe-faced dunnart *(S. macroura)* (Chapman et al., 2003) was screened for the *ELP* and *WAP* genes. The tammar *WAP* cDNA (Accession AJ005356) and *ELP* cDNA (Accession AJ000490) were labeled with [α-^{32}P]-dCTP and used for probing the libraries as described in Sambrook et al. (1989). Positive signals were identified and DNA extracted from the individual clones as described in Sambrook et al. (1989).

Identifying and sequencing the ELP and WAP genes in the BACs

DNA from the stripe-faced dunnart BACs with either the *ELP* or *WAP* gene were subcloned using the TOPO Shotgun kit (Invitrogen, USA) according to the manufacturer's instructions. Subclones were screened for *ELP* and *WAP* exons, and positives sequenced as described in Sambrook et al. (1989). The entire stripe-faced dunnart BAC containing the *ELP* gene was sequenced by a fluorescent shotgun method. Briefly, plasmid DNA was sequenced from both directions, and sequence reads were then assembled using the Phrap assembler (Ewing and Green, 1998). The 14 kb of DNA surrounding the stripe-faced dunnart *WAP* gene (including 7 kb of DNA 5′ of the first exon) was sequenced by the Australian Genome Research Facility (Queensland, Australia), using primers designed from the tammar *WAP* cDNA sequence.

Bioinformatics and BAC sequence analysis

Gene exon/intron structure was determined by comparing genomic DNA sequence with other marsupial *WAP* and *ELP* cDNAs in the GenBank database. In addition GENESCAN (http://genes.mit.edu/GENSCAN.html) was used to predict the putative cDNA and protein. Predicted protein alignments were carried out using ClustalW (http://clustalw.genome.jp/ or http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html).

A seven-fold coverage of short genomic sequences of approximately 1 kb were available for the didelphid marsupial, the Brazilian short-tailed gray opossum *(Monodelphis domestica)* (at www.ncbi.nlm.nih.gov/BLAST/tracemb.shtml). The gray short-tailed opossum *WAP* and *ELP* sequences were assembled from short genomic sequences retrieved using the cDNA of *WAP* and *ELP* from the stripe-faced dunnart, tammar (Accession AJ000490, AJ005356) and brush-tailed possum (Accession AF275314) to Blast the gray short-tailed opossum genome to obtain matching sequence reads and their mate pairs. The sequences were assembled using phrap (http://www.phrap.org/). Assembled fragments were used to retrieve further the gray short-tailed opossum overlapping sequences together with mate pairs in order to extend local assemblies around *WAP* and *ELP* genes.

RT PCR of cDNA for the milk protein genes

Tissue was excised from the lactating mammary gland of the fat-tailed dunnart, homogenized and total RNA extracted using the Qiagen RNeasy Lipid Tissue mini kit (Melbourne). The total RNA (2 μg) was reverse transcribed (RT) using a SuperScript^{TM} II kit (Invitrogen, Melbourne), following the manufacturer's recommended protocol.

Milk protein gene DNA primers were designed from both eutherian and marsupial β-lactoglobulin *(BLG)* and α-lactalbumin *(ALA)* by aligning sequences from GenBank. However, only marsupial *WAP* genes and marsupial *ELP* genes were aligned for primer design, because eutherian and marsupial *WAP* gene sequences differ greatly, and *ELP* has not been identified in eutherians. Highly conserved regions were selected for the priming region. Lactating mammary gland template cDNA (50 ng), 5 μl 10× buffer, 1 μl 10 mM each forward and reverse primer (Table 1), 1 μl 10 mM dNTPs and 2.5 units Taq Polymerase in a total volume of 50 μl were PCR amplified by initial denaturization at 94°C for 2 min, then 35 cycles at 94°C for 1 min, at 47°C for 30 s, at 72°C for 1 min, and a final extension of 72°C for 10 min. All PCR products showed a single band when electrophoresed on a 0.8% agarose gel and the bands were subsequently excised and DNA extracted using the Qiagen gel extraction kit. The DNA was ligated into pGEM T-easy and clones sequenced using primer sites T7 and SP6 in the vector.

Table 1. Primers used in PCR of cDNA from lactating mammary gland

Gene	Forward primer 5′-3′	Reverse primer 5′-3′
WAP	exon2F AGAAGGCTGGGCGCTGCCC	exon3R CATCGCAGTCCAGGTCAG
ELP	exon1F ATGAAATTCACCATCATTGCCCTCT	exon3R GCAGAAGAGGGAATCATAGAGGG
ALA	exon2F CTGCTCCTCTTTGCTCCTGCT	exon5R ACTGGTCCAGGTCCTCAATG
BLG	exon2F AATTCCCTTGTTTGACATGGAC	exon6R GCTTTCCCTAAAGARGGCAC

Expression of the WAP and ELP genes analyzed by Northern blotting

Total RNA (10 μg, as determined by spectrophotometer) was electrophoresed together with markers of known size and concentration in a 1% agarose gel with 1× MOPS and 1.1% formaldehyde. The RNA was transferred to Zeta probe membrane and the Northern blot was probed (Sambrook et al., 1989) with the full length tammar cDNA for *ELP* (Accession AJ000490), *WAP* (Accession AJ005356), *ALA* (Accession X15211) and *BLG* (Accession L14954-L14960) genes.

Fluorescence in situ hybridization (FISH)

Metaphase spreads were prepared from the stripe-faced dunnart and FISH was carried out as previously described (Toder et al., 1996). The stripe-faced dunnart BACs (237 L6-ELP, 002 F17-WAP) were used as probes. Slides were viewed under a Zeiss Axioplan microscope, images collected with software package Spot Advance and pictures processed using Image J, Confocal Assistant, Adobe Illustrator 7.0, and Adobe Photoshop. At least ten metaphase spreads were captured that had three or four signals of a possible maximum of four.

Results

Isolation and sequencing of the stripe-faced dunnart ELP and WAP genes

One *ELP* positive BAC (237 L6) was recovered from the stripe-faced dunnart BAC library and was completely sequenced by the shotgun method. The BAC was found to be 80 kb in length and to contain the entire *ELP* gene. One *WAP* positive BAC (002 F17) was detected by screening the stripe-faced dunnart library with the *WAP* probe. Approximately 14 kb was sequenced and found to include 7 kb of promoter and the first four exons of the stripe-faced dunnart *WAP* gene.

Comparison of the marsupial ELP sequences

Comparison of a large set of *ELP* EST (cDNA) sequences from the tammar mammary gland pooled from all phases of lactation (the database is held by the Collaborative Research Centre for Innovative Dairy Products, Melbourne, Australia) provided 510 bp of contiguous sequence for *ELP*.

Alignment of the stripe-faced dunnart and gray short-tailed opossum *ELP* genomic sequences with the tammar cDNA from the EST library was carried out to determine the exon/intron boundaries in the genomic regions (Fig. 1a). Stripe-faced dunnart and gray short-tailed opossum genomic *ELP* aligned with tammar *ELP* cDNA predicted that the complete coding sequence of the dunnart and opossum *ELP* is contained in three exons (Fig. 1a). In the stripe-faced dunnart the first exon contained a 5′ untranslated region plus a region coding for the first 24 amino acids of ELP, including a signal peptide sequence. The second exon comprised most of the coding region. Exons 2 and 3 encode the stop codon (TAG). The last nucleotide in exon 2 is T, and the first two nucleotides in exon 3 are AG. Exon 3 contains 154 bp up to the canonical transcription termination signal (AATAAA) at the 3′ end of the gene.

An NCBI BlastP search of the stripe-faced dunnart *ELP* predicted protein revealed the best alignment to be with ELP proteins of other marsupials including the tammar and the brush-tailed possum. The best alignment with the human genome was tissue factor pathway inhibitor 2 *(TFPI2)* located on human chromosome 7q22. The stripe-faced dunnart *ELP* sequence also has homology with eutherian genes coding for a Kunitz motif, including human *LOC 391253* and *WAP8*, bovine trypsin inhibitor, bovine colostrum trypsin inhibitor, and both sheep and bovine trophoblast Kunitz domain proteins.

Identification of other sequences in the stripe-faced dunnart ELP BAC

A second gene, phosphatidyl inositol glycan class T *(PIGT)*, was found in the stripe-faced dunnart *ELP* BAC by BlastP analysis of the GENESCAN predicted protein; stripe-faced dunnart PIGT shared nearly 72% protein identity with mouse, rat and human PIGT. The GENESCAN program searches for conserved start codons, 3′ and 5′ splice sites, canonical TATA boxes, and stop codons. Therefore, the prediction by GENESCAN that the *PIGT* sequence represents a functional gene, together with the sequence similarity between the stripe-faced dunnart and eutherians, strongly suggests that the stripe-faced dunnart *PIGT* is a functional gene rather than a pseudogene. The BAC also contained two identifiable repetitive elements: one with 42% protein identity with a long interspersed sequence element (LINE 1) in human and other eutherian genomes, and the second which showed no significant matches when subjected to BlastP, but a match of 100% with 24 bases of a human MER repeat element revealed by BlastN (Fig. 2).

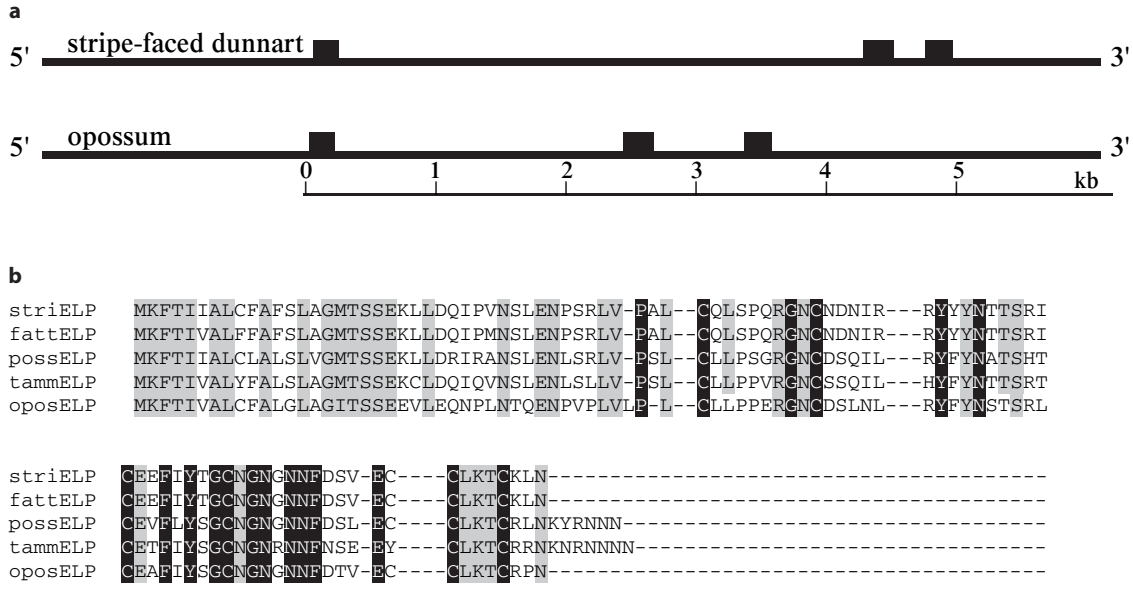


Fig. 1. Structure of the *ELP* gene. (**a**) Comparison of marsupial *ELP* gene structure. The black squares represent exons in the stripe-faced dunnart and opossum. Both species have three exons of almost identical size, but the introns vary in size. (**b**) Alignment of the stripe-faced dunnart ELP protein with other marsupial ELP proteins. The conserved Kunitz motif sequences are shown in white type with a black background. Other matches between marsupial ELP proteins are shown in black type with a grey background. striELP = Stripe-faced dunnart ELP conceptually translated from the BAC sequence; fattELP = fat-tailed dunnart ELP conceptually translated from the cDNA sequence; possELP = brush-tailed possum ELP gi1002806; tammELP = tammar wallaby ELP gi3046758; and oposELP = gray short-tailed opossum ELP, assembled from the short sequences available at www.ncbi.nlm.nih.gov/BLAST/tracemb.shtml.

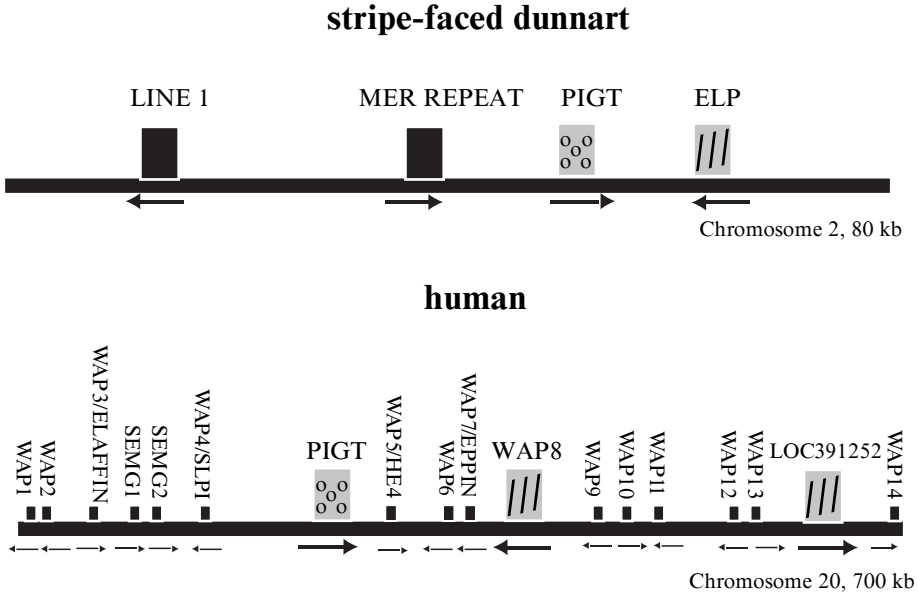
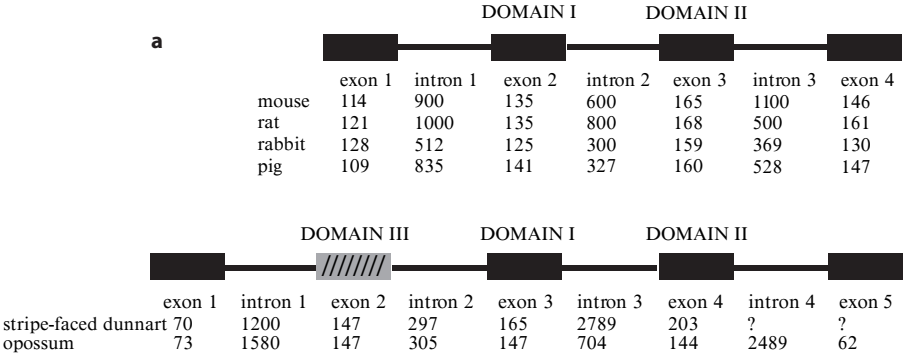


Fig. 2. Comparison of human and stripe-faced dunnart genomic regions surrounding the *ELP*/*PIGT* genes. The arrows represent the direction of transcription. *PIGT* is represented by a grey background with black spots. *ELP* in the stripe-faced dunnart is represented by a grey background with black stripes as are the human genes that have sequence homology to stripe-faced dunnart *ELP*. The black boxes represent other genes. The human map is adapted from Clauss et al. (2002). The stripe-faced dunnart chromosome position has been assigned to chromosome 2 by FISH mapping (see Fig. 4b).

Characterization of marsupial WAP genes

The structures of the stripe-faced dunnart and gray short-tailed opossum *WAP* genomic sequences were determined by comparing sequence with the tammar and possum *WAP* cDNAs in order to identify the five exons. The gray short-tailed opossum *WAP* gene genomic sequence was sourced from the database, as described in Materials and methods, and shown by alignment with the tammar and possum cDNAs to have five exons (Fig. 3a). The limited sequence available from the partial sequencing of the stripe-faced dunnart genomic *WAP* identified the first four of these five exons.



b

```
LEADER PEPTIDE
striWAP  M-QSVQILTLVLL-ALG-AWA-A-QNSVE------
possWAP  MMQSVQILTLVLL-ALG-AWA-A-QNSIE------
tammWAP  M-QPVQILTLVLL-ALG-AWA-A-QESTE------
oposWAP  MMQSVQILTLALL-ALG-SWA-A-QNLEE------

DOMAIN III
striWAP  KACRCPNFRQVLSDRTECKQMCHGDSSCPQDLKCCPQGCNWLCMNTTQE
fattWAP  ----CPNFRQVLSDRTECKQMCHGDSSCPQDLKCCPQGCNWLCMNTTQE
possWAP  KACYCPDFRHVLSDIRDCKQMCNNDASCPQNLRCCQRGCSWLCLNTTQE
tammWAP  KAGYCPDFRQVLLDRRDCKQLCNDDASCPQNMRCCQRGCSWLCMNTTQE
oposWAP  KSGYCPVLHPTLSDLIDCKRTCQNDTSCPGEMKCCRQGCSWMCKNPMKE
CONSENSUSK G CP          C  C    C    CC   C   C

DOMAIN I
striWAP  KDGSCPLATFSSSSSEEHHLNPLCG--KACNTDLDC-DGEAKCCV-SS-CGQ-TCTMPIRV-------
fattWAP  KDGSCPLATFSFSSSEEHHLNPLCG--KACHTDLDC-DGEAKCCV-SS-CGQ-TCTMPIRV-------
possWAP  KDGMCPMGTSSSSSSSEQWRNQLCN--STCKTDLDC-AGEAKCCA-SS-CGR-ICSMPIKA-------
tammWAP  KDGLCPVATSHSSSSEEQQRKQLCD--KTCKTDLGC-EGKAKCCA-SS-CGQ-TCFMPVKA-------
oposWAP  KACLCPLAT---SSSEERLP---CGIP--CETDKDCD-GEAKCCVRS--CGRT-CSMPIKV-------
CONSENSUSK G CP          C    C    C    CC   C   C

DOMAIN II
striWAP  KSCQCPAVTKV------CPRN-NWTHTCWLDDHGLGEKKCCF--SGYGRR-CME--PFRDKKELANLKCVSSFGRC
possWAP  KPGRCPAIAEV------CPKNMSWIDNCQSDDHCNRSKKCCS--SGCGQR-CMK--PLPEERGAVTRSFH------
tammWAP  KPGRCPAVTGI------CPEKKSWFHTCQRDDQCKENKKCCS--SACGRR-CTN--PFPEEYEASQDESTLLAL--
oposWAP  KPGRCPVVSEW------CPRNERFIG-CRGDDNCLGMKKCCF--SRCAMR-CMDAIPX------------------
CONSENSUSK G CP          C    C    C    CC   C   C
```

The five exons of the marsupial *WAP* genes contrast with the eutherian *WAP* genes, which have only four exons (Fig. 3a). Domains I and II of the WAP protein are encoded by exons 2 and 3 in eutherians, and exons 3 and 4 in marsupials. Exon 2 in marsupials codes for domain III (Fig. 3a), an extra domain described by studies using 3-dimensional modeling (Simpson et al., 2000). Domains I, II and III contain eight cysteine residues (represented by C in Fig. 3b) which constitute the 4-DSC domain, with the exception of domain II in the stripe-faced dunnart which has only six cysteines. The WAP motif (KXGXCP) is present at the 5′ end of marsupial domains I, II and III. However, the motif in domain II of the stripe-faced dunnart WAP is replaced by KXCXCP. In contrast, eutherians have the WAP motif only in domain II (Fig. 3b).

A BlastP search of the NCBI database with the stripe-faced dunnart predicted WAP protein revealed the best match to be with the eutherian milk WAPs from mouse, rat, camel, rabbit and pig. There is no functional human milk WAP protein as it has become a pseudogene in the human lineage (Clauss et al., 2002). A BlastP search of the human protein database (at NCBI) revealed the two best matches with the stripe-faced dunnart predicted WAP protein to be *WFDC3* and *WFDC5* which are both functional WAP genes located on human 20q13. The stripe-faced dunnart predicted WAP protein is compared to other marsupial milk WAPs in Fig. 3b. Homology is observed in the region of the WAP motif, the 4-DSC consensus sequences and the leader peptide sequences.

Chromosome mapping of marsupial ELP and WAP genes in marsupials

FISH onto metaphase spreads localized the *ELP* gene to chromosome 2q in the stripe-faced dunnart (Fig. 4b), and the *WAP* gene to chromosome 1p (Fig. 4a).

Milk protein gene expression

Studies were carried out to examine if there was asynchronous expression of the *ELP* and *WAP* genes during the lactation cycle in the stripe-faced dunnart. No lactating animals were available from a stripe-faced dunnart *(S. macroura)* colony. However, mammary glands were collected from lactating females of the closely related fat-tailed dunnart *(Sminthopsis crassicaudata)*.

Milk protein gene expression was analyzed by Northern blot analysis of RNA from mammary glands on day 37 of fat-tailed dunnart development, which is the early to middle stage of lactation where the young are permanently attached

Fig. 3. Structure of the *WAP* gene. (**a**) Comparative genomic organization of the *WAP* genes in eutherians and marsupials. The exons common to both marsupials and eutherians are shown in black, the extra exon in marsupials is shown by a grey background with black stripes. In eutherians exons 2 and 3 encode the two 4-DSC domains. In marsupials exons 2, 3 and 4 encode the three 4-DSC domains. Hence, domain III has been found in marsupials, but not in eutherians. The fifth marsupial exon has been identified in opossum, but the stripe-faced dunnart sequence does not extend far enough to reach the fifth exon. (**b**) Alignment of the stripe-faced dunnart WAP protein with other marsupial WAP proteins expressed in the mammary gland. In all marsupials there is homology between the protein domains as well as the leader sequence. The consensus sequence which includes the WAP motif (KXGXCP) and the four disulphide core domain represented by eight cysteines (C) is shown below each domain in bold. In the protein alignment the consensus sequence is shown in white type with a black background and other conserved proteins are shown in black type with a grey background. striWAP = stripe-faced dunnart WAP, conceptually translated from the BAC sequence; fattWAP = fat-tailed dunnart WAP conceptually translated from the partial cDNA sequence from domains III and I; possWAP = possum WAP gi 14582214; tammWAP = tammar WAP gi gi7711019, oposWAP = opossum WAP, assembled from the short sequences available at ww.ncbi.nlm.nih.gov/BLAST/tracemb.shtml.

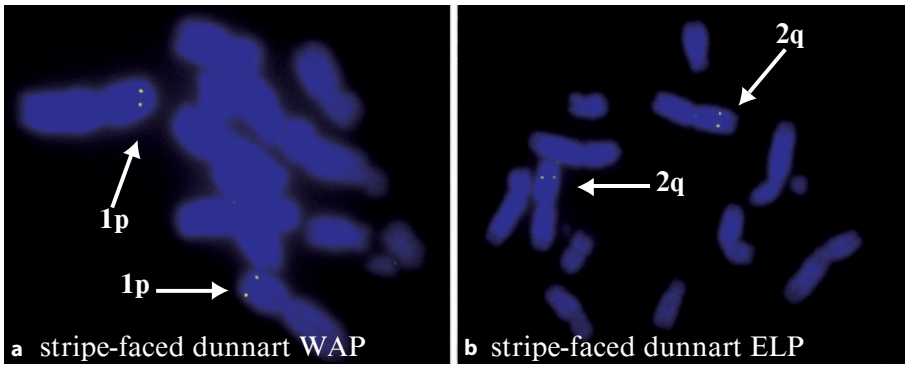


Fig. 4. Mapping of marsupial *WAP* and *ELP* genes on metaphase spreads. (**a**) *WAP* is located on chromosome 1p of the stripe-faced dunnart. (**b**) *ELP* is located on chromosome 2q of the stripe-faced dunnart.

to the teat, and on day 58 when the young are no longer permanently attached to the teat but are still suckling. Expression of the *ELP* gene was detected on day 37, and the *WAP* gene on day 58. In contrast, the *ALA* and *BLG* genes were expressed in lactating mammary gland on both day 37 and 58 (Fig. 5). The PCR products for *ALA, BLG, ELP* and *WAP* genes were sequenced for comparisons with other species.

Homology between ELP and WAP from the two dunnart species was greater than 97%, suggesting that the *ELP* and *WAP* genes in both dunnart species are homologous genes. Comparison of the *ELP* and *WAP* partial cDNA sequence from the fat-tailed dunnart with the *ELP* and *WAP* exon sequence predicted from the stripe-faced dunnart genomic sequence showed overall 1–3 base differences every 100 kb of sequence; seven differences in the 242 nucleotides sequenced from *WAP* (and two differences in the 80 amino acids of the predicted protein translation) and six differences in the 432 nucleotides sequenced from *ELP* (and two differences in the 96 amino acids of the predicted protein translation).

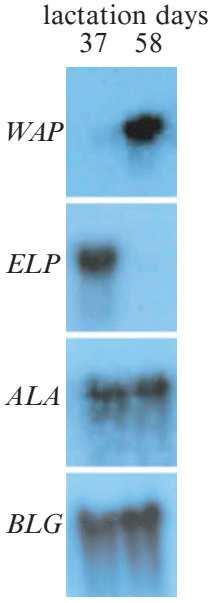


Fig. 5. Milk protein gene expression in the lactating mammary gland of the fat-tailed dunnart. The Northern blot shows that *ALA* and *BLG* are expressed on days 37 and 58. *WAP* is expressed on day 58 only and *ELP* is expressed on day 37 only. The *WAP, ALA* and *BLG* transcripts were detected as single bands at approximately 700 bp. The *ELP* transcript was detected as a single band at about 500 bp.

Discussion

Expression pattern of ELP, WAP, BLG and ALA genes during marsupial lactation cycles

The timing of gestation and lactation in the two dunnart species (the fat-tailed dunnart and the stripe-faced dunnart) is very similar. The stripe-faced dunnart has the shortest gestation (10.5–11 days) known for any mammal (Selwood and Woolley, 1991; Gemmell and Selwood, 1994), and a lactation length of 65–70 days (Frigo and Woolley, 1997). The newborn remains permanently attached to the teat for the first 30–40 days of lactation, and then on day 55 commences eating solid food while still intermittently suckling milk (Frigo and Woolley, 1997). The gestation time for the fat-tailed dunnart (13–16 days) is slightly longer and the duration of lactation is approximately 70 days. Permanent attachment to the teat ceases after 40 days, and after 60 days solid food is consumed by the young (Ewer, 1968; Godfrey and Crowcroft, 1971).

This study has shown that the fat-tailed dunnart *ELP* and *WAP* genes are expressed at different stages of lactation. *ELP* is expressed on day 37, when the young remain permanently attached to the teat and have a diet exclusively of milk, and *WAP* is expressed on day 58 when the young are no longer permanently attached to the teat and have a diet of both solids and milk. This expression pattern is similar to that in diprotodont marsupials (which include the tammar and possum) that diverged about 45 MYA (Woodburne et al., 2003). The tammar and brush-tailed possum *ELP* gene is expressed for 100 days after parturition, when the young remain permanently attached to the teat (Piotte and Grigor, 1996; Simpson et al., 1998b) (Fig. 6). In contrast, *WAP* is expressed in the tammar from day 100 to 200 when the young are in the pouch and intermittently suckling but are no longer permanently attached to the teat (Simpson et al., 2000) (Fig. 6). In possum, *WAP* is expressed during day 80–180, and then declines at permanent pouch exit (Demmer et al., 2001). In contrast to the tammar and fat-tailed dunnart, there is a switch period in the brush-tailed possum, during which *ELP* and *WAP* are both expressed in the mammary gland, then *WAP* continues to be expressed, for about an additional 60 days after solid food is consumed by

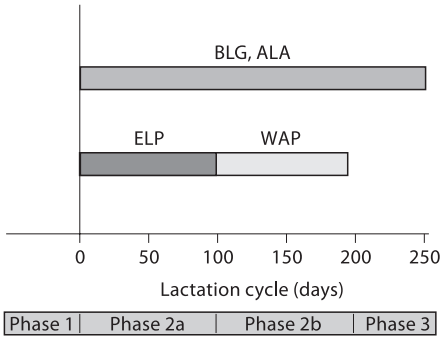


Fig. 6. The timing of expression of the ELP, WAP, BLG and ALA milk proteins during the lactation cycle of the tammar. BLG and ALA are expressed throughout lactation, but ELP and WAP are expressed at specific phases of lactation.

the young (Demmer et al., 2001). Due to the limited availability of tissue samples in the fat-tailed dunnart, the current study could not determine whether there is a similar switch period in this species.

It is interesting that *WAP* and *BLG* are concurrently expressed in the marsupials and monotremes studied to date. When *WAP* was first identified in the milk of mouse, rat (Hennighausen et al., 1982), rabbit (Beg et al., 1986) and camel (Devinoy et al., 1988), it was thought that animals secreted either *WAP* or *BLG* as their major whey protein, but not both. However, a more recent report showed the pig lactating mammary gland secretes *WAP* and *BLG* concurrently (Simpson et al., 1998a), suggesting that not all eutherians express only one of *WAP* or *BLG*. The current study suggests that *WAP* and *BLG* were expressed concurrently in the common mammalian ancestor, and the one or the other of *WAP* or *BLG* was lost subsequently in some eutherian lineages. In the human lineage both the 'milk *WAP*' gene and the *BLG* gene function have been lost and both are now pseudogenes (Clauss et al., 2002).

The evolution of the marsupial WAP and ELP genes

The sequence of the stripe-faced dunnart, tammar, and brush-tailed possum *ELP* and *WAP* genes showed more than 80% homology with the corresponding partial gene sequence from the fat-tailed dunnart. Sequence comparisons suggest that the gene structure, asynchronous expression and function of these genes is common at least to Australian marsupials. This study shows the genomic structure of the *ELP* and *WAP* genes in the gray short-tailed opossum is similar to that of Australian marsupials, therefore, it is likely that both genes are also present in the genome of other American marsupials. However, the expression pattern of these genes during lactation in the gray short-tailed opossum is yet to be studied.

The stripe-faced dunnart ELP protein contains a Kunitz motif with homology at the C terminal end (3′ end) of proteins to other Kunitz proteins. This suggests that the marsupial evolved additional sequences on the 5′ end of the *ELP* gene required for specific functions in marsupial lactation. However, a similar *ELP* gene in monotremes would suggest that the additional 5′ sequences were present in the common mammalian ancestor and lost in the eutherian lineage. Information on a potential platypus *ELP* should be available soon with the platypus genome nearly completely sequenced (http://www.ncbi.nlm.nih.gov/BLAST/tracemb.shtml). The Kunitz motif is common to one of the numerous families of serine protease inhibitors (NCBI protein database), including human TFPI2, trypstatin a mast cell inhibitor of trypsin in rats, isoinhibitor from the garden snail, a number of venom basic protease inhibitors, and basic protease inhibitor from sea turtle. Thus Kunitz motif genes are not exclusive to the mammary gland but are present in many genes expressed in different tissues with various functions.

The *WAP* genes expressed in the mammary gland of eutherians have two 4-DSC domains (Hennighausen et al., 1982; Simpson et al., 1998b), whereas WAP in all marsupials studied to date, as well as platypus and echidna, have only three (Simpson et al., 2000; Demmer et al., 2001). This extra domain is probably ancestral for all mammals, and was lost in the eutherian lineage. In eutherians, the identical distributions of cysteine residues within the two domains suggest that the *WAP* gene evolved by intergenic duplication (Hennighausen et al., 1982). The three domains of marsupials also have similar distributions of cysteine residues, suggesting they arose from two intergenic duplications.

Location of the stripe-faced dunnart WAP and ELP genes

The mouse milk *Wap* gene is located on chromosome 11, flanked by the genes *Ramp3* and *Tbrg4*. This region is equivalent to human chromosome 7p13→p12, where a non-functional human *WAP* relic is also flanked by *RAMP3* and *TBRG4* (Clauss et al., 2002; Rival-Gervier et al., 2003). Functional *WAP*-like genes in the human genome are clustered on chromosome 20q13, and are ubiquitously expressed, including non-lactating breast tissue (Clauss et al., 2002). However, *WAP* has not yet been identified in human milk, and the mouse milk *WAP* gene does not share orthology with the human *WAP* genes on chromosome 20 (Clauss et al., 2002). Therefore, the marsupial milk *WAP*s along with the mouse and other eutherian 'milk *WAP*s' are likely to share orthology with the human 'milk *WAP*' pseudogene. Hence, the finding of stripe-faced dunnart *WAP* on chromosome 1p suggests that stripe-faced dunnart 1p shares a common ancestry with human chromosome 7, where the human *WAP* pseudogene is located.

The finding that the stripe-faced dunnart *PIGT* gene is located next to the stripe-faced dunnart *ELP* gene, and that the sole human *PIGT* gene (and also the mouse *Pigt* gene) are near to two other genes (*WAP8* and *LOC391252*) that also share the Kunitz motif suggests that the putative therian common ancestor may have had a *PIGT* gene and a Kunitz protein motif gene neighboring each other on an ancestral chromosome. This shows that the chromosomal locations of this gene on stripe-faced dunnart chromosome 2q and the human chromosome 20q probably share common ancestry.

Cross-species painting between the stripe-faced dunnart, and gray short-tailed opossum (Rens et al., 2001) reveals which chromosome regions are shared between these species. The gene mapping from this study can therefore predict that in the gray short-tailed opossum genome the *ELP* gene should be found on chromosome 1q, and the *WAP* gene on chromosome 6q near the centromere. As the gray short-tailed opossum genome currently has only scaffold regions assembled (http://www.ensembl.org), which are not yet identified as belonging to specific chromosomes, the gene locations predicted in this study provide valuable reference points to anchor sequence to particular chromosomes in the opossum. The human/marsupial comparative map positions discovered in this milk protein gene study can now be used to advance the construction of a comparative map between human and marsupial genomes, which will be vital for exploring the evolution of the mammalian genome.

GenBank submissions

S. crassicaudata ALA partial cDNA AC DQ518240, *S. crassicaudata BLG* partial cDNA AC DQ518241, *S. crassicaudata ELP* partial cDNA AC DQ518242, *S. crassicaudata WAP* partial cDNA AC DQ5182403, *S. macroura* ELP BAC (237 L6) AC 186006, *S. macroura* WAP BAC (002F17) 14 kb AC DQ518244.

Acknowledgements

We thank Yi Peng and Shibani Mukherjee for performing the ELP BAC sequencing under the NHLBI's Program for Genomic Applications grant to J.F.C.

References

Bailey L, Lemon M: A specific protein difference in the milk from two mammary glands of a red kangaroo. Aust J Exp Biol Med Sci 44:705 (1966).

Ballard FJ, Grbovac S, Nicholas KR, et al: Differential changes in the milk concentrations of epidermal growth factor and insulin like growth factor-I during lactation in the tammar wallaby, *Macropus eugenii*. Gen Comp Endocrinol 98:262–268 (1995).

Beg OU, Von Bahr-Lindstrom H, Zaidi ZH, et al: A camel milk whey protein rich in half-cystine. Primary structure, assessment of variations, internal repeat patterns, and relationships with neurophysin and other active polypeptides. Eur J Biochem 159:195–201 (1986).

Chapman MA, Charchar FJ, Kinston S, et al: Comparative and functional analyses of *LYL1* loci establish marsupial sequences as a model for phylogenetic footprinting. Genomics 81:249–259 (2003).

Clauss A, Lija H, Lundwall A: A locus on human chromosome 20 contains several genes expressing protease inhibitor domains with homology to whey acidic protein. Biochem J 368:233–242 (2002).

De Leo AA, Guedelha N, Toder R, et al: Comparative chromosome painting between marsupial orders: relationships with a 2n = 14 ancestral marsupial karyotype. Chromosome Res 7:509–517 (1999).

Demmer J, Stasiuk SJ, Grigor MR, et al: Differential expression of the whey acidic protein gene during lactation in the brushtail possum *(Trichosurus vulpecula)*. Biochim Biophys Acta 1522: 187–194 (2001).

Devinoy E, Hubert C, Jolivet G, et al: Recent data on the structure of rabbit milk protein genes and on the mechanism of the hormonal control of their expression. Reprod Nutr Dev 28:1145–1164 (1988).

Ewer RF: A preliminary survey of the behaviour in captivity of the dasyurid marsupial, *Sminthopsis crassicaudata* (Gould). Z Tierpsychol 25: 319–365 (1968).

Ewing B, Green P: Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–194 (1998).

Frigo L, Woolley PA: Growth and development of pouch young of the stripefaced Dunnart, *Sminthopisis macroura* (Marsupialia:Dasyuridae), in captivity. Austral J Zool 45:157 (1997).

Gemmell RT, Selwood L: Structural development in the newborn marsupial, the stripe-faced dunnart, *Sminthopsis macroura*. Acta Anat 149:1–12 (1994).

Godfrey GK, Crowcroft P: Breeding the fat-tailed marsupial mouse, *Sminthopsis crassicaudata*, in captivity. Inter Zoo Yearbook 11:34 (1971).

Green B: Composition of milk and energetics of growth in marsupials. Zool Soc Lond Symp 51: 369 (1984).

Griffiths M, McIntosh DL, Leckie RMC: The mammary gland of the red kangaroo with observations on the fatty acid components of the milk triglycerides. J Zool Lond 166:265 (1972).

Hendry KA, Simpson KJ, Nicholas KR, et al: Autocrine inhibition of milk secretion in the lactating tammar wallaby *(Macropus eugenii)*. J Mol Endocrinol 21:169–177 (1998).

Hennighausen LG, Sippel AE, Hobbs AA, et al: Comparative sequence analysis of the mRNAs coding for mouse and rat whey protein. Nucleic Acids Res 10:3733–3744 (1982).

Jenness R: Lactational performance of various mammalian species. J Dairy Sci 69:869–885 (1986).

Nicholas KR: Asynchronous dual lactation in a marsupial, the tammar wallaby *(Macropus eugenii)*. Biochem Biophys Res Commun 154: 529–536 (1988).

Nicholas K, Simpson K, Wilson M, et al: The tammar wallaby: a model to study putative autocrine-induced changes in milk composition. J Mammary Gland Biol Neoplasia 2:299–310 (1997).

Nicholas KR, Fisher JA, Muths E, et al: Secretion of whey acidic protein and cystatin is down regulated at mid-lactation in the red kangaroo. Comp Biochem Physiol 129:851–858 (2001).

Piotte CP, Grigor MR: A novel marsupial protein expressed by the mammary gland only during the early lactation and related to the Kunitz proteinase inhibitors. Arch Biochem Biophys 330:59–64 (1996).

Ranganathan S, Simpson KJ, Shaw DC, et al: The whey acidic protein family: a new signature motif and three-dimensional structure by comparative modeling. J Mol Graph Model 17:106–113 (1999).

Ren J, Stuart DI, Acharya KR: Alpha-lactalbumin possesses a distinct zinc binding site. J Biol Chem 268:19292–19298 (1993).

Rens W, O'Brien PC, Yang F, et al: Karyotype relationships between distantly related marsupials from South America and Australia. Chromosome Res 9:301–308 (2001).

Rival-Gervier S, Thepot D, Jolivet G, et al: Pig whey acidic protein gene is surrounded by two ubiquitously expressed genes. Biochim Biophys Acta 1627:7–14 (2003).

Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning, a Laboratory Manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

Selwood L, Woolley PA: A timetable of embryonic development, and ovarian and uterine changes during pregnancy, in the stripe-faced dunnart, *Sminthopsis macroura* (Marsupialia: Dasyuridae). J Reprod Fertil 91:213–227 (1991).

Simpson KJ, Bird P, Shaw D, et al: Molecular characterisation and hormone dependent expression of the porcine whey acidic protein gene. J Mol Endocrinol 20:27–35 (1998a).

Simpson K, Shaw D, Nicholas KR: Developmentally regulated expression of a putative protease inhibitor gene in the lactating mammary gland of the tammar wallaby, *Macropus eugenii*. Comp Biochem Physiol 120:535–541 (1998b).

Simpson KJ, Ranganathan S, Fisher JA, et al: The gene for a novel member of the whey acidic protein family encodes three four-disulfide core domains and is asynchronously expressed during lactation. J Biol Chem 275:23074–23081 (2000).

Toder R, Wilcox SA, Smithwick M, et al: The human/mouse imprinted genes *IGF2, H19, SNRPN* and *ZNF127* map to two conserved autosomal clusters in a marsupial. Chromosome Res 4:295–300 (1996).

Triplett AA, Sakamoto K, Matulka LA, et al: Expression of the whey acidic protein (Wap) is necessary for adequate nourishment of the offspring but not functional differentiation of mammary epithelial cells. Genesis 43:1–11 (2005).

Trott JF, Simpson KJ, Moyle RL, et al: Maternal regulation of milk composition, milk production, and pouch young development during lactation in the tammar wallaby *(Macropus eugenii)*. Biol Reprod 68:929–936 (2003).

Tyndale-Biscoe CH, Janssens PA: Introduction, in Tyndale-Biscoe CH, Janssens PA (eds): The Developing Marsupial: Models for Biomedical Research, p 1 (Springer-Verlag, Heidelberg 1988).

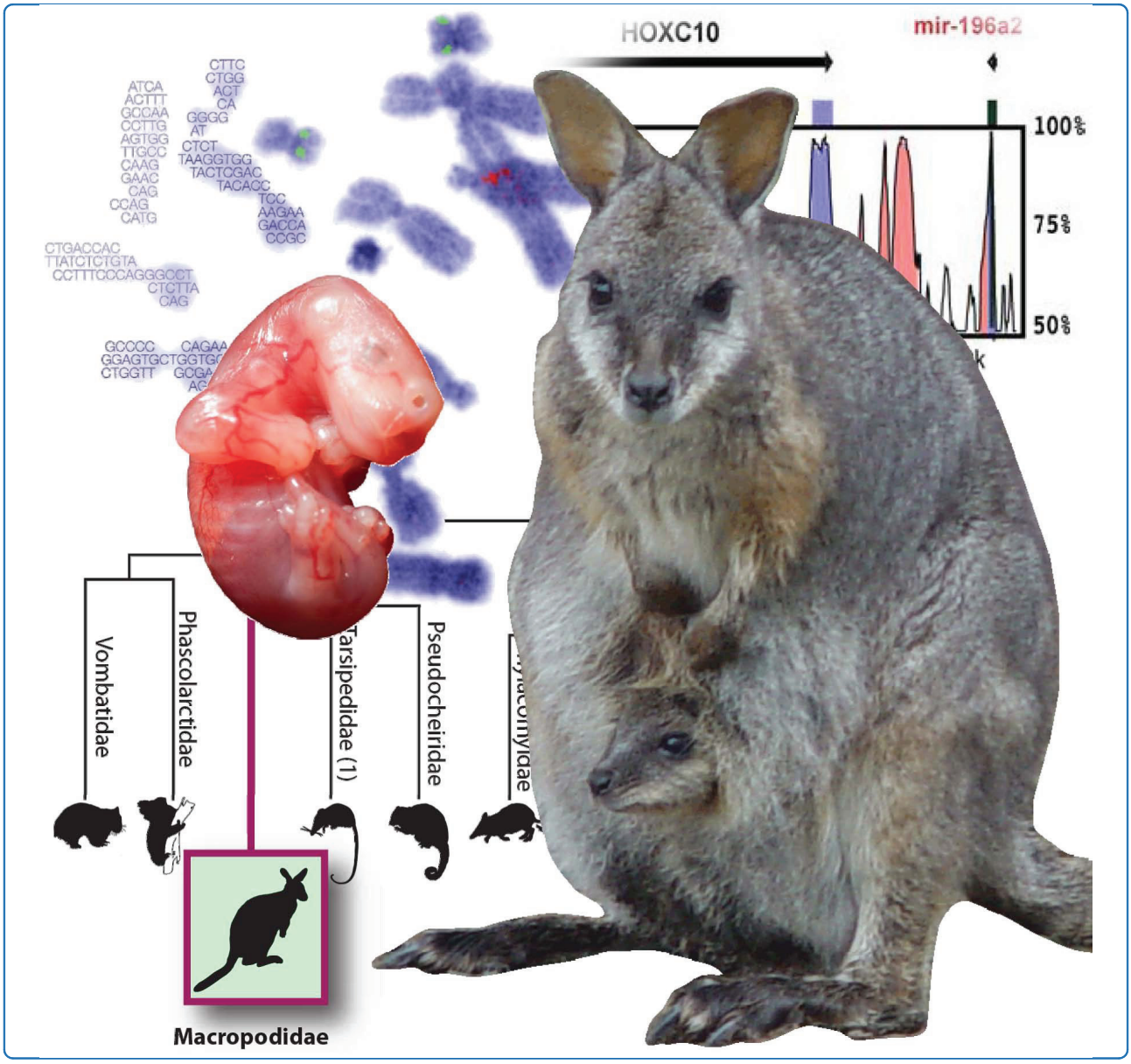
Woodburne MO, Rich TH, Springer MS: The evolution of tribospheny and the antiquity of mammalian clades. Mol Phylogenet Evol 28:360–385 (2003).

Genome Biology
Biology for the post-genomic era
2011 Volume 12 Issue 8
URL http://genomebiology.com/2011/12/8/
YEAR VOLUME ISSUE



BioMed Central
The Open Access Publisher

Genome Biology



Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development

Renfree *et al.*

BioMed Central

Renfree *et al. Genome Biology* 2011, **12**:R81
http://genomebiology.com/2011/12/8/R81 (29 August 2011)

Genome Biology

RESEARCH Open Access

Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development

Marilyn B Renfree[1,2*†], Anthony T Papenfuss[1,3,4*†], Janine E Deakin[1,5], James Lindsay[6], Thomas Heider[6], Katherine Belov[1,7], Willem Rens[8], Paul D Waters[1,5], Elizabeth A Pharo[2], Geoff Shaw[1,2], Emily SW Wong[1,7], Christophe M Lefèvre[9], Kevin R Nicholas[9], Yoko Kuroki[10], Matthew J Wakefield[1,3], Kyall R Zenger[1,7,11], Chenwei Wang[1,7], Malcolm Ferguson-Smith[8], Frank W Nicholas[7], Danielle Hickford[1,2], Hongshi Yu[1,2], Kirsty R Short[12], Hannah V Siddle[1,7], Stephen R Frankenberg[1,2], Keng Yih Chew[1,2], Brandon R Menzies[1,2,13], Jessica M Stringer[1,2], Shunsuke Suzuki[1,2], Timothy A Hore[1,14], Margaret L Delbridge[1,5], Amir Mohammadi[1,5], Nanette Y Schneider[1,2,15], Yanqiu Hu[1,2], William O'Hara[6], Shafagh Al Nadaf[1,5], Chen Wu[7], Zhi-Ping Feng[3,16], Benjamin G Cocks[17], Jianghui Wang[17], Paul Flicek[18], Stephen MJ Searle[19], Susan Fairley[19], Kathryn Beal[18], Javier Herrero[18], Dawn M Carone[6,20], Yutaka Suzuki[21], Sumio Sugano[21], Atsushi Toyoda[22], Yoshiyuki Sakaki[10], Shinji Kondo[10], Yuichiro Nishida[10], Shoji Tatsumoto[10], Ion Mandiou[23], Arthur Hsu[3,16], Kaighin A McColl[3], Benjamin Lansdell[3], George Weinstock[24], Elizabeth Kuczek[1,25,26], Annette McGrath[25], Peter Wilson[25], Artem Men[25], Mehlika Hazar-Rethinam[25], Allison Hall[25], John Davis[25], David Wood[25], Sarah Williams[25], Yogi Sundaravadanam[25], Donna M Muzny[24], Shalini N Jhangiani[24], Lora R Lewis[24], Margaret B Morgan[24], Geoffrey O Okwuonu[24], San Juana Ruiz[24], Jireh Santibanez[24], Lynne Nazareth[24], Andrew Cree[24], Gerald Fowler[24], Christie L Kovar[24], Huyen H Dinh[24], Vandita Joshi[24], Chyn Jing[24], Fremiet Lara[24], Rebecca Thornton[24], Lei Chen[24], Jixin Deng[24], Yue Liu[24], Joshua Y Shen[24], Xing-Zhi Song[24], Janette Edson[25], Carmen Troon[25], Daniel Thomas[25], Amber Stephens[25], Lankesha Yapa[25], Tanya Levchenko[25], Richard A Gibbs[24], Desmond W Cooper[1,28], Terence P Speed[1,3], Asao Fujiyama[22,27], Jennifer A M Graves[1,5], Rachel J O'Neill[6], Andrew J Pask[1,2,6], Susan M Forrest[1,25] and Kim C Worley[24]

Abstract

Background: We present the genome sequence of the tammar wallaby, *Macropus eugenii*, which is a member of the kangaroo family and the first representative of the iconic hopping mammals that symbolize Australia to be sequenced. The tammar has many unusual biological characteristics, including the longest period of embryonic diapause of any mammal, extremely synchronized seasonal breeding and prolonged and sophisticated lactation within a well-defined pouch. Like other marsupials, it gives birth to highly altricial young, and has a small number of very large chromosomes, making it a valuable model for genomics, reproduction and development.

Results: The genome has been sequenced to 2 × coverage using Sanger sequencing, enhanced with additional next generation sequencing and the integration of extensive physical and linkage maps to build the genome assembly. We also sequenced the tammar transcriptome across many tissues and developmental time points.

* Correspondence: m.renfree@unimelb.edu.au; papenfuss@wehi.edu.au
† Contributed equally
[1]The Australian Research Council Centre of Excellence in Kangaroo Genomics, Australia
Full list of author information is available at the end of the article

BioMed Central

© 2011 Renfree et al.; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Our analyses of these data shed light on mammalian reproduction, development and genome evolution: there is innovation in reproductive and lactational genes, rapid evolution of germ cell genes, and incomplete, locus-specific X inactivation. We also observe novel retrotransposons and a highly rearranged major histocompatibility complex, with many class I genes located outside the complex. Novel microRNAs in the tammar HOX clusters uncover new potential mammalian HOX regulatory elements.

Conclusions: Analyses of these resources enhance our understanding of marsupial gene evolution, identify marsupial-specific conserved non-coding elements and critical genes across a range of biological systems, including reproduction, development and immunity, and provide new insight into marsupial and mammalian biology and genome evolution.

Background

The tammar wallaby holds a unique place in the natural history of Australia, for it was the first Australian marsupial discovered, and the first in which its special mode of reproduction was noted: '*their manner of procreation is exceeding strange and highly worth observing; below the belly the female carries a pouch into which you may put your hand; inside the pouch are her nipples, and we have found that the young ones grow up in this pouch with the nipples in their mouths. We have seen some young ones lying there, which were only the size of a bean, though at the same time perfectly proportioned so that it seems certain that they grow there out of the nipples of the mammae from which they draw their food, until they are grown up*' [1]. These observations were made by Francisco Pelseart, Captain of the ill-fated and mutinous Dutch East Indies ship Batavia in 1629, whilst shipwrecked on the Abrolhos Islands off the coast of Geraldton in Western Australia. It is therefore appropriate that the tammar should be the first Australian marsupial subject to an in-depth genome analysis.

Marsupials are distantly related to eutherian mammals, having shared a common ancestor between 130 and 148 million years ago [2-4]. The tammar wallaby *Macropus eugenii* is a small member of the kangaroo family, the Macropodidae, within the genus *Macropus*, which comprises 14 species [5] (Figure 1). The macropodids are the most specialized of all marsupials. Mature females weigh about 5 to 6 kg, and males up to 9 kg. The tammar is highly abundant in its habitat on Kangaroo Island in South Australia, and is also found on the Abrolhos Islands, Garden Island and the Recherche Archipelago, all in Western Australia, as well as a few small areas in the south-west corner of the continental mainland. These populations have been separated for at least 40,000 years. Its size, availability and ease of handling have made it the most intensively studied model marsupial for a wide variety of genetic, developmental, reproductive, physiological, biochemical, neurobiological and ecological studies [6-13].

In the wild, female Kangaroo Island tammars have a highly synchronized breeding cycle and deliver a single young on or about 22 January (one gestation period after the longest day in the Southern hemisphere, 21 to 22 December) that remains in the pouch for 9 to 10 months. The mother mates within a few hours after birth but development of the resulting embryo is delayed during an 11 month period of suspended animation (embryonic diapause). Initially diapause is maintained by a lactation-mediated inhibition, and in the second half of the year by photoperiod-mediated inhibition that is removed as day length decreases [14]. The anatomy, physiology, embryology, endocrinology and genetics of the tammar have been described in detail throughout development [6,11-13,15].

The marsupial mode of reproduction exemplified by the tammar with a short gestation and a long lactation does not imply inferiority, nor does it represent a transitory evolutionary stage, as was originally thought. It is a successful and adaptable lifestyle. The maternal investment is minimal during the relatively brief pregnancy and in early lactation, allowing the mother to respond to altered environmental conditions [11,12,15]. The tammar, like all marsupials, has a fully functional placenta that makes hormones to modulate pregnancy and parturition, control the growth of the young, and provide signals for the maternal recognition of pregnancy [14,16-18]. The tammar embryo develops for only 26 days after diapause, and is born when only 16 to 17 mm long and weighing about 440 mg at a developmental stage roughly equivalent to a 40-day human or 15-day mouse embryo. The kidney bean-sized newborn has well-developed forelimbs that allow it to climb up to the mother's pouch, where it attaches to one of four available teats. It has functional, though not fully developed, olfactory, respiratory, circulatory and digestive systems, but it is born with an embryonic kidney and undifferentiated immune, thermoregulatory and reproductive systems, all of which become functionally differentiated during the lengthy pouch life. Most major structures and organs, including the hindlimbs, eyes, gonads and a significant portion of the brain, differentiate while the young is in the pouch and are therefore readily available for study [11,12,19-24]. They also have a sophisticated lactational

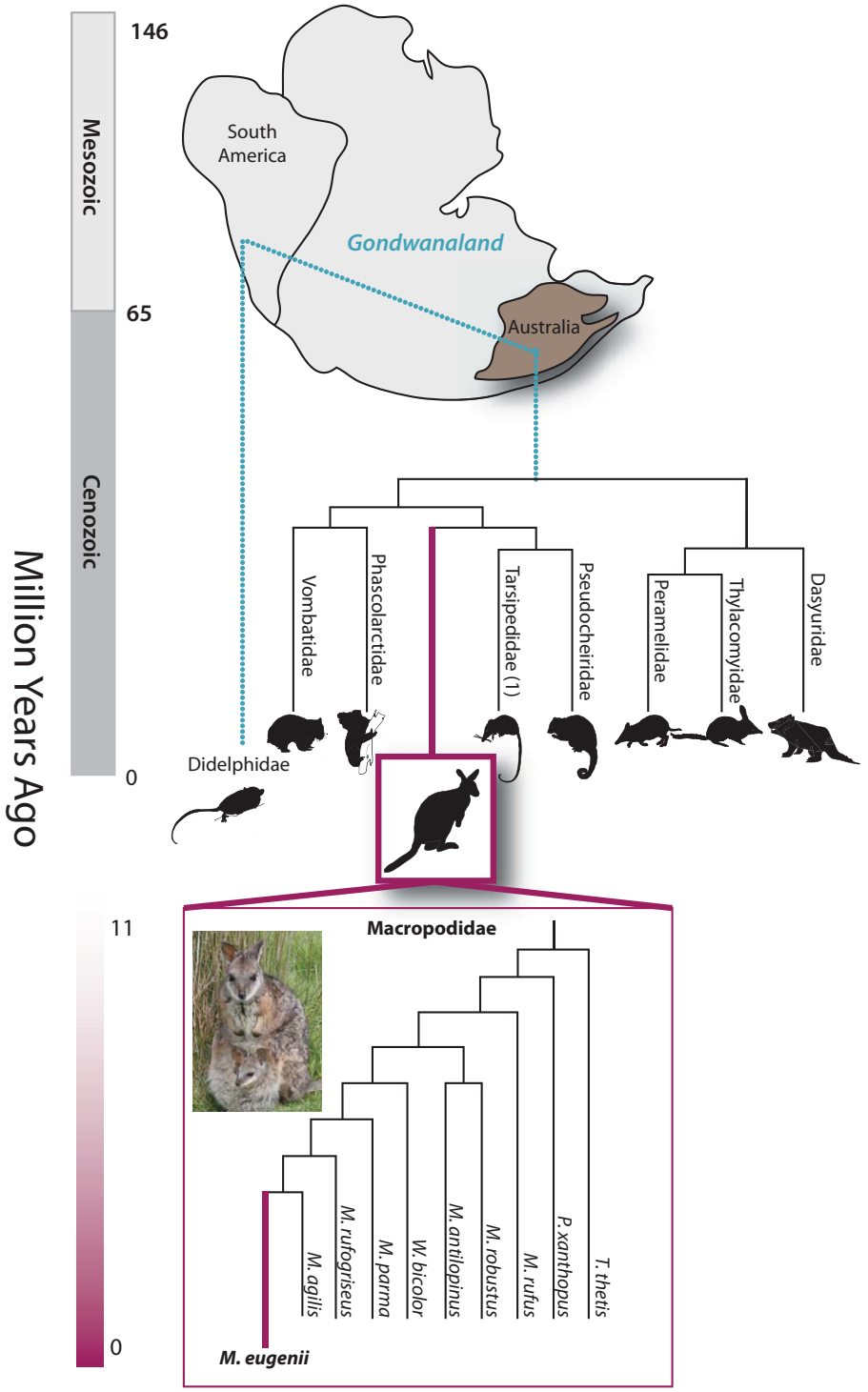


Figure 1 Phylogeny of the marsupials. Phylogenetic relationships of the orders of Marsupialia. Top: the placement of the contemporary continents of South America and Australia within Gondwanaland and the split of the American and Australian marsupials. Relative divergence in millions of years shown to the left in the context of geological periods. The relationship of the Macropodide within the Australian marsupial phylogeny shown is in purple with estimated divergence dates in millions of years [5,162,163]. Representative species from each clade are illustrated. Inset: phylogeny of the genus *Macropus* within the Macropodidae showing the placement of the model species *M. eugenii* (purple) based on [59]. Outgroup species are *Thylogale thetis* and *Petrogale xanthopus*.

physiology with a milk composition that changes throughout pouch life, ensuring that nutrient supply is perfectly matched for each stage of development [25]. Adjacent teats in a pouch can deliver milk of differing composition appropriate for a pouch young and a young-at-foot [26].

Kangaroo chromosomes excited some of the earliest comparative cytological studies of mammals. Like other kangaroos, the tammar has a low diploid number (2n = 16) and very large chromosomes that are easily distinguished by size and morphology. The low diploid number of marsupials makes it easy to study mitosis, cell cycles [27], DNA replication [28], radiation sensitivity [29], genome stability [30], chromosome elimination [31,32] and chromosome evolution [33,34]. Marsupial sex chromosomes are particularly informative. The X and Y chromosomes are small; the basic X chromosome constitutes only 3% of the haploid genome (compared with 5% in eutherians) and the Y is tiny. Comparative studies show that the marsupial X and Y are representative of the ancestral mammalian X and Y chromosomes [35]. However, in the kangaroos, a large heterochromatic nucleolus organizer region became fused to the X and Y. Chromosome painting confirms the extreme conservation of kangaroo chromosomes [36] and their close relationship with karyotypes of more distantly related marsupials [37-40] so that genome studies are likely to be highly transferable across marsupial species.

The tammar is a member of the Australian marsupial clade and, as a macropodid marsupial, is maximally divergent from the only other sequenced model marsupial, the didelphid Brazilian grey short-tailed opossum, *Monodelphis domestica* [41]. The South American and Australasian marsupials followed independent evolutionary pathways after the separation of Gondwana into the new continents of South America and Australia about 80 million years ago and after the divergence of tammar and opossum (Figure 1) [2,4]. The Australasian marsupials have many unique specializations. Detailed knowledge of the biology of the tammar has informed our interpretation of its genome and highlighted many novel aspects of marsupial evolution.

Sequencing and assembly (Meug_1)

The genome of a female tammar of Kangaroo Island, South Australia origin was sequenced using the whole-genome shotgun (WGS) approach and Sanger sequencing. DNA isolated from the lung tissue of a single tammar was used to generate WGS libraries with inserts of 2 to 6 kb (Tables S1 and S2 in Additional file 1). Sanger DNA sequencing was performed at the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC), and the Australian Genome Research Facility using ABI3730xl sequencers (Applied BioSystems, Foster City, CA, USA). Approximately 10 million Sanger WGS reads, representing about 2 × sequence coverage, were submitted to the NCBI trace archives (NCBI BioProject PRJNA12586; NCBI Taxonomy ID 9315). An additional 5.9 × sequence coverage was generated on an ABI SOLiD sequencer at BCM-HGSC. These 25-bp paired-end data with average mate-pair distance of 1.4 kb (Table S3 in Additional file 1) [SRA:SRX011374] were used to correct contigs and perform super-scaffolding. The initial tammar genome assembly (Meug_1.0) was constructed using only the low coverage Sanger sequences. This was then improved with additional scaffolding using sequences generated with the ABI SOLiD (Meug_1.1; Table 1; Tables S4 to S7 in Additional file 1). The Meug_1.1 assembly had a contig N50 of 2.6 kb and a scaffold N50 of 41.8 kb [GenBank:GL044074-GL172636].

The completeness of the assembly was assessed by comparison to the available cDNA data. Using 758,062 454 FLX cDNA sequences [SRA:SRX019249, SRA:SRX019250], 76% are found to some extent in the assembly and 30% are found with more than 80% of their length represented (Table S6 in Additional file 1). Compared to 14,878 Sanger-sequenced ESTs [GenBank:EX195538-EX203564, GenBank:EX203644-EX210452], more than 85% are found in the assembly with at least one half their length aligned (Table S7 in Additional file 1).

Table 1 Comparison of Meug genome assemblies

	Assembly version		
	1.0	**1.1**	**2.0**
Contigs (million)	1.211	1.174	1.111
N50 (kb)	2.5	2.6	2.91
Bases (Mb)	2546	2,536	2,574
Scaffolds	616,418	277,711	379,858
Max scaffold size	NA	472,108	324,751
Gaps (Mb)	NA	539	619
N50 (kb)	NA	41.8	34.3
Complex scaffolds	NA	128,563	124,674
Singleton scaffolds	NA	149,148	255,184
Co-linear with BACs	NA	87.2% (418)	93.4% (298)
Co-linear with ESTs	NA	82.3% (704)	86.7% (454)

Summary statistics for the tammar genome assemblies. These statistics indicate the extension and merging of contigs done to improve the assembly. The larger number of scaffolds and smaller scaffold N50 is a consequence of higher stringency in the 2.0 scaffolding workflow. The higher stringency isolated many contigs. However, the number of complex (that is, useful) scaffolds is similar between the assemblies. For co-linear estimates, the scaffolds were linearized and BACs and cDNA libraries were mapped against them. The 1.1 and 2.0 assemblies were validated against 169 BAC contigs and 84,718 ESTs (that were not incorporated into either genome assembly). We determined the percentage of contigs where the scaffolding matched the order and orientation when compared to BACs or ESTs (co-linear with BACs/ESTs). Parentheses indicate the total number of contigs identified after alignment to BAC contigs or ESTs.

Additional sequencing and assembly improvement (Meug_2)

Contig improvement

The tammar genome assembly was further improved using additional data consisting of 0.3 × coverage by paired and unpaired 454 GS-FLX Titanium reads [SRA:SRX080604, SRA:SRX085177] and 5 × coverage by paired Illumina GAIIx reads [SRA:SRX085178, SRA:SRX081248] (Table S8 in Additional file 1). A local reassembly strategy mapped the additional 454 and Illumina data against Meug_1.1 contigs. Added data were used to improve the accuracy of base calls and to extend and merge contigs. The Meug_2.0 assembly [GenBank:ABQO000000000] (see also 'Data availability' section) has 1.111 million contigs with an N50 of 2.9 kb. Contigs were validated directly by PCR on ten randomly selected contigs. The assembly was also assessed by aligning 84,718 ESTs and 169 BAC sequences to the genome. The amount of sequence aligning correctly to the genome assembly showed modest improvement between Meug_1.1 and Meug_2.0 (Table 1; Table S9 in Additional file 1).

Scaffolding and anchoring using the virtual map

Scaffolds were constructed using the previously mentioned Illumina paired-end libraries with insert sizes of 3.1 kb (8,301,018 reads) and 7.1 kb (12,203,204 reads), 454 paired-end library with an insert size of 6 kb and SOLiD mate pair library. The mean insertion distances for each library were empirically determined using paired reads where both ends mapped within the same contig and only those within three standard deviations from the mean were used for scaffolding. The contigs were ordered and oriented using Bambus [42], through three iterations of scaffolding to maximize the accuracy of the assembly. The highest priority was given to the library with the smallest standard deviation in the paired end distances, and the remaining libraries arranged in descending order. Initial scaffolding by Bambus was performed using five links as a threshold [43]. Overlapping contigs were identified and set aside before reiteration. This step was performed twice and the overlapping contigs pooled. The non-overlapping and overlapping contigs were then scaffolded independently. Any scaffolds found to still contain overlap were split apart. The resulting assembly has 324,751 scaffolds with an N50 of 34,279 bp (Table 1). Scaffolds were assigned to chromosomes by aligning them to markers from the virtual map [44], represented using sequences obtained from the opossum and human genomes [45]. We assigned 6,979 non-overlapping scaffolds (163 Mb or 6% of the genome assembly) to the seven autosomes. The vast majority of the genome sequence remained unmapped.

Tammar genome size

The tammar genome size was estimated using three independent methods: direct assessment by quantitative PCR [46]; bivariate flow karyotyping and standard flow cytometry; and genome analyses based in the Sanger WGS reads, using the Atlas-Genometer [47]. These three approaches produced quite different genome size estimates (Tables S11 to S13 in Additional file 1) so the average size estimate, 2.9 Gb, was used for the purposes of constructing the Meug_2.0 integrated genome assembly. The smaller genome size of tammar compared to human is unlikely to be due to fewer genes or changes in gene size (Figure S1 in Additional file 2), but may be accounted for by the greatly reduced centromere size of 450 kb/chromosome and number (n = 8) [48] compared to the human centromere size of 4 to 10 Mb/chromosome (n = 23).

Physical and linkage mapping

Novel strategies were developed for the construction of physical and linkage maps covering the entire genome. The physical map consists of 520 loci mapped by fluorescence *in situ* hybridization (FISH) and was constructed by mapping the ends of gene blocks conserved between human and opossum, thereby allowing the location of genes within these conserved blocks to be extrapolated from the opossum genome onto tammar chromosomes [37] (JE Deakin, ML Delbridge, E Koina, N Harley, DA McMillan, AE Alsop, C Wang, VS Patel, and JAM Graves, unpublished results). Three different approaches were used to generate a linkage map consisting of 148 loci spanning 1,402.4 cM or 82.6% of the genome [49]. These approaches made the most of the available tammar sequence (genome, BACs or BAC ends) to identify markers to increase coverage in specific regions of the genome. Many of these markers were also physically mapped, providing anchors for the creation of an integrated map comprising all 553 distinct loci included in the physical and/or linkage maps. Interpolation of segments of conserved synteny (mainly from the opossum assembly) into the integrated map then made it possible to predict the genomic content and organization of the tammar genome through the construction of a virtual genome map comprising 14,336 markers [44].

Mapping data were used to construct tammar-human (Figure 2) and tammar-opossum comparative maps in order to study genome evolution. Regions of the genome were identified that have undergone extensive rearrangement when comparisons between tammar and opossum are made. These are in addition to previously known rearrangements based on chromosome-specific paints [50]. For example, tammar chromosome 3, consisting of genes that are on nine human chromosomes (3, 5, 7, 9,

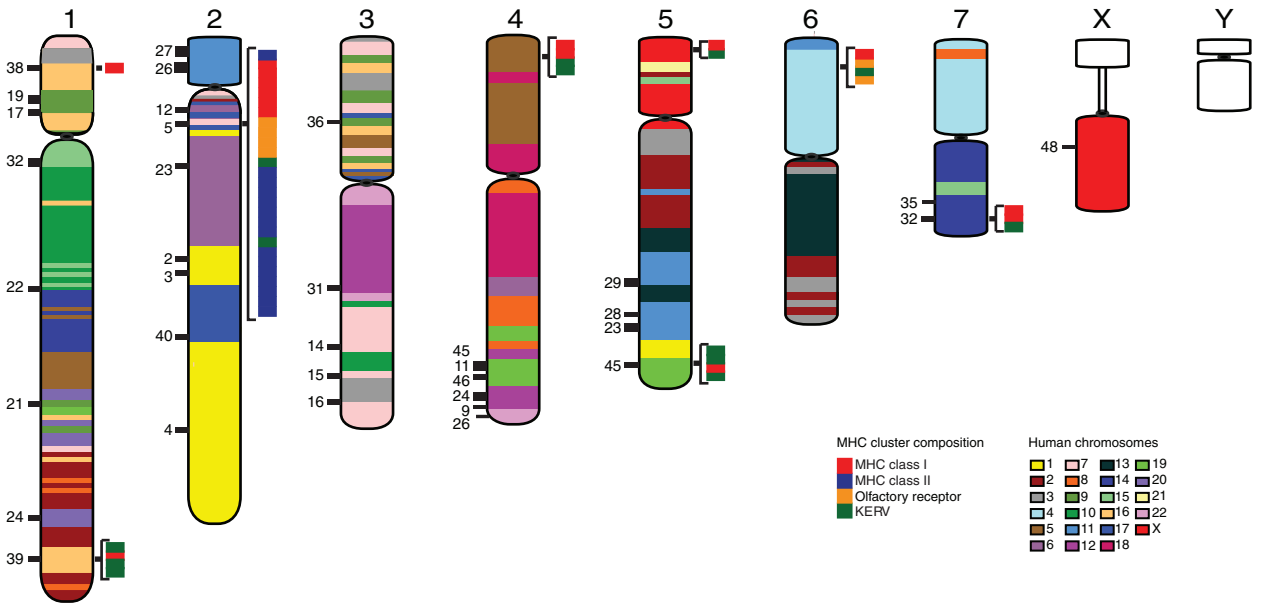


Figure 2 Homology of tammar regions to the human karyotype, and location of major histocompatibility complex, classical class I genes and olfactory receptor gene. Colored blocks represent the syntenic blocks with human chromosomes as shown in the key. A map of the locations of the tammar major histocompatibility complex (MHC) is shown on the right-hand side of each chromosome. The rearranged MHCs are on chromosome 2 and clusters of MHC class I genes (red) near the telomeric regions of chromosomes 1, 4, 5, 6, and 7. MHC class II genes are shown in blue, olfactory receptors are shown in orange and Kangaroo endogenous retroviral elements found within these clusters are shown in green. The location of the conserved mammalian OR gene clusters in the tammar genome are shown on the left-hand side of each chromosome. OR genes are found on every chromosome, except for chromosome 6 but including the X. The location of the OR gene clusters (numbers) are shown, and their approximate size is represented by lines of different thickness.

10, 12, 16, 17, 22; Figure 2) and the X have an extensive reshuffling of the gene order. Rearrangements on the remaining chromosomes are mostly the result of large-scale inversions. This enabled us to predict the ancestral marsupial karyotype, revealing that inversions and micro-inversions have played a major role in shaping the genomes of marsupials (JE Deakin, ML Delbridge, E Koina, N Harley, DA McMillan, AE Alsop, C Wang, VS Patel, and JAM Graves, unpublished results).

Genome annotation

The Ensembl genebuild (release 63) for the Meug_1.0 assembly identified 18,258 genes by projection from high quality reference genomes. Of these, 15,290 are protein coding, 1,496 are predicted pseudo-genes, 525 are micro-RNA (miRNA) genes, and 42 are long non-coding RNA genes, though these are composed of just 7 different families: 7SK, human accelerated region 1F, CPEB3 ribozyme, ncRNA repressor of NFAT, nuclear RNase P, RNase MRP and Y RNA.

Since the coverage is low, many genes may be fragmented in the assembly or even unsequenced. The Ensembl genebuild pipeline scaffolds fragmented genes using comparative data and constructs 'GeneScaffolds'. There are 10,257 GeneScaffolds containing 13,037 genes. The annotation also contains 9,454 genes interrupted by Ns. To partially ameliorate the problems of missing genes, a number of BACs from targeted locations have been sequenced and annotated, including the HOX gene clusters (H Yu, Z-P Feng, RJ O'Neill, Y Hu, AJ Pask, D Carone, J Lindsay, G Shaw, AT Papenfuss, and MB Renfree, unpublished results), major histocompatibility complex (MHC) [51], X chromosome (ML Delbridge, B Landsdell, MT Ross, TP Speed, AT Papenfuss, JAM Graves, unpublished results), pluripotency genes, germ cell genes, spermatogenesis genes [52,53] and X chromosome genes. Findings from these are summarized in later sections of this paper.

Expansion of gene families

Many genes evolve and acquire novel function through duplication and divergence. We identified genes that have undergone expansions in the marsupial lineage but remain largely unduplicated in eutherians and reptiles (Table S15 in Additional file 1). Both the tammar and opossum have undergone expansion of MHC class II genes, critical in the immune recognition of extracellular pathogens, and *TAP* genes that are responsible for loading endogenously derived antigens onto MHC class I proteins. Three marsupial-specific class II gene families exist: DA, DB and DC. Class II genes have undergone further duplications in the tammar and form two genomic clusters, adjacent to the antigen-processing genes

[51]. The opossum has one *TAP1* and two *TAP2* genes, while the tammar has expanded *TAP1* (two genes) and *TAP2* (three genes) genes [51]. We also detected marsupial expansions linked to apoptosis (*NET1*, *CASP3*, *TMBIM6*) and sensory perception (olfactory receptors).

Genomic landscape

Sequence conservation

We next explored sequence conservation between tammar and opossum using sequence similarity as a sensitive model of conservation. We found that 38% of nucleotides in the tammar genome (Meug_1.0) could be aligned to the high-quality opossum genome (7.3×). Of the aligned sequence, 72% was unannotated, reflecting a high proportion of conserved non-coding regions between the marsupial species. The level of conservation between opossum and tammar varied from 36.0 to 40.9% across the different opossum chromosomes (Table S16 in Additional file 1). This variation seems modest and may be largely stochastic, but it is interesting to examine further. Opossum chromosome 1 has 40.6% sequence conservation with the tammar. The gene order between tammar and opossum chromosome 1 is also highly conserved. This may mean that within the tammar genome assembly scaffolds, the alignment is well anchored by conserved protein-coding genes, making the intergenic sequence easier to align. Thus this 'high' conservation may be largely due to inherent biases in the approach. Opossum chromosome X has the most conserved sequence compared to tammar (40.9%), despite the high level of rearrangement between the tammar and opossum X. Intriguingly, the proportion of conserved sequence on opossum chromosome X that is located in unannotated regions is also the highest of any chromosome (28.2%; Table S16 in Additional file 1) despite the level of rearrangement. This may indicate a significant number of non-coding regulatory elements on the X chromosome. The mechanism of X inactivation in marsupials is not well understood. Examination of transcription within individual nuclei shows that there is at least regional coordinated expression of genes on the partially inactive X [54-56]. It would be interesting to determine whether these conserved non-coding sequences are involved.

GC content

The average GC content based upon the assembly Meug_2.0 is 38.8% (Table 2), while the GC content based upon cytometry is 34%. This is lower than the GC content for human (41%) but similar to opossum (38%). The tammar X also has a GC content (34%) lower than that of the opossum X (42%). Thus, tammar chromosomes are relatively GC poor. The proportion of CpGs in the tammar genome is higher than that of the opossum, but similar to human (Table 2). The GC content was also calculated from RIKEN full-length cDNA pools and varied from 44% to 49% across tissue types (Table S17 in Additional file 1), indicating that the lower GC content of the tammar genome is contained within non-exonic regions.

Repeats

The repeat content of the tammar wallaby genome was assessed using RepeatMasker, RepeatModeler and *ab initio* repeat prediction programs. The Repbase database of consensus repeat sequences was used to identify repeats in the genome derived from known classes of elements [57] (Table 2). RepeatModeler uses a variety of *ab initio* tools to identify repetitive sequences irrespective of known classes [58]. After identification, the putative *de novo* repeats were mapped against the Repbase repeat annotations using BLAST. Any *de novo* repeat with at least 50% identity and coverage was annotated as that specific Repbase element. All putative *de novo* repeats that could not be annotated were considered *bona fide*, *de novo* repeats. The results from the database and *de novo* RepeatMasker annotations were combined, and any overlapping annotations were merged if they were of the same class of repeat element. Overlapping repeats from different classes were reported; therefore, each position

Table 2 Comparison of repeat landscape in tammar and other mammals

	Tammar	Opossum	Platypus	Human	Mouse
Total assembly size (Gb)	2.7	3.48	2.3	2.88	2.55
Interspersed repeats (%)					
Total	52.8	52.2	44.6	45.5	40.9
LINE/non-LTR retroelements	28.6	29.2	21.0	20.0	19.6
SINE	11.7	10.4	22.4	12.6	7.2
ERV	3.9	10.6	0.47	8.1	9.8
DNA transposon	2.9	1.7	1.1	2.8	0.8
C+G (%)	38.8	37.7	45.5	40.9	41.8
CpG (%)	3.5	2.3	NA	3.7	3.9

Comparative analyses of the interspersed repeat content in the tammar and other sequenced mammalian genomes. Repeat modeller combined dataset includes *ab initio* annotation of *de novo* repeats. ERV, endogenous retroviral element; LTR, long terminal repeat; NA, not available.

in the genome may have more than one unique annotation.

The total proportion of repetitive sequence in the tammar was found to be 52.8%, although this is probably an underestimate resulting from the low coverage. This is similar to the repeat content of the opossum genome (52.2%). The proportion of LINEs and SINEs was also similar between opossum and tammar; however, the overall content for long terminal repeat (LTR) elements was significantly below that observed for any other mammal (only 3.91%) with the exception of the platypus (about 0.47%). Interestingly, 36 elements were identified that were tammar-specific, including novel LTR elements (25), SINEs (1), LINEs (4) and DNA elements (3). Moreover, analyses of the small RNA pools that emanate from repeats (see below) allowed for identification of a novel SINE class that is rRNA derived and shared among all mammals (J Lindsay, DM Carone, E Murchison, G Hannon, AJ Pask, MB Renfree, and RJ O'Neill, unpublished results; MS Longo, LE Hall, S Trusiak, MJ O'Neill, and RJ O'Neill, unpublished results).

Given the unique small size of the tammar centromere, estimated to cover only 450 kb [48], the genome was further scanned for putative pericentric regions using our previously annotated centromere repeat elements [59]. We identified 66,256 contigs in 53,241 scaffolds as having centromeric sequences and these were further examined for repeat structure. Analyses of these regions confirms the proposed punctate distribution of repeats within pericentromeric regions of the tammar [48,60] and indicate the absence of monomeric satellite repeats in the centromeres of this species (J Lindsay, S Al Seesi, RJ O'Neill, unpublished results) compared with many others (reviewed in [61,62]).

The tammar transcriptome

Sequencing of the tammar genome has been augmented by extensive transcriptomic sequencing from multiple tissues using both Sanger sequencing and the Roche 454 platform by a number of different groups. Transcriptome datasets collected are summarized in Table S17 in Additional file 1 and are described in more detail in several companion papers. Sequences from the multiple tissues have been combined to assess the assembly and annotation, and to provide a resource that supplements the low coverage tammar genome by identifying and adding unsequenced and unannotated genes.

Transcriptomes of the testis [DDBJ:FY644883-FY736474], ovary [DDBJ:FY602565-FY644882], mammary gland [GenBank:EX195538-EX203564, GenBank:EX203644-EX210452], gravid uterus [DDBJ:FY469875-FY560833], hypothalamus [DDBJ:FY560834-FY602565) and cervical and thoracic thymus [SRA:SRX019249, SRA:SRX019250] were sequenced. Each dataset was aligned to the assembly (Meug_1.0) using BLASTN. The proportion of reads that mapped varied between approximately 50% and 90% depending on the tissues of origin (Figure S2a Additional file 3). Of the successfully mapped reads, the proportion aligning to annotated genes (Ensembl annotation or 2 kb up- or downstream) were more similar between libraries (Figure S2b in Additional file 3). However, the lowest rates at which reads mapped to annotated genes in the genome were observed in transcripts from the two thymuses and the mammary gland. The former is unsurprising as a large number of immune genes are expressed in the thymus and are likely to be more difficult to annotate by projection due to their rapid evolution. The lower rate at which these ESTs aligned to annotated genes in mammary gland may reflect the highly sophisticated and complex lactation of marsupials (reviewed in [12]), a conclusion supported by the large number of unique genes identified with whey acidic protein and lipid domains (Figure 3). The mammary transcriptome may also contain a large number of immune transcripts. Together, these findings suggest a high degree of innovation in immune and lactation genes in the tammar. Previous analyses revealed that about 10% of transcripts in the mammary transcriptome were marsupial-specific and up to 15% are therian-specific [63]. Conversely, the high proportion of reads mapping to annotated genes in the testis and ovary (> 80%) suggest that there is significant conservation of active genes involved in reproduction between mammalian species (see section on 'Reproductive genes'

The testis, ovary, hypothalamus and gravid uterus full-length cDNA libraries were end-sequenced at RIKEN to evaluate composition and complexity of each transcriptome. We produced 360,350 Sanger reads in total (Table S18a in Additional file 1). Reads were clustered and the ratio of the clusters to reads was used as an estimate of the tissue's transcriptomic complexity. The hypothalamus showed the highest complexity (44.3%), whereas ovary showed the lowest (18.8%). We then looked for representative genes in each library by aligning reads to the Refseq database using BLASTN. For example, homologues of *KLH10* and *ODF1/2*, both of which function in spermatogenesis and male fertility, were found to be highly represented in the testis library (4.3% and 3.5% respectively). The hypothalamus library was rich in tubulin family genes (7.9% of reads), and hormone-related genes such as *SST* (somatostatin; 1.8% of reads) (see Table S18b in Additional file 1 for details).

Highly divergent or tammar-specific transcripts

Based upon stringent alignments to Kyoto Encyclopedia of Genes and Genomes genes (E-value < 10^{-30}), it was initially estimated that up to 17% of ovary clusters, 22% of testis clusters, 29% of gravid uterus clusters and 52%

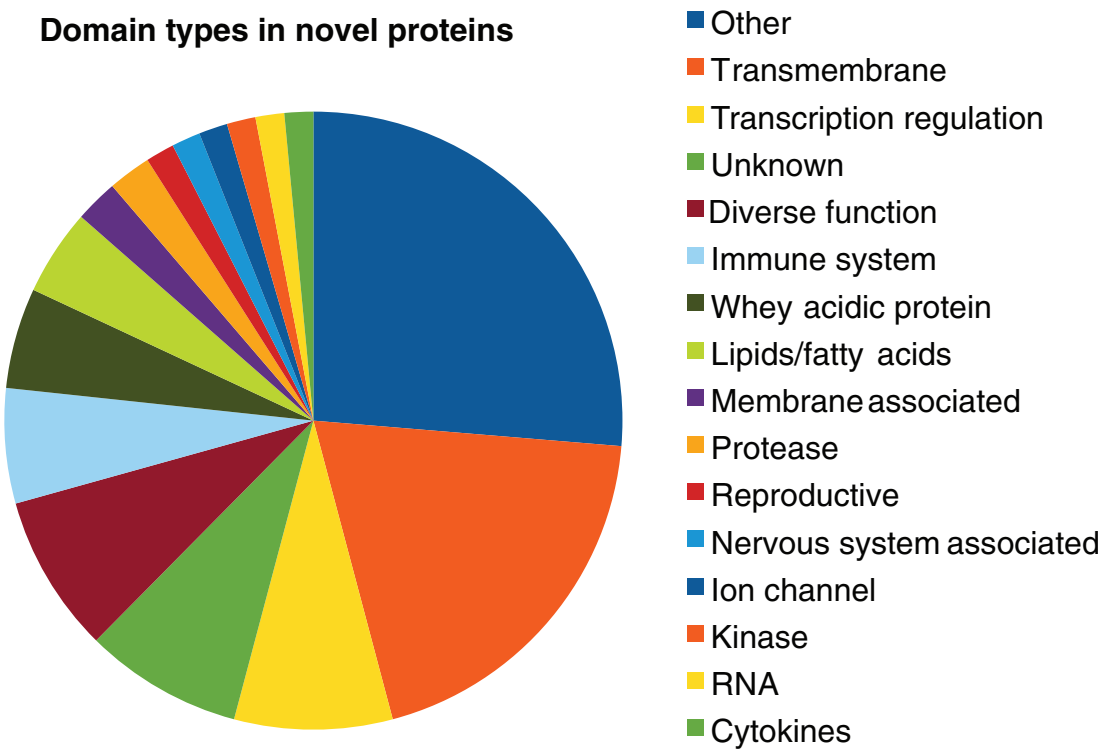


Figure 3 Classification of novel tammar genes. Summary of protein domains contained within translated novel ESTs isolated from the tammar transcriptomes. A large proportion of unique genes contain receptor or transcriptional regulator domains. The next largest classes of unique ESTs were immune genes, whey acidic protein and lipid domain containing genes. These findings suggest a rapid diversification of genes associated with immune function and lactation in the tammar.

of hypothalamus clusters were tammar-specific or highly divergent. Unique genes were identified by clustering of the EST libraries (to remove redundancy) followed by alignment of the unique reads to dbEST (NCBI) with BLASTN [64] using an E-value threshold of 10^{-5}. We identified 4,678 unique ESTs (6.1%) from a total of 76,171 input ESTs (following clustering) and used these for further analyses. Sequences were translated using OrfPredictor [65] and passed through PfamA [66] for classification. Of the unique genes that could be classified using this approach, many appear to be receptors or transcriptional regulators (Figure 3). A large number of unique ESTs contained whey acidic protein and lipid domains, common in milk proteins, suggesting a rapid diversification of these genes in the tammar genome. An EST containing a unique zona pellucida domain was also identified. Detailed expression was examined for 32 unique genes isolated from the RIKEN testis RNA-Seq pool. Of the initial 32, 11 were gonad-specific. Spatial expression of five of these genes was examined by *in situ* hybridization in adult testes and ovaries. One gene was germ cell-specific, two genes had weak signals in the somatic tissue and the remaining two genes were not detected.

Small RNAs

Recently, it has become clear that small RNAs are essential regulatory molecules involved in a variety of pathways, including gene regulation, chromatin dynamics and genome defense. While many small RNA classes appear to be well conserved, such as the miRNAs, it has become evident that small RNA classes can also evolve rapidly and contribute to species incompatibilities [67-70]. Our analyses of the tammar small RNAs focused on known classes of small RNAs, miRNAs, and Piwi-interacting RNAs (piRNAs), as well as a novel class first identified in the tammar wallaby, centromere repeat-associated short interacting RNAs (crasiRNAs) [48] (Figure 4a).

Small RNAs in the size range 18 to 25 nucleotides, including miRNAs, from neonatal fibroblasts, liver, ovary, testis and brain were sequenced [GEO:GSE30370, SRA: SRP007394] and annotated. Following the mapping pipeline (Supplementary methods in Additional file 1), hairpin predictions for the precursor sequence within the tammar genome for each small RNA in this class were used. Those small RNAs derived from a genomic location with a *bona fide* hairpin were classified as miRNA genes and further analyzed for both conserved and novel miRNAs. Of those annotated in Ensembl, one was confirmed as a novel

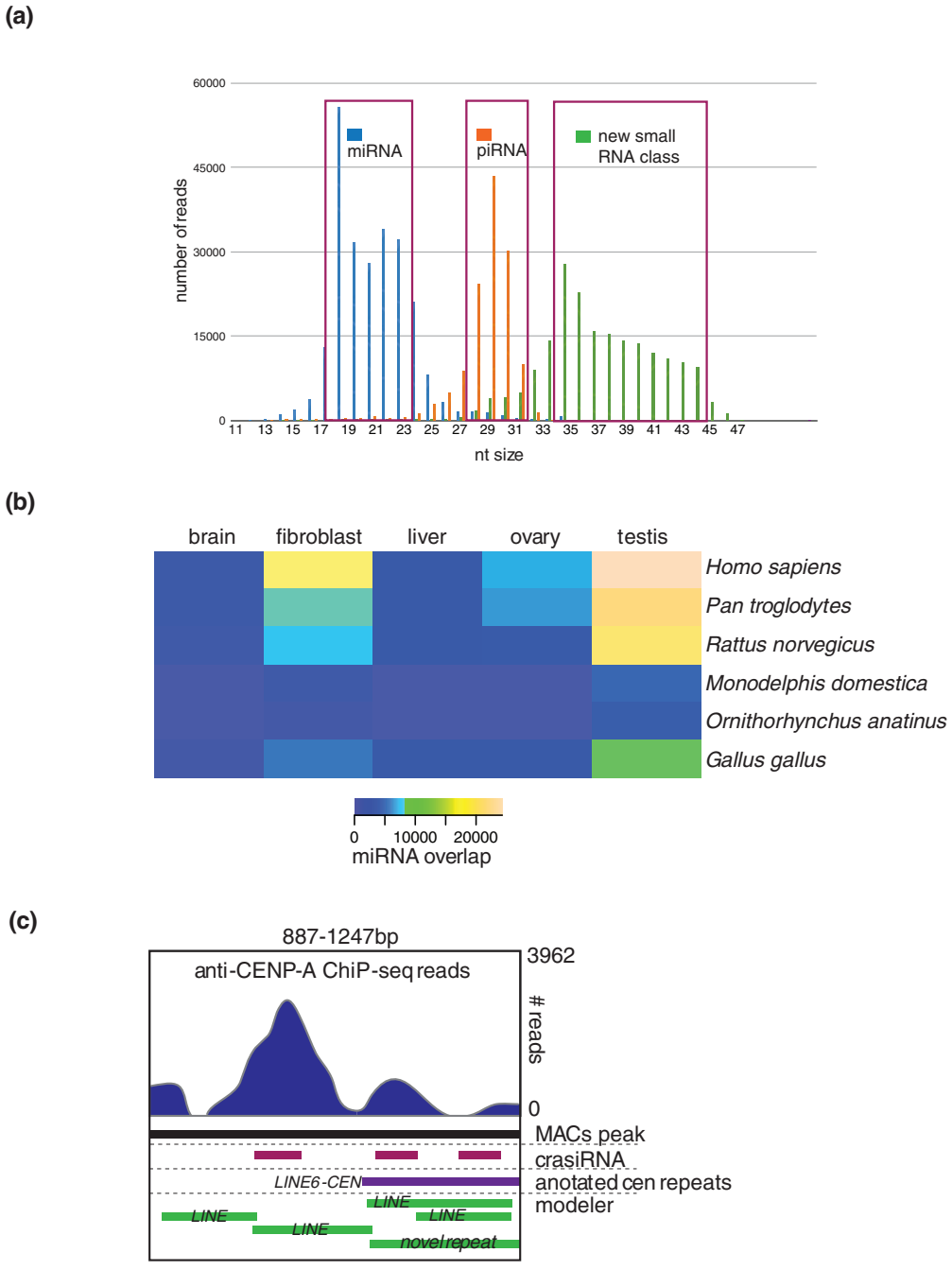


Figure 4 A survey of both conserved and novel small RNAs in the tammar genome. **(a)** Size ranges of the major classes of small RNAs. The x-axis shows number of reads mapped to the tammar genome while the size of the read in nucleotides is on the y-axis. Boxes denote each major class analyzed in the tammar. Classes targeted for sequencing and full annotation include the miRNAs (18 to 22 nucleotides), the piRNAs (28 to 32 nucleotides) and the newly discovered crasiRNAs (35 to 45 nucleotides). **(b)** Five tammar miRNA libraries (brain, liver, fibroblast, ovary and testis) were pooled and mapped to the tammar genome. miRNAs with a complete overlap with miRBase entries mapped to the tammar genome were considered conserved and annotated according to species. Heat map showing the frequency of conserved mirBase entries per tissue and per species as identified in the tammar. A high degree of overlap (that is, conservation) was observed between tammar and human for fibroblast and testis, but a relatively low degree of overlap was observed for the brain. **(c)** The complex tammar centromere. Genome browser view of chromatin immunoprecipitation-sequencing (ChIP-Seq) for DNA bound by the centromere-specific histone CENP-A mapped to a centromeric contig (top, blue). Nucleotide position on the contig is shown on the x-axis and depth of reads shown on the y-axis. Tracks illustrated: MACs peak (model-based analyses of Chip-Seq (black); locations for mapped reads of crasiRNAs (red); location of annotated centromere sequences (in this example, the centromeric LINE L6; purple); modeler repeat prediction track (green). crasiRNAs co-localize to DNA found in CENP-A-containing nucleosomes and are enriched in regions containing known centromere sequences.

tammar miRNA gene and a further 56 as putative miRNA genes. Using a cross-database mapping scheme targeting both miRBase [71-74] and the tammar genome assembly (Supplementary methods in Additional file 1), 11% of miRNAs in the tammar tissues analyzed were related to previously annotated miRNAs (Figure 4b). However, the majority of miRNA alignments in the genome did not overlap with previously identified miRNAs and are thus considered novel. Combining these datasets with the gene annotations, 147 target genes were conserved with other mammals. Of these, four were shared between mouse and tammar and twelve were shared between human and tammar, thus indicating that the tammar miRNA repository might provide new targets for study in these species. Moreover, there were nine novel target genes in the tammar genome, pointing to both tammar-specific miRNA regulation as well as potentially novel targets in human that were previously unknown. Small RNAs were also identified in the HOX clusters (see 'HOX gene patterning in the limb' section below).

piRNAs are predominantly found in ovaries and testes [69,75,76]. Global comparisons to RepBase and our *de novo* repeat database show that the overall composition of tammar piRNAs in testis is similar in terms of repeat element type (that is, SINEs, LINEs, and so on) to that observed for other species. In addition, there were ovary-specific piRNAs derived from *de novo* tammar repeats, which may contribute to the observed hybrid incompatibility observed in this group of marsupial mammals [60,77-79].

The first identification of crasiRNAs (35 to 42 nucleotides) found that they contain centromere repeat-derived sequences specific to the retroelement KERV (kangaroo endogenous retrovirus) [48,60]. Approximately 68% of repeat-associated crasiRNAs mapped within viral-derived repeats (such as KERV) [80], SINE, and LINE elements (J Lindsay, S Al Seesi, RJ O'Neill, unpublished results). Many of these elements mapped to centromeres using primed *in situ* labeling (PRINS), and mapped to scaffolds enriched for centromere-specific repeats and CENP-A-containing nucleosomes (as determined by ChIP-seq) [GEO:GSE30371, SRA:SRP007562], confirming that this pool consists of centromeric elements (Figure 4c). Closer examination of this sequence pool and the progenitor sequences within the genome uncovered a distinct motif specific to the crasiRNAs, which may indicate novel biogenesis (J Lindsay, S Al Seesi, and RJ O'Neill, unpublished results).

Immunity

The organization of the tammar MHC is vastly different from that of other mammals [81,82]. Rather than forming a single cluster, MHC genes are found on every chromosome, except the sex chromosomes (Figure 2). The MHC itself is found on chromosome 2q and contains 132 genes spanning 4 Mb [51]. This region was sequenced using a BAC-based Sanger sequencing strategy as it did not assemble well from the low-coverage sequencing. An expansion of MHC class II genes is accompanied by duplication of antigen-processing genes. The seven classical MHC class I genes are all found outside the core MHC region. KERVs may have contributed to this re-organization (Figure 2).

The tammar wallaby has two thymuses: a thoracic thymus (typically found in all mammals) and a dominant cervical thymus. Based on digital gene expression profiles both thymuses appear functionally equivalent and drive T-cell development [83]. Transcriptomic sequencing also shows that both thymuses express genes that mediate distinct phases of T-cell differentiation, including the initial commitment of blood stem cells to the T lineage (for example, *IL-7R*, *NOTCH1*, *GATA3*, *SPI1*, *IKZF1*), the generation of T-cell receptor diversity and development of the thymic environment (for example, *TRAF6*, *TP63* and *LTBR*). In the thymus transcriptomes, we identified and annotated 34 cytokines and their receptors (10 chemokines, 22 interleukins and 2 interferons), 22 natural killer cell receptors (20 leukocyte receptor complex (LRC) genes and 2 natural killer complex (NKC) genes), 3 antimicrobial peptides (2 beta-defensins and 1 cathelicidin), post-switch immunoglobulin isotypes *IgA* and *IgG* and *CD4* and *CD8* T-cell markers.

At birth, the altricial pouch young is exposed to a variety of different bacterial species in the pouch. These include *Acinetobacter* spp., *Escherichia coli* and *Corynebacteria* spp. [84]. These bacteria remain in the pouch despite the female tammar extensively cleaning the pouch by licking before birth. To survive in this pathogen-laden environment, the immunologically naive neonate is reliant on immune factors, which are transmitted from the mother through the milk. The sequencing of the genome uncovered a family of cathelicidin genes, which are expressed in the mammary gland during lactation and encode powerful antimicrobial peptides. These peptides may provide unique opportunities to develop novel therapeutics against emerging multidrug-resistant superbugs.

Due to the rapid evolution of immune genes, a high proportion of tammar immune genes were not annotated using automated annotation pipelines. For this reason an Immunome Database for Marsupials and Monotremes has been established [85]. This database contains over 5,000 marsupial and monotreme immune sequences from a variety of EST projects, as well as expert-curated gene predictions. Marsupial chemokine, interleukin, natural killer cell receptor, surface receptor and antimicrobial peptide gene sequences are also available. Genomic evidence confirms that the marsupial immune system is on par with the eutherian immune system in terms of complexity.

Sex chromosomes

Marsupial sex chromosomes have been shown to represent the ancestral sex chromosomes, to which an autosomal region was fused early in the eutherian radiation. Thus, the basic marsupial X shares homology with the long arm and pericentric region of the human X [35,36]. The tammar Y shares only five genes with the degraded eutherian Y [86] (Figure 5).

Marsupial sex chromosomes lack the autosomal addition and so are expected to be smaller than those of eutherian mammals. The opossum X is about 97 Mb (Table S12 in Additional file 1). The larger size of the tammar X (150 Mb) reflects the addition of a heterochromatic arm containing satellite repeats and the nucleolus-organizing region [59]. Of the 451 protein coding genes on the opossum X chromosome, 302 have orthologues in the tammar Ensembl gene build. Gene mapping indicates that the gene order within the tammar X is scrambled with respect to both the opossum and human X chromosomes [37]. This scrambling of the marsupial X contrasts to the eutherian X chromosome, which is almost identical in gene content and order between even the most distantly related taxa [87,88]. The rigid conservation of the eutherian X was hypothesized to be the result of strong purifying selection against rearrangements that might interrupt a chromosome-wide mechanism to effect X-chromosome inactivation. Consistent with this hypothesis, inactivation on the scrambled marsupial X is incomplete, locus-specific, and does not appear to be controlled by an inactivation center [54,56].

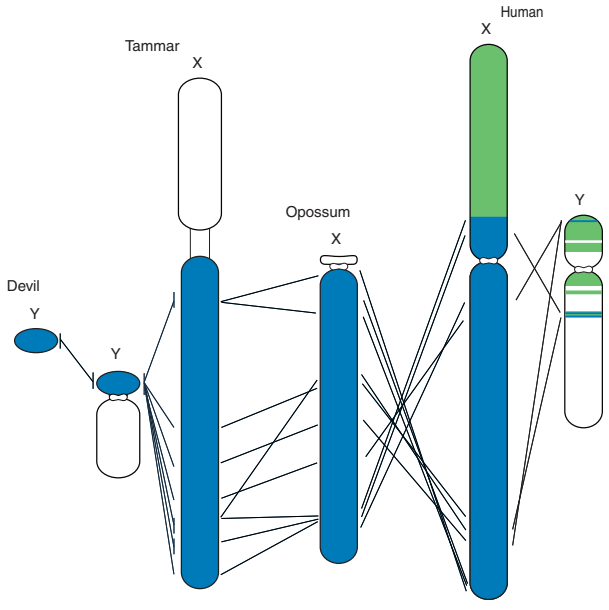


Figure 5 Comparative map of X and Y chromosomes. Comparison of X/Y shared gene locations on the tammar wallaby, grey short-tailed opossum and human X chromosomes. Blue represents the X conserved region, which is common to all therian X chromosomes. Green represents the X added region, which is on the X in eutherian mammals, but autosomal in marsupial mammals. Ten genes have been identified on the short arm of the tammar Y chromosome, all with a partner on the X, and an orthologue on the Tasmanian devil Y. In contrast, only four genes on the human Y have a partner on the conserved region of the X.

In many marsupial species the Y chromosome is a minute element of about 12 Mb. The tammar Y is larger, as the result of the addition to the X and Y in the early macropodid radiation of a heterochromatic long arm that contained the nucleolar organizing region (NOR) and NOR-associated repeats [59]. Degradation of the Y removed active rDNA genes but left repetitive sequences with homology to the NOR-bearing short arm of the X [89,90]. The tammar Y chromosome bears at least ten genes, which are all located on the tiny short arm of the Y (reviewed in [91]) (V Murtagh, N Sankovic, ML Delbridge, Y Kuroki, JJ Boore, A Toyoda, KS Jordan, AJ Pask, MB Renfree, A Fujiyama, JAM Graves and PD Waters, unpublished results). All ten have orthologues on the Y of a distantly related Australian dasyurid marsupial, the Tasmanian devil, implying that the marsupial Y chromosome is conserved (Figure 5). It has degraded more slowly than the eutherian Y, which retains only four (human) or five (other mammals) genes from the ancient XY pair [91,92].

Like most genes on the human Y, all of these tammar Y genes have an X partner, from which they clearly diverged. Some tammar Y genes are expressed exclusively in the testis (for example, the marsupial-specific *ATRY* [93]), but most have widespread expression. Phylogenetic analysis of the X and Y copies of these ten tammar XY genes indicate that marsupial Y genes have a complex evolutionary history.

X chromosome inactivation

Epigenetic silencing of one X chromosome occurs in female mammals as a means of dosage compensation between XX females and XY males. Classic work on kangaroos established that X inactivation occurs in marsupials, but is paternal, incomplete and tissue-specific [94] and apparently occurs in the absence of the *XIST* controlling element [95,96]. Using tammar sequence to isolate X-borne genes and study their expression at the level of individual nuclei using RNA *in situ* hybridization, it has been found that different genes have a characteristic frequency of expression from one or both loci, suggesting that it is the probability of expression rather than the rate of transcription that is controlled [54]. The absence of clustering of high- or low-expressing genes has not so far provided evidence for an inactivation center. It appears that X inactivation in marsupials, like eutherians, uses a repressive histone-mediated gene silencing, and although inactive marks are not identical [55,56], they do have

H3K27 trimethylation and targeting to the perinucleolar compartment [97].

Reproductive genes

Marsupials differ from eutherian mammals primarily in their unique mode of reproduction. In contrast to mice and humans, in which sexual differentiation occurs *in utero*, the altricial 440 mg tammar neonate has indifferent gonads on the day of birth and does not undergo gonadal sex determination until approximately 2 days later (testis) and 8 days later (ovary) [22]. This postnatal differentiation of the gonads therefore provides an unparalleled model for studying sex determination and sexual differentiation and enables experimental manipulation not possible in eutherian species. We have shown that almost all genes critical for testis and ovarian development are highly conserved between the tammar, mouse and human at the molecular level [98,99], but their precise role in gonadogenesis may differ between the mammalian groups.

Gonadal differentiation genes

ATRX is an ultra-conserved, X-linked gene essential for normal testis development in humans. Marsupials are unique among the mammals in that they have orthologues of this gene on both their X and Y chromosomes (*ATRX* and *ATRY*, respectively). Almost all X-linked genes once shared a partner on the Y, but the vast majority of these have been lost during its progressive degeneration. The Y-linked *ATRX* orthologue was lost in the eutherian lineage before their radiation, but was retained in the marsupial lineage. *ATRY* shows functional specialization, and is exclusively expressed in the developing and adult testis of the tammar, while tammar *ATRX* is broadly expressed, but is absent in the developing testis, unlike eutherians [93]. The distribution of *ATRX* mRNA and protein in the developing gonads is ultra-conserved between the tammar and the mouse [100], and is found within the germ cells and somatic cells. *ATRX* therefore appears to have a critical and conserved role in normal development of the testis and ovary that has remained unchanged for up to 148 million years of mammalian evolution [100].

Desert hedgehog (*DHH*) is another essential signaling molecule required for normal testicular patterning in mice and humans. Members of the hedgehog family of secreted proteins act as intercellular transducers that control tissue patterning across the entire embryo. Like other hedgehog proteins, DHH signals through the PTCH receptors 1 and 2 [101]. *DHH*, *PTCH1* and *PTCH2* in the tammar are highly conserved with their eutherian orthologues. However, unlike in eutherian mammals, *DHH* expression is not restricted to the testes during tammar development, but is also detected in the developing ovary (WA O'Hara, WJ Azar, RR Behringer, MB Renfree, and AJ Pask, unpublished results). Furthermore, hedgehog-signaling inhibitors disrupt both testicular and ovarian differentiation [101]. Together, these data confirm a highly conserved role for DHH in the formation of both the male and female tammar gonad.

Most interestingly, *DHH* is clearly a mammal-specific gonadal development gene. Hedgehog orthologues that are described as DHH in non-mammalian vertebrates actually form a distinct lineage no more closely related to mammalian DHH than they are to Sonic hedgehog (SHH) or Indian hedgehog (IHH) orthologues (Figure 6). Thus, *DHH* is the only mammal-specific gonadal development gene other than *SRY* so far discovered. In the tammar *PTCH2* a novel exon (exon 21a) was detected that is not annotated in any eutherian PTCH2 proteins (WA O'Hara, WJ Azar, RR Behringer, MB Renfree, and AJ Pask, unpublished results). These analyses suggest that DHH evolved recently in vertebrates, yet acquired a critical role in mammalian gonadal development before the eutherian-marsupial divergence. However, the role of DHH in gonadogenesis has become more specialized to the testis in the eutherian lineage.

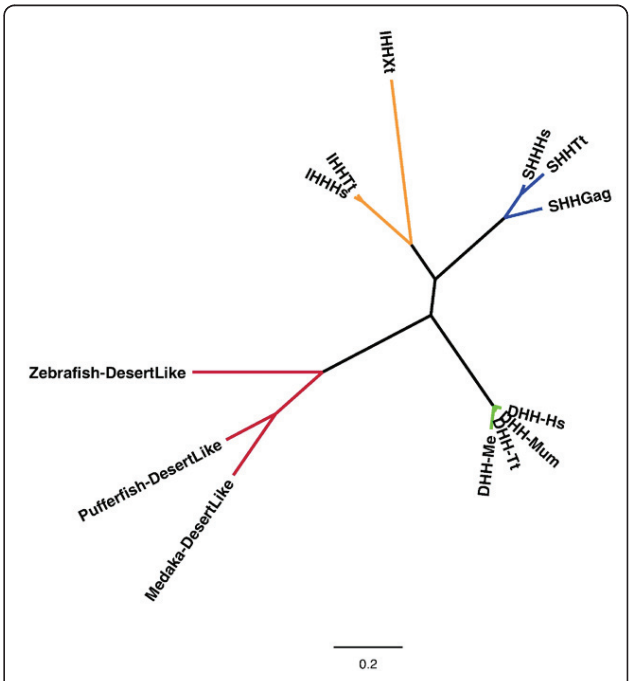


Figure 6 *Desert hedgehog* phylogeny. A phylogenetic tree showing the relationship of the SHH, IHH, DHH, and fish desert-like genes. Each group is composed of representatives from mammalian and non-mammalian species. The mammalian DHH group (green) clusters tightly and forms a separate linage to the fish DHH-like genes (red), which are no more closely related to DHH than they are to vertebrate IHH (yellow) and SHH (blue). Hs, human; Tt, dolphin; Xt, *Xenopus*; Gag, chicken; Mum, mouse; Me, tammar.

Germ cell genes

The differentiation of the somatic cell lineages in the ovary and testis, mediated by the pathways described above, is critical for the subsequent development of the germ cells. Germ cells carry the genetic information from one generation to the next, making them arguably the most important cell lineage in the body. Comparative analyses of the genes essential for mouse and human germ cell development using the tammar genome presented an unexpected paradox. It was presumed that the genes mediating germ cell specification and development in mammals would be highly conserved because this cell lineage is critical for species' survival. However, our analyses indicate that many genes are rapidly evolving and likely to be controlled by specific elements in each mammalian lineage.

Orthologues of genes critical for the specification and development of eutherian germ cells, including *BMP4*, *PRDM1* and *PRDM14*, were identified in the tammar genome. The tammar genome also contains transcripts for *DDX4* (*VASA*) [102]. One transcript encodes a full length protein and the other has exon 4 spliced out. *In silico* analysis and 3' RACE showed that tammar *DDX4* also utilizes more than one polyA signal [102]. The significance of these differentially spliced and alternatively polyadenylated *DDX4* transcripts is unknown but may represent alternative mechanisms for controlling *DDX4* expression; the 3' untranslated region of *DDX4* in many species controls the localization, stabilization and translation of the gene [103]. Some genes expressed in murine primordial germ cells (PGCs) but not essential for their development lack marsupial orthologues. *Stella* is expressed in PGCs and in pluripotent cells but mice lacking *Stella* do not have any defects in germ cell specification or development [104]. In humans, *STELLA* is located on chromosome 12p13, a region known for structural chromosomal changes that are commonly associated with germ cell tumor formation. This region contains a cluster of genes, including *NANOG* and *GDF3* [105], that are expressed in pluripotent cells. The syntenic region in the tammar and opossum contains *NANOG* and *GDF3* but *STELLA* is absent, suggesting it evolved only recently in the eutherian lineage. Similarly, interferon inducible transmembrane protein (Ifitm)3 is produced in cells competent to form PGCs in mice [106], and both Ifitm3 and Ifitm1 are thought to mediate migration of PGCs from the posterior mesoderm into the endoderm [107]. Ifitm proteins 1 and 3 are expressed in early murine PGCs [106,108] but deletion of the locus containing Ifitm1 and Ifitm3 has no apparent effect on germ cell specification or migration [109]. The tammar genome contains several IFITM orthologues, some expressed in the early embryo, as in the mouse. The low sequence conservation between marsupial and eutherian IFITM orthologues suggests that the IFITMs may not be critical for mammalian germ cell development.

Spermatogenesis genes

The genes regulating the later differentiation of the germ cells into mature oocytes and spermatocytes, especially those controlling spermatogenesis, are much more conserved between marsupials and eutherians than the signals that trigger their initial development. In eutherian mammals, there are a disproportionately high number of genes involved in spermatogenesis located on the X chromosome [110]. From the genome analyses in the tammar, it is clear that some of these genes were originally autosomal, and others appear to be on the ancestral X of the therian ancestor.

AKAP4, a scaffold protein essential for fibrous sheath assembly during spermatogenesis, is X-linked in the tammar as it is in eutherian mammals and maintains a highly conserved role in spermatogenesis [111]. In contrast, the *Kallman syndrome gene 1* (*KAL1*) is X-linked in eutherians but autosomal in the tammar, located on chromosome 5p in a block of genes transposed to the X chromosome in an ancestral eutherian [52]. Despite its different chromosomal location, *KAL1* is highly conserved and expressed in neuronal tissues as well as in the developing and adult gonads throughout spermatogenesis. Thus *KAL1* probably evolved its role in mammalian gametogenesis before its relocation to the eutherian X [52]. Another eutherian X-linked gene, *TGIFLX* is absent from the tammar genome, but its progenitor, *TGIF2*, is present and appears to function in gametogenesis. Once again, this suggests that the gene had a role in spermatogenesis before its retrotransposition to the eutherian X [53]. These genomic and functional analyses not only shed light on the control of mammalian spermatogenesis, but also on genome evolution. These data support the theory that the X chromosome has selectively recruited and maintained spermatogenesis genes during eutherian evolution.

Developmental genes

The segregation of the first cell lineages and specification of embryonic and extra-embryonic cell lineages have been studied extensively in the mouse. However, the mouse has a highly specialized embryogenesis, quite different from that of other mammals. Unlike a typical eutherian blastocyst with its inner cell mass, the tammar conceptus forms a unilaminar blastocyst of approximately 100 cells that lacks a readily defined pluriblast in the form of an inner cell mass. It can undergo a prolonged period of diapause. Thus, these differences highlight the developmental plasticity of mammalian embryos

and genome analysis may provide comparative data that clarify the underlying control mechanisms of early mammalian development.

Pluripotency genes

The tammar embryo develops when the embryonic disc forms on the blastocyst surface. The difference in embryo specification raises many interesting questions about early marsupial and mammalian development in general. After the differentiation of the embryonic area, the tammar embryo proper develops in a planar fashion on the surface of the embryonic vesicle. This makes the study of early embryonic events and morphogenesis easier to observe and manipulate than in the complicated egg cylinder formed in the mouse.

It is still unknown how the cells are specified in the unilaminar blastocyst that will go on to form the embryo in the tammar, but in the polyovular dasyurid marsupials, and also in the opossum, there appears to be cellular polarity in cleavage stages (reviewed in [112]). Whether the signals that regulate specification and induction are the same or different from those that regulate the specification of the eutherian mammal inner cell mass is under investigation. However, *POU5F1* expression is limited to pluripotent cell types in the tammar as in eutherians. Marsupials additionally have a *POU2* orthologue that is similarly expressed in pluripotent tissues but is also expressed in a broad range of adult tissues, suggesting that unlike *POU5F1*, the role of POU2 may function in maintaining multipotency in adult stem cells [113]. In the tammar, opossum and platypus genomes, but not in eutherian genomes, *POU2* is an ancient vertebrate paralogue of *POU5F1* [113,114]. Tammar wallaby *POU2* is co-expressed in embryonic pluripotent tissues with *POU5F1* but is also expressed in a broad range of adult tissues, suggesting it may also additionally function in maintaining multipotency in adult marsupial stem cells [113].

Orthologues of the vast majority of early developmental genes characterized in the mouse were identified in the tammar genome, including those encoding key transcription factors, such as *POU5F1*, *SOX2*, *NANOG*, *CDX2*, *EOMES*, *GATA4*, *GATA6* and *BRACHYURY*. Genes encoding components of key signaling pathways in early development are largely conserved between tammar and mouse. One exception is *TDGF1* (also called *CRIPTO*), which is present in eutherians but absent from the genome in tammar (as well as in those of opossum, platypus and non-mammalian vertebrates). *TDGF1* encodes a co-receptor of NODAL signaling, which has a central role in early germ layer formation and axial specification in the mouse and in self-renewal of human embryonic stem cells [115]. Thus, *TDGF1* is eutherian-specific, while the related paralogue *CFC1* (also called *CRYPTIC*) is widely conserved in all vertebrates. This suggests the evolution of partly divergent roles for NODAL signaling in early embryonic patterning among mammals.

Embryonic patterning

Once the early embryo is formed, the body plan must be established. The HOX genes are essential regulators of embryonic patterning in all animals, mediating the specification of structures along the anterior-posterior axis. In the tammar, as in all vertebrates, the HOX genes are arranged in four clusters. The clusters are low in repetitive elements compared to the rest of the genome (H Yu, Z-P Feng, RJ O'Neill, Y Hu, AJ Pask, D Carone, J Lindsay, G Shaw, AT Papenfuss, and MB Renfree, unpublished results). The tammar HOX clusters have a high degree of both conservation and innovation in the protein-coding and non-coding functional elements relative to eutherian mammals (Figure 7). Intronic regions are mostly divergent, but have isolated regions of high similarity corresponding to important enhancer elements. In eutherians, the clusters contain conserved intronic non-coding RNAs that are

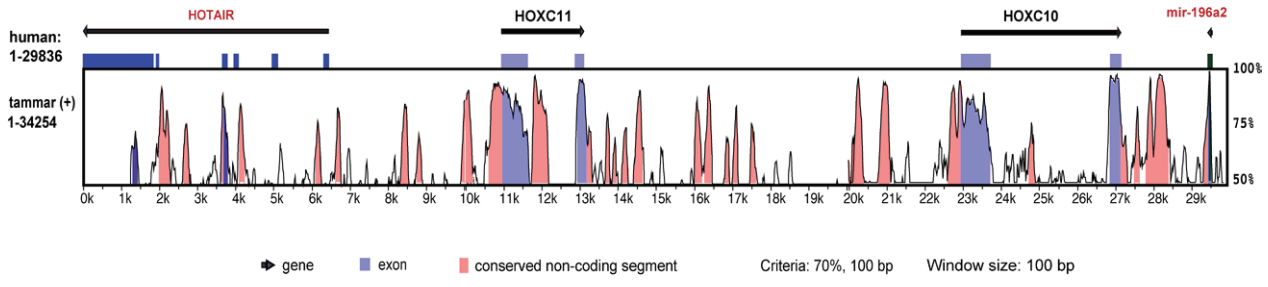


Figure 7 *HOX* genes in the tammar. mVISTA comparison of partial HOXC cluster highlights conserved HOX genes and non-coding RNAs between human and tammar. In the coding regions, *HOXC11* and *HOXC10* are highly conserved between human and tammar. In the intergenic regions, some conserved regions shown are non-coding RNAs (long non-coding RNA such as *HOTAIR*, and miRNAs such as mir-196) or unknown motifs participating in gene expression and regulation. The percentage of identities (50 to 100%) (Vertical axis) is displayed in the coordinates of the genomic sequence (horizontal axis).

likely to participate in gene regulation [116]. Using the tammar genome, a new tetrapod miRNA was identified by conservation analysis and confirmed by RT-PCR to be expressed in fibroblasts (H Yu, Z-P Feng, RJ O'Neill, Y Hu, AJ Pask, D Carone, J Lindsay, G Shaw, AT Papenfuss, and MB Renfree, unpublished results). In addition, two novel miRNAs were characterized that are not conserved in eutherian mammals (Figure 7).

The HOX clusters also contain a number of genes that are transcribed into long non-coding RNAs [117,118]. Three long non-coding RNAs previously identified in the mouse were identified in the tammar HOX gene clusters. *HOX* antisense intergenic RNA myeloid 1 (*HOTAIRM1*), located between *HOXA1* and *HOXA2*, is conserved in mammals and shows myeloid-specific expression [119]. Similarly, *HOXA11* antisense (*HOXA11AS*), located between *HOXA13* and *HOXA11*, is only conserved in mammals and is expressed during the human menstrual cycle [120]. Interestingly, antisense intergenic RNA (*HOTAIR*), located between *HOXC12* and *HOXC11*, was conserved between human, mouse and tammar only in exons 3 and 6 (Figure 7). *HOTAIR* is an important trans-regulator that controls *HOXD* but not *HOXC* gene expression during limb development [116,121] and participates in reprogramming chromatin state to promote cancer metastasis [122]. The expression of *HOTAIR* was confirmed by RT-PCR in the tammar, suggesting an important and conserved regulatory role for this gene. The functional consequences of the marsupial-specific miRNAs and variation in the long non-coding RNAs are yet to be determined, but indicate mammalian lineage-specific regulation of HOX genes that could be responsible for species phenotypic differences.

HOX gene patterning in the limb

Macropodid marsupials have very specialized limbs. The forelimb is developed at birth to allow the neonate to climb to the pouch to locate and attach to one of the four available teats [123] but the hind limb, which eventually becomes the dominant feature of this hopping family, is barely formed at birth. Despite its embryonic nature, it is already possible to see the syndactylus arrangement of digits in which digits 2 and 3 are fused, digit 4 is enlarged and digit 5 is reduced. HOX genes play an important role in this arrangement. In particular, *HOXA13* and *HOXD13* play essential roles in digit development (reviewed in [119]). *HOXA13* and *HOXD13* in the developing tammar limb have both a conserved and divergent expression pattern (KY Chew, H Yu, AJ Pask, G Shaw, and MB Renfree, unpublished results). Tammar HOXA13 has a transient expression compared to the chicken and mouse, while tammar *HOXD13* is expressed in distal limb elements, as in other vertebrate species [124,125]. Early differences in the expression pattern were observed in the specialized tammar hindlimb compared to other species. These subtle differences could direct the morphological specialization of the tammar hindlimb to allow for the hopping mode of locomotion.

Pre-natal growth and placental genes

Mammals require genes that regulate growth both pre- and postnatally. Genes of the growth hormone/insulin-like growth factor-I (GH-IGF-I) axis are highly conserved in marsupials owing to their important function in pre- and postnatal growth. Sequencing and expression analysis of the GH receptor gene shows that exon 3, which is associated with variable growth and IGF-1 physiology in humans, is specific to the eutherian lineage and has undergone more rapid evolution in species with placental variants of GH and prolactin, indicating a possible fetal-specific role for the GH receptor in these species [126].

Prenatally, the placenta is a critical regulator of fetal growth. Genes involved in growth regulation in eutherian mammals (GH, GH receptor, prolactin, luteinizing hormone, IGF-1, IGF-2, insulin and their receptors) are all highly conserved in the tammar and all are expressed in the yolk sac placenta of the tammar wallaby, suggesting a conserved role for these hormones and growth factors during pregnancy in therian mammals [127]. GH and its receptor appear to be under tight regulation in the placenta, with expression increasing dramatically after close attachment of the placenta to the endometrium. Placental expression of both *GH* and *GHR* peaks at the end of pregnancy during the most rapid phase of fetal growth. These data indicate that GH and other pituitary hormones and growth factors are as essential for growth and development of the placenta in the tammar as in eutherian mammals.

Postnatally, maturation of GH-regulated growth in marsupials occurs during late lactation at a developmental stage equivalent to that of birth in precocial eutherian mammals (B Menzies, G Shaw, T Fletcher, AJ Pask, and MB Renfree, unpublished results) and it appears that this process is not associated with birth in mammals but instead with relative maturation of the young. This emphasizes the importance of nutrition in controlling early development in all mammals as they transition to independence. The neonatal tammar expresses ghrelin, a peptide that stimulates both hunger and GH release, in the stomach, ensuring that it can feed from a relatively early developmental stage [128].

Genomic imprinting

Genomic imprinting is a widespread epigenetic phenomenon characterized by differential expression of alleles, depending on their parent of origin. Imprinted genes in eutherian mammals regulate many aspects of early growth and development, especially those occurring in the

placenta. Most, but not all, genes that are imprinted in mouse and human have orthologues in the tammar genome; an exception is the Prader-Willi-Angelman syndrome region containing *SNRPN* and *UBE3A*, which does not exist in tammar, nor in monotremes, so was evidently recently constructed in eutherians by fusion and retrotransposition [129]. Some tammar orthologues of genes that are imprinted in eutherians are not imprinted [130,131]. So far the orthologues of 13 eutherian imprinted genes examined have a conserved expression in the marsupial placenta, but only 6 of these are imprinted in marsupials [132,133].

Marsupial orthologues of the classically imprinted IGF-2 receptor (*IGF2R*), insulin (*INS*) or paternally expressed gene 1/mesoderm specific transcript (*PEG1/*MEST) also show parent-of-origin expression in marsupials. However, some genes that are imprinted in eutherians, such as *Phlda2* in the *KCNQ1* domain, a negative regulator of placental growth, are not imprinted in the tammar [134]. This demonstrates that acquisition of genomic imprinting in the *KCNQ1* domain occurred specifically in the eutherian lineage after the divergence of marsupials, even though imprinting of the adjacent *H19-IGF2* domain [135] arose before the marsupial-eutherian split. A similar scenario applies to *DLK1*, *DIO3* and *RTL1* (*PEG11*), which are not imprinted in marsupials [130,136].

Differentially methylated regions (DMRs) are the most common signals controlling genomic imprinting in eutherian mammals. However, no DMRs were found near the tammar orthologues of the classically imprinted genes *IGF2R*, *INS* or *PEG1/MEST*, although these genes still showed parent of origin specific expression differences. Other marsupial imprinted genes (*H19*, *IGF2* and *PEG10*) do have DMRs, indicating that this mechanism of gene control evolved in the common therian ancestor at least 140 million years ago [133]. Using comparisons with the tammar genome, we have been able to reconstruct the emergence of an imprinted gene - *PEG10* [137]. *PEG10* is derived from a retrotransposon of the suchi-ichi family and was inserted after the prototherian-therian mammal divergence. This demonstrates that retrotransposition can drive the evolution of an imprinted region with a DMR [137]. In contrast, another retrotransposed gene also of the suchi-ichi family, *SIRH12*, has been identified specifically in the tammar genome but is not seen in eutherians. It appears to be tammar-specific since it is absent from the opossum genome. Its imprint status has yet to be ascertained [138].

The insulator genes *CTCF* (CCCTC-binding factor) and its paralogue *BORIS* (brother of regulator of imprinted sites) have orthologues in the tammar genome, and as in mouse, *CTCF* is expressed ubiquitously and *BORIS* is expressed in gonads. The existence of both genes in the monotreme and reptile genomes but the ubiquitous expression of *BORIS* in these species suggests that this gene became gonad-specific in therian mammals, coincident with the evolution of imprinting [139].

Although all imprinted genes so far identified in the mouse are expressed in the placenta, the few mouse genes that have been knocked out (for example, *Grb10*, *Peg3*) that are also imprinted in the fetal brain have marked behavioral effects [140]. We now know that there are additional autosomal genes in the cortex and hypothalamus with sex-specific imprinting [141,142], so we can expect an increase in the identification of imprinted brain genes that influence behavior. Since a large proportion of known imprinted genes also have a role in postnatal growth and nutrient supply, and marsupials depend much more on lactation than most other mammals (see below), it is possible that genomic imprinting might function in the marsupial mammary gland as it does in the placenta. Transcription analysis has confirmed that two genes critical for the onset of lactation in the tammar, *IGF2* and *INS*, are imprinted in the tammar mammary gland throughout the long period of lactation (JM Stringer, S Suzuki, G Shaw, AJ Pask, and MB Renfree, unpublished observations).

Olfaction

Vomeronasal organ

Pheromone detection in vertebrates is mostly mediated by the vomeronasal organ (VNO). The VNO organ is well developed in the tammar [123]. Pheromone detection occurs via two large families of vomeronasal receptors (VNRs). VN1Rs are associated with the protein Giα2 and VN2Rs with Goα using a signaling cascade dependent on transient receptor potential channel, subfamily C, member 2, encoded by the *TRPC2* gene. Previous characterizations of *TRPC2* in rodents led to confusion regarding its functionally relevant transcripts. Expression analysis and characterization of transcripts in the tammar have now shown that the locus consists of two distinct genes, one that is VNO-specific (*TRPC2* proper) and a previously unidentified copy that is ubiquitously expressed (*XNDR*) [143]. *XNDR* has homology with *XRCC1*, suggesting a role in DNA base excision repair due to homology with *XRCC1* [144]. *Giα2* and *Goα* have high sequence conservation and both are expressed in the tammar VNO and accessory olfactory bulb (NY Schneider, G Shaw, PT Fletcher, and MB Renfree, unpublished results). The projection pattern of the tammar Giα2 and Goα expressing receptor cells differs from that of the goat (uniform type) and the mouse (segregated type) and so may represent a new intermediate type (Figure 8a), with Goα not being confined to the rostral or caudal part of the accessory olfactory bulb, respectively, but found throughout (for example, [145]). Immunostaining results further suggest that Giα2 may follow the same pattern, but confirmation awaits the availability of a more specific antibody.

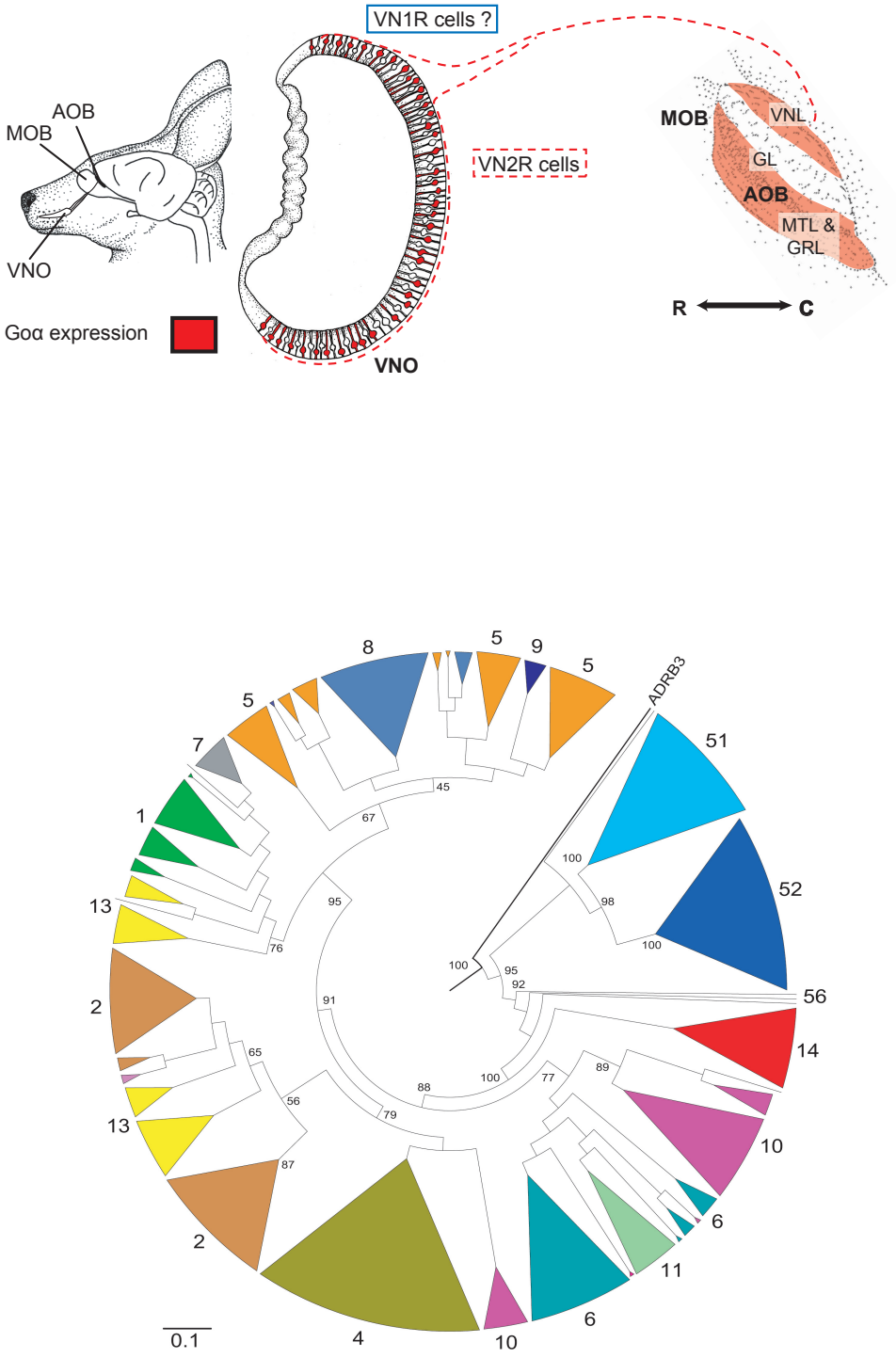


Figure 8 Olfaction in the tammar. **(a)** The olfactory apparatus of the tammar showing the pattern of vomeronasal receptor projections to the accessory olfactory bulb with the VN2 receptor cells (expressing Goα) projecting to all parts of the vomeronasal nerve layer (which may also be the case for the VN1 receptor cells (expressing Giα2). This projection pattern may reflect an intermediate type to the 'segregated type' and the 'uniform type' so far described. AOB, accessory olfactory bulb; GL, glomerular layer; GRL, granule cell layer; MOB, main olfactory bulb; MTL, mitral tufted cell layer; VNL, vomeronasal nerve layer; VNO, vomeronasal organ; VN1R and VN2R, vomeronasal receptors 1 and 2. **(b)** Olfactory receptor (OR) gene family in the tammar. The families of the OR gene repertoire. Neighbor joining tree of 456 full-length functional OR genes was rooted with opossum adrenergic β receptor. Only a few OR gene families (14, 51 and 52) have members that are most closely related to each other, whilst most other families have a high degree of relatedness to other families.

Olfactory receptor family genes

The marsupial genome has one of the largest mammalian olfactory receptor gene families, containing up to 1,500 olfactory receptor (OR) genes that apparently provide the tammar with a remarkably large range of odor detection in both the VNO and the main olfactory epithelium (A Mohammadi, H Patel, ML Delbridge, and JAM Graves, unpublished results) (Figure 8b). Certainly the neonate uses odor to locate the teat within the pouch [146]. There are 286 OR gene families in the tammar genome, with duplications especially in class I OR families OR8, -11, -13 and -51. However, the class II family OR14 has only one-third of the number found in the platypus genome, and eutherians have lost them altogether [147] (A Mohammadi, H Patel, ML Delbridge, and JAM Graves, unpublished results; Figure 8b). We found that class I OR families, particularly OR8, -11, -13 and -51, have undergone expansion in the tammar lineage, whereas the class II family OR14 has only one-third of the number found in the platypus genome and eutherians have lost them altogether [147].

The tammar and opossum have remarkably similar OR gene repertoires despite the significant variation in OR genes found in eutherian species that diverged about the same time. The OR genes are observed in gene clusters across all chromosomes, except chromosome 6 (Figure 2). The tammar Y chromosome has not yet been fully characterized but OR genes are not found on the Y of tammar or other mammals. They are found in the same conserved syntenic blocks as in the human (Figure 2) and opossum (data not shown), except clusters 9, 11 and 24, which have moved to 4q, and part of cluster 23, which is on 2q.

Lactation

Lactation is a defining character of mammals [148,149]. This is especially true of marsupials that give birth to highly altricial young that depend upon milk for growth and development during a relatively long lactation period. The marsupial mother dramatically alters milk production and composition throughout lactation, specifically for each stage of development of the pouch young [26,150,151]. They are even able to produce milk of differing compositions from adjacent mammary glands, a phenomenon known as concurrent asynchronous lactation (reviewed in [152]).

Lactation in the tammar extends for approximately 300 days and is divided into 3 phases based upon the sucking pattern of the young (phase 1 (late pregnancy-birth), lactogenesis; phase 2A (day 0 to 100), permanently attached to the teat; phase 2B (day 100 to 200), intermittently sucking and confined to the pouch; phase 3 (day 200 to 300), in and out of the pouch), accompanied by changes in milk composition and mammary gland gene expression [26]. The tammar mammary gland transcriptome consists of two groups of genes [63]. One group is induced at parturition and expressed throughout lactation, as in eutherians. These genes include the milk protein genes encoding α-, β-, and κ-casein (*CSN1*, *CSN2* and *CSN3*) and the α-lactalbumin (*LALBA*) and β-lactoglobulin (*LGB*) whey protein genes. However, the tammar genome lacks additional copies of α- or β-like caseins that are present in monotremes and eutherians (Figure S3 in Additional file 4).

The second group of mammary genes is expressed only during specific phases of lactation. This group includes marsupial-specific milk protein genes such as the late lactation proteins (*LLPA* and *LLPB*) as well as others such as whey acidic protein (*WAP*) [153] that are also found in milk of many eutherians [154] but lacking in humans, goat and ewe [155]. Evidence is now emerging that changes in composition of the major milk proteins and many bioactives [156,157] contribute to a more central role of milk in regulating development and function of the mammary gland [158] to provide protection from bacterial infection in the gut of the young and the mammary gland [159] (A Watt and KR Nicholas, unpublished results) and to deliver specific signals to the young that regulate growth and development of specific tissues such as the gut [160]. There is also a novel putative non-coding RNA (*PTNC-1*) expressed in the mammary gland throughout lactation. *PTNC-1* is derived from a region of the genome that is highly conserved in mammals, suggesting it may have an important functional role [63]. Tammar *ELP* (early lactation protein), originally thought to be marsupial-specific (phase 2A) [63], has a eutherian orthologue, colostrum trypsin inhibitor (*CTI*), which is present in some eutherians but is reduced to a pseudogene in others (EA Pharo, AA De Leo, MB Renfree, and KR Nicholas, unpublished results). The *ELP*/*CTI* gene is flanked by single-copy genes that map to orthologous regions of the genome - strong evidence that *ELP/CTI* evolved from the same ancestral gene. *ELP*/*CTI* has not yet been detected in monotremes. Other marsupial-specific milk protein genes identified include trichosurin and the putative tammar milk proteins *PTMP-1* and *PTMP-2* [63]. Remarkably, the tammar *PTMP-1* gene has been identified in the tammar genome sequence, but does not seem to occur in the genome sequence of the short-tail grey opossum. Thus, *PTMP-1* may be macropodid-specific.

Conclusions

The tammar, a small kangaroo species, is the model Australian marsupial that has played a particularly important role in the study of reproduction, development, immunity and the evolution of the mammalian sex chromosomes. Here, we have presented its genome sequence and associated resources, including transcriptome

sequence data from a range of tissues. Together these data have provided new insights into a host of important gene families. We identified novel tammar-specific, as well as conserved but previously undiscovered, miRNAs that regulate the *HOX* genes, a novel SINE class that is rRNA-derived and a novel class of small RNAs. We show that there has been expansion of several gene families, especially of the MHC and OR genes, that there are features that are of specific importance to marsupials, such as the innovation of genes in lactation and the presence of genomic imprinting in the mammary gland. However, there is high conservation in testicular and ovarian genes, one of which, *DHH*, is only the second mammal-specific gonadal development gene so far identified. The Y chromosome is minute but relatively gene rich and conserved in marsupials. The X chromosome reflects the ancestral mammalian X and perhaps an ancestral stochastic dosage compensation that operates without an X chromosome inactivation center. These initial tammar genome analyses have already provided many unique insights into the evolution of the mammalian genome and highlight the importance of this emerging model system for understanding mammalian biology.

Materials and methods

Materials and methods are briefly described in the body of the paper and extensively in the supplementary methods (Additional file 1).

Data availability

Public database accessions are provided for all raw datasets where they are first mentioned in the text. The latest version of the genome assembly is available in NCBI under the GenBank accession ABQO000000000; Meug_1.1 has accession ABQO010000000; Meug_2.0 has accession ABQO020000000. All versions of the genome assembly are also accessible via the web [161].

Additional material

Additional file 1: Supplementary material. Supplementary materials and methods, results and tables [39,42,46,47,58,74,164-192].

Additional file 2: Figure S1 - comparison of gene sizes in *Monodelphis domestica* and *Macropus eugenii*. One-to-one opossum orthologues of tammar genes located more than 1 kb from the end of a scaffold were downloaded from Ensembl v62. The genomic lengths of the genes are plotted as a scatter plot on the $\log_2$ scale. A 1:1 linear relationship between gene sizes is present for genes less than the average scaffold size, suggesting that no major change in genome size has occurred in genic regions. A trend towards larger genes in opossum with $\log_2$ length > 15 is driven primarily by incompleteness of tammar genes when the gene size is larger than the average scaffold size.

Additional file 3: Figure S2 - analysis of the alignment of transcriptomic reads from different tissues to the tammar genome. **(a)** Proportion of reads that align to unannotated regions, annotated genes, within 2 kb upstream or downstream of a gene, or fail to align to the tammar genome. **(b)** Proportion of mapped reads that align to unannotated regions, annotated genes, or within 2 kb upstream or downstream of a gene in the tammar genome.

Additional file 4: Figure S3 - Comparative analysis of the mammalian casein locus showing the expansion of the casein locus in mammals. Comparison of the casein locus organization in the platypus, tammar, opossum, cattle, mouse and human genomes. Drawn to scale and aligned on the β-casein gene. Genes are represented by a box with a tail arrow pointing in the direction of gene transcription. Gene models for confirmed genes were generated from mammary gland EST data (platypus and tammar) or retrieved from Ensembl (others) when available. The tammar locus is not fully resolved and sequence scaffolds (indicated by black bars and scaffold numbers) have been aligned with the opossum sequence. Gaps in the tammar genome mainly fall in regions containing a repeated transposon type I in the opossum (black arrows), probably compounding the assembly of the tammar genome. Blank boxes represent putative genes based on similarity, grey boxes represent genes with observed expression. Note the close proximity of α- (*CSN1*, csna) and β- (*CSN2*, csnb) casein genes in reverse orientation on the left and the expansion of the region between β- and kappa- (*CSN3*, csnk) casein on the right. Except for β-casein, all genes are transcribed from left to right. In monotremes, a recent duplication of *CSN2* has led to *CSN2b*, whereas in eutherians, an ancient duplication produced *CSN1S2*, which has been duplicated in some species to produce *CSN1S2b*, now a pseudogene in human but not in mouse. In the marsupial locus, there is no casein duplication and the spacing region contains several copies of an invading repetitive element (black arrows), suggesting active rearrangement of this region in the ancient marsupial lineage, probably resulting in the deletion of a putative ancient casein duplicate in the area.

Abbreviations

BAC: bacterial artificial chromosome; BCM-HGSC: Baylor College of Medicine Human Genome Sequencing Center; bp: base pair; crasiRNA: centromere repeat-associated short interacting RNA; DHH: Desert hedgehog; DMR: differentially methylated region; EST: expressed sequence tag; GH: growth hormone; IFITM: interferon inducible transmembrane protein; IGF: insulin-like growth factor; KERV: kangaroo endogenous retrovirus; LINE: long interspersed nuclear element; LTR: long terminal repeat; MHC: major histocompatibility complex; miRNA: microRNA; NOR: nucleolar organizing region; OR: olfactory receptor; PGC: primordial germ cell; piRNA: Piwi-interacting RNA; SINE: short interspersed nuclear element; VNO: vomeronasal organ; VNR: vomeronasal receptor; WGS: whole-genome shotgun.

Acknowledgements

This study was supported by the State Government of Victoria, the National Human Genome Research Institute of the National Institutes of Health (grant number U54-HG003273), the Australian Genome Research Facility (AGRF), the Jack Brockhoff Foundation, Applied Biosystems, the Australian Research Council Centre of Excellence in Kangaroo Genomics, the National Science Foundation and the Center for Applied Genetics and Technology at University of Connecticut (UConn). This work was supported in part by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas 'Comparative Genomics' from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and by the Director's Special Grant from RIKEN-GSC. The Baylor College of Medicine Human genome Sequencing Center (HGSC) acknowledges the following production staff of the HGSC: Jennifer Hume, John Lopez, Kashif Hirani, Lingling Pu, Marvin D Dao, Mimi N Chandrabose, Ngoc B Nguyen, Ramatu A Gabisi, Rita A Wright, Sandra Hines, Yih-Shin Liu, Ugonna Anosike, Tony Attaway, Dilrukshi Bandaranaike, Ashton Bell, Blake Beltran, Carla Bickham, Temika Caleb, Kelvin Carter, Joseph Chacko, Alejandra Chavez, Hau-Seng Chu, Raynard Cockrell, Mary Louise Davila, Latarsha Davy-Carroll, Shawn Denson, Victor Ebong, Veronica Espinosa, Sonia Fernandez, Pushpa Fernando, Nicole Flagg, Jason Ganer, Ricardo Garcia III, Toni Garner, Tiffany Garrett, Brandy Harbes, Ebere Hawkins, Barbara Hollins, Bennie Johnson, Johann Kalu, Haika Kisamo, Michael Lago, Liza Lago, Chuan-Yar Lai, Thanh-Kim Le, Fitzherbert Legall III, Semethia Lemon, Renita Madu, Kevin Malloy, Evangelina Martinez, Christian Mercado, Iracema Mercado, Mala Munidasa, Phong Nguyen, Ogechi Nwaokelemeh, Melissa Obregon,

Chibueze Onwere, Andrea Parra, Agapito Perez, Yolanda Perez, Christopher Pham, Eltrick Primus, Maria Puazo, Juana Quiroz, Eric Rachlin, Marcos Ruiz, Brian Schneider, Denard Simmons, Ida Sisson, Rosenie Thelus, Nicole Thomas, Rachel Thorn, Reshaunda Thornton, Zulma Trejos, Kamran Usmani, Davian Walker, Keqing Wang, Suzhen Wang, Courtney White, Aneisa Williams and Jerrell Woodworth.

Author details

[1]The Australian Research Council Centre of Excellence in Kangaroo Genomics, Australia. [2]Department of Zoology, The University of Melbourne, Melbourne, Victoria 3010, Australia. [3]Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia. [4]Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Victoria 3010, Australia. [5]Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia. [6]Department of Molecular and Cell Biology, Center for Applied Genetics and Technology, University of Connecticut, Storrs, CT 06269, USA. [7]Faculty of Veterinary Science, University of Sydney, Sydney, NSW 2006, Australia. [8]Department of Veterinary Medicine, University of Cambridge, Madingley Rd, Cambridge, CB3 0ES, UK. [9]Institute for Technology Research and Innovation, Deakin University, Geelong, Victoria, 3214, Australia. [10]RIKEN Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [11]School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia. [12]Department of Microbiology and Immunology, The University of Melbourne, Melbourne, Victoria 3010, Australia. [13]Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, Berlin 10315, Germany. [14]Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge, CB22 3AT, UK. [15]Department of Molecular Genetics, German Institute of Human Nutrition, Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany. [16]Department of Medical Biology, The University of Melbourne, Melbourne, Victoria 3010, Australia. [17]Biosciences Research Division, Department of Primary Industries, Victoria, 1 Park Drive, Bundoora 3083, Australia. [18]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. [19]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. [20]Department of Cell Biology, University of Massachusetts Medical School, Worcester, MA 01655, USA. [21]Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8560, Japan. [22]National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan. [23]Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269, USA. [24]Human Genome Sequencing Center, Department of Molecular and Human Genetics Baylor College of Medicine, Houston, TX 77030, USA. [25]Australian Genome Research Facility, Melbourne, Victoria, 3052 and the University of Queensland, St Lucia, Queensland 4072, Australia. [26]Westmead Institute for Cancer Research, University of Sydney, Westmead, New South Wales 2145, Australia. [27]National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. [28]Department of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, NSW 2052, Australia.

Authors' contributions

All authors were members of the tammar wallaby genome sequencing consortium. Authors contributed to sequencing, assembly, analysis, experiments or writing as follows (team leaders are in bold). Joint lead authors: **MBR, ATP**. Principal investigators: JAMG, SMF, RAG, DWC, MBR, TPS, AF. Tammar genome size: **AJP, WR**, RJO'N, KCW, MFS. Physical and linkage mapping: **JED**, CW, FWN, KRZ, JAMG. Highly divergent or tammar-specific transcripts: **AJP**, AF, TH, YK, HY, MBR. Transcriptome: **ATP, AF**, AJP, MBR, YK, Z-PF, YSuz, SS, AT, YSak, SK, YN, ST. Genebuild: **SMJS**, SF. Sequence conservation and gene family expansion: **PF, ESWW**, KBl, JH. GC content: **WR, RJO'N**, JL. Analysis of Repeats and small RNAs: **RJO'N**, JL, DMC. Immunity: **KB**, ESWW, HVS, JED, MBR, KRS, CW, JW, BGC, ATP. Sex chromosomes: **PDW, JAMG**, JED. X chromosome inactivation: **JAMG, PDW**, JED, SAN. Reproductive genes: **AJP, MBR, GS**, DH, WO'H, YH. Developmental genes: **MBR, AJP**, GS, SRF, HY, K-YC, BRM, RJO'N, JL. Genomic imprinting: **MBR, AJP, JAMG**, GS, JS, SS, TAH. Olfaction: **MBR, JAMG**, SRF, AM, MLD, NYS, GS. Lactation: **CML, KRN**, EAP. Additional bioinformatics contributions: AH, BL, KAM, MJW. Australian Genome Research Facility Sanger sequencing: **SMF**, EK, AMG, PW, AM, JE, CT, DT, AS, LY, TL, MH-R, AH, JD, DW, SW, YSun. HGSC leadership: KCW, DMM, RAG. HGSC Sanger production: SNJ, LRL, MBM, GOO, SJR, JS, LN, AC, GF, CLK, HHD, VJ. HGSC SOLiD production team: HHD, CJ. HGSC 454 production: CLK, FL, RT. HGSC genome assembly and analysis: LC, JD, YL, JYS, X-ZS, GW, KCW. UConn sequencing and assembly improvement (Meug_2): **RJO'N, AJP**, JL, TH, IM. Senior authors: SMF, JAMG, RJO'N, AJP, KCW. Senior authors contributed equally and principal investigators contributed equally.

Received: 23 May 2011 Revised: 22 July 2011 Published: 20 August 2011

References

1. Drake-Brockman H: *Voyage to Disaster* Sydney: Angus and Robertson; 1963.
2. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: **The delayed rise of present-day mammals.** *Nature* 2007, **446**:507-512.
3. Luo Z-X, Ji Q, Wible JR, Yuan C-X: **An Early Cretaceous tribosphenic mammal and metatherian evolution.** *Science* 2003, **302**:1934-1940.
4. Nilsson MA, Churakov G, Sommer M, Tran NV, Zemann A, Brosius J, Schmitz J: **Tracking marsupial evolution using archaic genomic retroposon insertions.** *PLoS Biol* 2010, **8**:e1000436.
5. Meredith RW, Westerman M, Springer MS: **A phylogeny and timescale for the living genera of kangaroos and kin (Macropodiformes: Marsupialia) based on nuclear DNA sequences.** *Aust J Zool* 2008, **56**:395-410.
6. Deakin JE, Waters PD, Graves JAM: *Marsupial Genetics and Genomics* Dordrecht, Heidleberg, London, New York: Springer; 2010.
7. Hickford D, Frankenberg S, Renfree MB: **The tammar wallaby, *Macropus eugenii*: a model kangaroo for the study of developmental and reproductive biology.** In *Emerging Model Organisms: A Laboratory Manual. Volume 2.* Edited by: Crotty DA, Gann A. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2010:449-494.
8. Hume ID: *Marsupial Nutrition* New York: Cambridge University Press; 1999.
9. Lee AK, Cockburn A: *Evolutionary Ecology of Marsupials* New York, Cambridge: Cambridge University Press; 1985.
10. Richardson SJ, Aldred AR, Leng SL, Renfree MB, Hulbert AJ, Schreiber G: **Developmental profile of thyroid hormone distributor proteins in a marsupial, the tammar wallaby *Macropus eugenii*.** *Gen Comp Endocrinol* 2002, **125**:92-103.
11. Tyndale-Biscoe CH, Janssens PA: *The Developing Marsupial - Models for Biomedical Research* Berlin, Germany: Springer-Verlag; 1988.
12. Tyndale-Biscoe CH, Renfree MB: *Reproductive Physiology of Marsupials* Cambridge, UK: Cambridge University Press; 1987.
13. Renfree MB: **Endocrinology of pregnancy, parturition and lactation in marsupials.** In *Marshall's Physiology of Reproduction Pregnancy and Lactation. Volume 3*.. Fourth edition. Edited by: Lamming GE. London, UK: Chapman 1994:677-766.
14. Renfree MB, Shaw G: **Diapause.** *Annu Rev Physiol* 2000, **62**:353-375.
15. Tyndale-Biscoe CH: *Life of Marsupials* Collingwood, VIC, Australia: CSIRO Publishing; 2005.
16. Freyer C, Zeller U, Renfree MB: **The marsupial placenta: a phylogenetic analysis.** *J Exp Zool A Comp Exp Biol* 2003, **299**:59-77.
17. Renfree MB: **Maternal recognition of pregnancy in marsupials.** *Rev Reprod* 2000, **5**:6-11.
18. Renfree MB: **Marsupials: placental mammals with a difference.** *Placenta* 2010, **31(Suppl)**:S21-26.
19. Mark RF, Marotte LR: **Australian marsupials as models for the developing mammalian visual system.** *Trends Neurosci* 1992, **15**:51-57.
20. Pask AJ, Renfree MB: **Sex determining genes and sexual differentiation in a marsupial.** *J Exp Zool* 2001, **290**:586-596.
21. Renfree MB, Harry JL, Shaw G: **The marsupial male: a role model for sexual development.** *Phil Trans R Soc Lond Series B Biol Sci* 1995, **350**:243-251.
22. Renfree MB, O WS, Short RV, Shaw G: **Sexual differentiation of the urogenital system of the fetal and neonatal tammar wallaby, *Macropus eugenii*.** *Anat Embryol (Berl)* 1996, **194**:111-134.
23. Reynolds ML, Cavanagh ME, Dziegielewska KM, Hinds LA, Saunders NR, Tyndale-Biscoe CH: **Postnatal development of the telencephalon of the tammar wallaby (*Macropus eugenii*). An accessible model of neocortical differentiation.** *Anat Embryol (Berl)* 1985, **173**:81-94.
24. Renfree MB, Wilson JD, Shaw G: **The hormonal control of sexual development.** *Novartis Found Symp* 2002, **244**:136-152, discussion 152-136, 203-136, 253-137.

25. Green B, Merchant JC: **The composition of marsupial milk.** In *The Developing Marsupial Models for Biomedical Research.* Edited by: Tyndale-Bisco CH, Janssens PA. Berlin, Germany: Springer-Verlag; 1988:41-54.
26. Nicholas K, Simpson K, Wilson M, Trott J, Shaw D: **The tammar wallaby: a model to study putative autocrine-induced changes in milk composition.** *J Mammary Gland Biol Neoplasia* 1997, **2**:299-310.
27. Schneider LK: **Marsupial chromosomes, cell cycles and cytogenetics.** In *The Biology of Marsupials.* Edited by: Hunsaker D. New York: Academic Press; 1977:51-93.
28. Graves JAM: **DNA synthesis in chromosomes of cultured leucocytes from two marsupial species.** *Exp Cell Res* 1967, **46**:37-57.
29. Yao KT: **Haploid cells in rat kangaroo corneal endothelium cultures and x-ray-induced chromosome aberrations.** *Genetics* 1971, **67**:399-409.
30. O'Neill RJ, Eldridge MD, Toder R, Ferguson-Smith MA, O'Brien PC, Graves JAM: **Chromosome evolution in kangaroos (Marsupialia: Macropodidae): cross species chromosome painting between the tammar wallaby and rock wallaby spp. with the 2n = 22 ancestral macropodid karyotype.** *Genome* 1999, **42**:525-530.
31. Hayman DL, Martin PG: **Sex chromosome mosaicism in the marsupial genera *Isoodon* and *Perameles*.** *Genetics* 1965, **52**:1201-1206.
32. Watson CM, Margan SH, Johnston PG: **Sex-chromosome elimination in the bandicoot *Isoodon macrourus* using Y-linked markers.** *Cytogenet Cell Genet* 1998, **81**:54-59.
33. Hayman DL, Martin PG: *Animal Cytogenetics. Mammalia I: Monotremata and Marsupialia* Berlin: Gebrud er Borntraeger; 1974.
34. Rofe R, Hayman D: **G-banding evidence for a conserved complement in the Marsupialia.** *Cytogenet Cell Genet* 1985, **39**:40-50.
35. Graves JAM: **The origin and function of the mammalian Y chromosome and Y-borne genes - an evolving understanding.** *Bioessays* 1995, **17**:311-320.
36. Glas R, Leo A, Delbridge M, Reid K, Ferguson-smith M, O'Brien P, Westerman M, Graves JAM: **Chromosome painting in marsupials: genome conservation in the kangaroo family.** *Chromosome Res* 1999, **7**:167-176.
37. Deakin JE, Koina E, Waters PD, Doherty R, Patel VS, Delbridge ML, Dobson B, Fong J, Hu Y, van den Hurk C, Pask AJ, Shaw G, Smith C, Thompson K, Wakefield MJ, Yu H, Renfree MB, Graves JAM: **Physical map of two tammar wallaby chromosomes: a strategy for mapping in non-model mammals.** *Chromosome Res* 2008, **16**:1159-1175.
38. Rens W, O'Brien PC, Yang F, Graves JAM, Ferguson-Smith MA: **Karyotype relationships between four distantly related marsupials revealed by reciprocal chromosome painting.** *Chromosome Res* 1999, **7**:461-474.
39. Rens W, O'Brien PC, Yang F, Solanky N, Perelman P, Graphodatsky AS, Ferguson MW, Svartman M, De Leo AA, Graves JAM, Ferguson-Smith MA: **Karyotype relationships between distantly related marsupials from South America and Australia.** *Chromosome Res* 2001, **9**:301-308.
40. Alsop AE, Miethke P, Rofe R, Koina E, Sankovic N, Deakin JE, Haines H, Rapkins RW, Graves JAM: **Characterizing the chromosomes of the Australian model marsupial *Macropus eugenii* (tammar wallaby).** *Chromosome Res* 2005, **13**:627-636.
41. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SM, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Greally JM, Gu W, *et al*: **Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences.** *Nature* 2007, **447**:167-177.
42. Pop M, Kosack DS, Salzberg SL: **Hierarchical scaffolding with Bambus.** *Genome Res* 2004, **14**:149-159.
43. Sundquist A, Ronaghi M, Tang H, Pevzner P, Batzoglou S: **Whole-genome sequencing and assembly with high-throughput, short-read technologies.** *PLoS One* 2007, **2**:e484.
44. Wang C, Deakin JE, Rens W, Zenger KR, Belov K, Graves JAM, Nicholas FW: **An integrated tammar wallaby map and its use in creating a virtual tammar wallaby genome map.** *BMC Genomics* .
45. Heider TN, Lindsay J, Wang C, O'Neill RJ, Pask AJ: **Enhancing genome assemblies by integrating non-sequence based data.** *BMC Proc* 2011, **5(Suppl 2)**:S7.
46. Wilhelm J, Pingoud A, Hahn M: **Validation of an algorithm for automatic quantification of nucleic acid copy numbers by real-time polymerase chain reaction.** *Anal Biochem* 2003, **317**:218-225.
47. Havlak P, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Weinstock GM, Gibbs RA: **The Atlas genome assembly system.** *Genome Res* 2004, **14**:721-732.
48. Carone DM, Longo MS, Ferreri GC, Hall L, Harris M, Shook N, Bulazel KV, Carone BR, Obergfell C, O'Neill MJ, O'Neill RJ: **A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres.** *Chromosoma* 2009, **118**:113-125.
49. Wang C, Webley L, Wei K, Wakefield MJ, Patel HR, Deakin JE, Alsop AE, Graves JAM, Cooper DW, Nicholas FW, Zenger KR: **A second-generation anchored genetic linkage map of the tammar wallaby (*Macropus eugenii*).** *BMC Genet* 2011, **12**:72.
50. Rens W, O'Brien PC, Fairclough H, Harman L, Graves JAM, Ferguson-Smith MA: **Reversal and convergence in marsupial chromosome evolution.** *Cytogenet Genome Res* 2003, **102**:282-290.
51. Siddle HV, Deakin JE, Coggill P, Whilming L, Harrow J, Kaufman J, Beck S, Belov K: **The tammar wallaby major histocompatibility complex shows evidence of past genomic instability.** *BMC Genomics* 2011, **12**:421.
52. Hu Y, Yu H, Shaw G, Pask AJ, Renfree MB: **Kallmann syndrome 1 gene is expressed in the marsupial gonad.** *Biol Reprod* 2011, **84**:595-603.
53. Hu Y, Yu H, Shaw G, Renfree MB, Pask AJ: **Differential roles of TGIF family genes in mammalian reproduction.** *BMC Dev Biol* .
54. Al Nadaf S, Waters PD, Koina E, Deakin JE, Jordan KS, Graves JAM: **Activity map of the tammar X chromosome shows that marsupial X inactivation is incomplete and escape is stochastic.** *Genome Biol* 2010, **11**:R122.
55. Chaumeil J, Waters PD, Koina E, Gilbert C, Robinson TJ, Graves JAM: **Evolution from XIST-independent to XIST-controlled X-chromosome inactivation: epigenetic modifications in distantly related mammals.** *PLoS One* 2011, **6**:e19040.
56. Rens W, Wallduck MS, Lovell FL, Ferguson-Smith MA, Ferguson-Smith AC: **Epigenetic modifications on X chromosomes in marsupial and monotreme mammals and implications for evolution of dosage compensation.** *Proc Natl Acad Sci USA* 2010, **107**:17657-17662.
57. **RepeatMasker Open 3.0.** [http://www.repeatmasker.org/RMDownload.html].
58. **RepeatModeler Open 1.0.** [http://www.repeatmasker.org/RepeatModeler.html].
59. Bulazel KV, Ferreri GC, Eldridge MD, O'Neill RJ: **Species-specific shifts in centromere sequence composition are coincident with breakpoint reuse in karyotypically divergent lineages.** *Genome Biol* 2007, **8**:R170.
60. Carone DM, O'Neill RJ: **Marsupial centromeres and telomeres: dynamic chromosome domains.** In *Marsupial Genetics and Genomics.* Edited by: Deakin JE, Waters PD, Graves JAM. Dordrecht, Heidleberg, London, New York: Springer; 2010:55-74.
61. Allshire RC, Karpen GH: **Epigenetic regulation of centromeric chromatin: old dogs, new tricks?** *Nat Rev Genet* 2008, **9**:923-937.
62. Sullivan BA, Blower MD, Karpen GH: **Determining centromere identity: cyclical stories and forking paths.** *Nat Rev Genet* 2001, **2**:584-596.
63. Lefèvre CM, Digby MR, Whitley JC, Strahm Y, Nicholas KR: **Lactation transcriptomics in the Australian marsupial, *Macropus eugenii*: transcript sequencing and quantification.** *BMC Genomics* 2007, **8**:417.
64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
65. Min XJ, Butler G, Storms R, Tsang A: **OrfPredictor: predicting protein-coding regions in EST-derived sequences.** *Nucleic Acids Res* 2005, **33**: W677-680.
66. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A: **The Pfam protein families database.** *Nucleic Acids Res* 2010, **38**:D211-222.
67. Aravin AA, Hannon GJ, Brennecke J: **The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race.** *Science* 2007, **318**:761-764.
68. Blumenstiel JP, Hartl DL: **Evidence for maternally transmitted small interfering RNA in the repression of transposition in *Drosophila virilis*.** *Proc Natl Acad Sci USA* 2005, **102**:15965-15970.
69. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ: **Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*.** *Cell* 2007, **128**:1089-1103.
70. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ: **An epigenetic role for maternally inherited piRNAs in transposon silencing.** *Science* 2008, **322**:1387-1392.
71. Griffiths-Jones S: **The microRNA Registry.** *Nucleic Acids Res* 2004, **32**: D109-111.

72. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006, **34**:D140-144.
73. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Res* 2008, **36**:D154-158.
74. Kozomara A, Griffiths-Jones S: **miRBase: integrating microRNA annotation and deep-sequencing data.** *Nucleic Acids Res* 2011, **39**:D152-157.
75. Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, Minami N, Imai H: **Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes.** *Genes Dev* 2006, **20**:1732-1743.
76. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H: **Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes.** *Nature* 2008, **453**:539-543.
77. Close R, Lowry P: **Hybrids in marsupial research.** *Aust J Zool* 1989, **37**:259-267.
78. O'Neill RJ, Eldridge MD, Graves JAM: **Chromosome heterozygosity and *de novo* chromosome rearrangements in mammalian interspecies hybrids.** *Mamm Genome* 2001, **12**:256-259.
79. O'Neill RJ, O'Neill MJ, Graves JAM: **Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid.** *Nature* 1998, **393**:68-72.
80. Ferreri GC, Brown JD, Obergfell C, Jue N, Finn CE, O'Neill MJ, O'Neill RJ: **Recent amplification of the kangaroo endogenous retrovirus, KERV, limited to the centromere.** *J Virol* 2011, **85**:4761-4771.
81. Deakin JE, Siddle HV, Cross JG, Belov K, Graves JAM: **Class I genes have split from the MHC in the tammar wallaby.** *Cytogenet Genome Res* 2007, **116**:205-211.
82. Siddle HV, Deakin JE, Coggill P, Hart E, Cheng Y, Wong ESW, Harrow J, Beck S, Belov K: **MHC-linked and un-linked class I genes in the wallaby.** *BMC Genomics* 2009, **10**:310.
83. Wong ESW, Papenfuss AT, Heger A, Hsu AL, Ponting CP, Miller RD, Fenelon JC, Renfree MB, Gibbs RA, Belov K: **Transcriptomic analysis supports similar functional roles for the two thymuses of the tammar wallaby.** *BMC Genomics* 2011, **12**:420.
84. Wang J, Wong ESW, Whitley JC, Li J, Stringer JM, Short KR, Renfree MB, Belov K, Cocks BG: **Antibiotic-resistant pathogens killed by ancient peptides from the genomes of Australian mammals.** *PLoS One* 2011, **6**: e24030.
85. Wong ESW, Papenfuss AT, Belov K: **Immunome database for marsupials and monotremes.** *BMC Immunol* 2011, **12**:48.
86. Graves JAM: **Sex chromosome specialization and degeneration in mammals.** *Cell* 2006, **124**:901-914.
87. Raudsepp T, Lee EJ, Kata SR, Brinkmeyer C, Mickelson JR, Skow LC, Womack JE, Chowdhary BP: **Exceptional conservation of horse-human gene order on X chromosome revealed by high-resolution radiation hybrid mapping.** *Proc Natl Acad Sci USA* 2004, **101**:2386-2391.
88. Rodríguez Delgado CL, Waters PD, Gilbert C, Robinson TJ, Graves JAM: **Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years.** *Chromosome Res* 2009, **17**:917-926.
89. Graves JAM, Wakefield MJ, Toder R: **The origin and evolution of the pseudoautosomal regions of human sex chromosomes.** *Hum Mol Genet* 1998, **7**:1991-1996.
90. Toder R, Wienberg J, Voullaire L, O'Brien PC, Maccarone P, Graves JAM: **Shared DNA sequences between the X and Y chromosomes in the tammar wallaby - evidence for independent additions to eutherian and marsupial sex chromosomes.** *Chromosoma* 1997, **106**:94-98.
91. Waters PD, Wallis MC, Graves JAM: **Mammalian sex - Origin and evolution of the Y chromosome and SRY.** *Semin Cell Dev Biol* 2007, **18**:389-400.
92. Delbridge ML, Harry JL, Toder R, O'Neill RJ, Ma K, Chandley AC, Graves JAM: **A human candidate spermatogenesis gene, *RBM1*, is conserved and amplified on the marsupial Y chromosome.** *Nat Genet* 1997, **15**:131-136.
93. Pask AJ, Renfree MB, Graves JAM: **The human sex-reversing *ATRX* gene has a homologue on the marsupial Y chromosome, *ATRY*: implications for the evolution of mammalian sex determination.** *Proc Natl Acad Sci USA* 2000, **97**:13198-13202.
94. Cooper DW, Johnston PG, Watson JM, Graves JAM: **X-inactivation in marsupials and monotremes.** *Semin Dev Biol* 1993, **4**:117-128.
95. Duret L, Chureau C, Samain S, Weissenbach J, Avner P: **The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene.** *Science* 2006, **312**:1653-1655.
96. Hore TA, Koina E, Wakefield MJ, Graves JAM: **The region homologous to the X-chromosome inactivation centre has been disrupted in marsupial and monotreme mammals.** *Chromosome Res* 2007, **15**:147-161.
97. Mahadevaiah SK, Royo H, VandeBerg JL, McCarrey JR, Mackay S, Turner JM: **Key features of the X inactivation process are conserved between marsupials and eutherians.** *Curr Biol* 2009, **19**:1478-1484.
98. Pask AJ, Renfree MB: **Molecular regulation of marsupial reproduction and development.** In *Marsupial Genetics and Genomics*. Edited by: Deakin JE, Waters PD, Graves JAM. Dordrecht, Heidleberg, London, New York: Springer; 2010:285-316.
99. Chung JW, Pask AJ, Yu H, Renfree MB: **Fibroblast growth factor-9 in marsupial testicular development.** *Sex Dev* 2011, **5**:131-140.
100. Huynh K, Renfree MB, Graves JAM, Pask AJ: ***ATRX* has a critical and conserved role in mammalian sexual differentiation.** *BMC Dev Biol* 2011, **11**:39.
101. Goodrich LV, Johnson RL, Milenkovic L, McMahon JA, Scott MP: **Conservation of the hedgehog/patched signaling pathway from flies to mice: induction of a mouse patched gene by Hedgehog.** *Genes Dev* 1996, **10**:301-312.
102. Hickford DE, Frankenberg S, Pask AJ, Shaw G, Renfree MB: **DDX4 (VASA) is conserved in germ cell development in marsupials and monotremes.** *Biol Reprod* 2011.
103. Gustafson EA, Wessel GM: **Vasa genes: emerging roles in the germ line and in multipotent cells.** *Bioessays* 2010, **32**:626-637.
104. Payer B, Saitou M, Barton SC, Thresher R, Dixon JP, Zahn D, Colledge WH, Carlton MB, Nakano T, Surani MA: ***stella* is a maternal effect gene required for normal early development in mice.** *Curr Biol* 2003, **13**:2110-2117.
105. Clark AT, Rodriguez RT, Bodnar MS, Abeyta MJ, Cedars MI, Turek PJ, Firpo MT, Reijo Pera RA: **Human *STELLAR*, *NANOG*, and *GDF3* genes are expressed in pluripotent cells and map to chromosome 12p13, a hotspot for teratocarcinoma.** *Stem Cells* 2004, **22**:169-179.
106. Saitou M, Barton SC, Surani MA: **A molecular programme for the specification of germ cell fate in mice.** *Nature* 2002, **418**:293-300.
107. Tanaka SS, Yamaguchi YL, Tsoi B, Lickert H, Tam PP: **IFITM/Mil/fragilis family proteins IFITM1 and IFITM3 play distinct roles in mouse primordial germ cell homing and repulsion.** *Dev Cell* 2005, **9**:745-756.
108. Tanaka SS, Matsui Y: **Developmentally regulated expression of *mil-1* and *mil-2*, mouse interferon-induced transmembrane protein like genes, during formation and differentiation of primordial germ cells.** *Mech Dev* 2002, **119(Suppl 1)**:S261-267.
109. Lange UC, Adams DJ, Lee C, Barton S, Schneider R, Bradley A, Surani MA: **Normal germ line establishment in mice carrying a deletion of the *Ifitm/Fragilis* gene family cluster.** *Mol Cell Biol* 2008, **28**:4688-4696.
110. Graves JAM, Gecz J, Hameister H: **Evolution of the human X - a smart and sexy chromosome that controls speciation and development.** *Cytogenet Genome Res* 2002, **99**:141-145.
111. Hu YQ, Yu HS, Pask AJ, O'Brien DA, Shaw G, Renfree MB: **A-kinase anchoring protein 4 has a conserved role in mammalian spermatogenesis.** *Reproduction* 2009, **137**:645-653.
112. Selwood L: **Development of early cell lineages in marsupial embryos: an overview.** *Reprod Fertil Dev* 1994, **6**:507-527.
113. Frankenberg SR, Pask AJ, Renfree MB: **The evolution of class V POU domain transcription factors in vertebrates and their characterisation in a marsupial.** *Dev Biol* 2010, **337**:162-170.
114. Niwa H, Sekita Y, Tsend-Ayush E, Grutzner F: **Platypus *Pou5f1* reveals the first steps in the evolution of trophectoderm differentiation and pluripotency in mammals.** *Evol Dev* 2008, **10**:671-682.
115. Shen MM: **Nodal signaling: developmental roles and regulation.** *Development* 2007, **134**:1023-1034.
116. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.** *Cell* 2007, **129**:1311-1323.
117. Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB: **Annotating non-coding regions of the genome.** *Nat Rev Genet* 2010, **11**:559-571.
118. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY: **Long noncoding RNA as modular scaffold of histone modification complexes.** *Science* 2010, **329**:689-693.
119. Zakany J, Duboule D: **The role of Hox genes during vertebrate limb development.** *Curr Opin Genet Dev* 2007, **17**:359-366.

120. Chau YM, Pando S, Taylor HS: **HOXA11 silencing and endogenous HOXA11 antisense ribonucleic acid in the uterine endometrium.** *J Clin Endocrinol Metab* 2002, **87**:2674-2680.
121. Woo CJ, Kingston RE: ***HOTAIR* lifts noncoding RNAs to new levels.** *Cell* 2007, **129**:1257-1259.
122. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY: **Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis.** *Nature* 2010, **464**:1071-1076.
123. Schneider NY, Fletcher TP, Shaw G, Renfree MB: **The vomeronasal organ of the tammar wallaby.** *J Anat* 2008, **213**:93-105.
124. Chen CH, Cretekos CJ, Rasweiler JJt, Behringer RR: ***Hoxd13* expression in the developing limbs of the short-tailed fruit bat, *Carollia perspicillata*.** *Evol Dev* 2005, **7**:130-141.
125. Nelson CE, Morgan BA, Burke AC, Laufer E, DiMambro E, Murtaugh LC, Gonzales E, Tessarollo L, Parada LF, Tabin C: **Analysis of *Hox* gene expression in the chick limb bud.** *Development* 1996, **122**:1449-1466.
126. Menzies BR, Shaw G, Fletcher TP, Pask AJ, Renfree MB: **Exon 3 of the growth hormone receptor (GH-R) is specific to eutherian mammals.** *Mol Cell Endocrinol* 2008, **296**:64-68.
127. Menzies BR, Pask AJ, Renfree MB: **Placental expression of pituitary hormones may have an ancient ancestry.** *EVO DEVO* .
128. Menzies BR, Shaw G, Fletcher TP, Renfree MB: **Early onset of ghrelin production in a marsupial.** *Mol Cell Endocrinol* 2009, **299**:266-273.
129. Rapkins RW, Hore TA, Smithwick M, Ager E, Pask AJ, Renfree MB, Kohn M, Hameister H, Nicholls RD, Deakin JE, Graves JAM: **Recent assembly of an imprinted domain from non-imprinted components.** *PLoS Genet* 2006, **2**:e182.
130. Edwards CA, Mungall AJ, Matthews L, Ryder E, Gray DJ, Pask AJ, Shaw G, Graves JAM, Rogers J, Dunham I, Renfree MB, Ferguson-Smith AC: **The evolution of the *DLK1-DIO3* imprinted domain in mammals.** *PLoS Biol* 2008, **6**:e135.
131. Edwards CA, Rens W, Clarke O, Mungall AJ, Hore TA, Graves JAM, Dunham I, Ferguson-Smith AC, Ferguson-Smith MA: **The evolution of imprinting: chromosomal mapping of orthologues of mammalian imprinted domains in monotreme and marsupial mammals.** *BMC Evol Biol* 2007, **7**:157.
132. Hore TA, Renfree MB, Pask AJ, Graves JAM: **The evolution of genomic imprinting - a marsupial perspective.** In *Marsupial Genetics and Genomics.* Edited by: Deakin JE, Waters PD, Graves JAM. Dordrecht, Heidleberg, London, New York: Springer; 2010:233-257.
133. Renfree MB, Hore TA, Shaw G, Graves JAM, Pask AJ: **Evolution of genomic imprinting: insights from marsupials and monotremes.** *Annu Rev Genomics Hum Genet* 2009, **10**:241-262.
134. Suzuki S, Shaw G, Kaneko-Ishino T, Ishino F, Renfree MB: **Characterisation of the marsupial *PHLDA2* reveals eutherian specific acquisition of imprinting.** *BMC Evol Biol* 2011, **11**:244.
135. Smits G, Mungall AJ, Griffiths-Jones S, Smith P, Beury D, Matthews L, Rogers J, Pask AJ, Shaw G, VandeBerg JL, McCarrey JR, Renfree MB, Reik W, Dunham I: **Conservation of the *H19* noncoding RNA and *H19-IGF2* imprinting mechanism in therians.** *Nat Genet* 2008, **40**:971-976.
136. Sekita Y, Wagatsuma H, Nakamura K, Ono R, Kagami M, Wakisaka N, Hino T, Suzuki-Migishima R, Kohda T, Ogura A, Ogata T, Yokoyama M, Kaneko-Ishino T, Ishino F: **Role of retrotransposon-derived imprinted gene, *Rtl1*, in the feto-maternal interface of mouse placenta.** *Nat Genet* 2008, **40**:243-248.
137. Suzuki S, Ono R, Narita T, Pask AJ, Shaw G, Wang C, Kohda T, Alsop AE, Graves JAM, Kohara Y, Ishino F, Renfree MB, Kaneko-Ishino T: **Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting.** *PLoS Genet* 2007, **3**:e55.
138. Ono R, Kuroki Y, Narusea M, Ishii M, Iwasaki S, Toyoda A, Fujiyama A, Shaw G, Renfree MB, Kaneko-Ishino T, Ishino F: **Identification of tammar wallaby *SIRH12*, derived from a marsupial-specific retrotransposition event.** *DNA Res* .
139. Hore TA, Deakin JE, Graves JAM: **The evolution of epigenetic regulators *CTCF* and *BORIS/CTCFL* in amniotes.** *PLoS Genet* 2008, **4**:e1000169.
140. Keverne EB, Curley JP: **Epigenetics, brain evolution and behaviour.** *Front Neuroendocrinol* 2008, **29**:398-412.
141. Wang X, Sun Q, McGrath SD, Mardis ER, Soloway PD, Clark AG: **Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain.** *PLoS One* 2008, **3**:e3839.
142. Gregg C, Zhang J, Butler JE, Haig D, Dulac C: **Sex-specific parent-of-origin allelic expression in the mouse brain.** *Science* 2010, **329**:682-685.
143. Frankenberg SR, Schneider NY, Fletcher TP, Shaw G, Renfree MB: **Identification of two distinct genes at the vertebrate *TRPC2* locus and their characterisation in a marsupial and a monotreme.** *BMC Mol Biol* 2011, **12**:39.
144. Caldecott KW: **XRCC1 and DNA strand break repair.** *DNA Repair (Amst)* 2003, **2**:955-969.
145. Takigami S, Mori Y, Tanioka Y, Ichikawa M: **Morphological evidence for two types of mammalian vomeronasal system.** *Chem Senses* 2004, **29**:301-310.
146. Schneider NY, Fletcher TP, Shaw G, Renfree MB: **The olfactory system of the tammar wallaby is developed at birth and directs the neonate to its mother's pouch odours.** *Reproduction* 2009, **138**:849-857.
147. Delbridge ML, Mohammadi A, Graves JAM: **The olfactory receptor gene family of marsupials.** In *Marsupial Genetics and Genomics.* Edited by: Deakin JE, Waters PD, Graves JAM. Dordrecht, Heidleberg, London, New York: Springer Netherlands; 2010:435-456.
148. Lefèvre CM, Sharp JA, Nicholas KR: **Evolution of lactation: ancient origin and extreme adaptations of the lactation system.** *Annu Rev Genomics Hum Genet* 2010, **11**:219-238.
149. Oftedal OT: **The mammary gland and its origin during synapsid evolution.** *J Mamm Gland Biol Neoplasia* 2002, **7**:225-252.
150. Green B, Griffiths M, Leckie RM: **Qualitative and quantitative changes in milk fat during lactation in the tammar wallaby (*Macropus eugenii*).** *Aust J Biol Sci* 1983, **36**:455-461.
151. Joss JL, Molloy MP, Hinds L, Deane E: **A longitudinal study of the protein components of marsupial milk from birth to weaning in the tammar wallaby (*Macropus eugenii*).** *Dev Comp Immunol* 2009, **33**:152-161.
152. Nicholas KR: **Asynchronous dual lactation in a marsupial, the tammar wallaby (*Macropus eugenii*).** *Biochem Biophys Res Commun* 1988, **154**:529-536.
153. Simpson KJ, Ranganathan S, Fisher JA, Janssens PA, Shaw DC, Nicholas KR: **The gene for a novel member of the whey acidic protein family encodes three four-disulfide core domains and is asynchronously expressed during lactation.** *J Biol Chem* 2000, **275**:23074-23081.
154. Hennighausen LG, Sippel AE: **Characterization and cloning of the mRNAs specific for the lactating mouse mammary gland.** *Eur J Biochem* 1982, **125**:131-141.
155. Hajjoubi S, Rival-Gervier S, Hayes H, Floriot S, Eggen A, Piumi F, Chardon P, Houdebine LM, Thepot D: **Ruminants genome no longer contains Whey Acidic Protein gene but only a pseudogene.** *Gene* 2006, **370**:104-112.
156. Sharp JA, Digby M, Lefèvre C, Mailer S, Khalil E, Topcic D, Auguste A, Kwek JH, Brennan AJ, Familari M, Nicholas KR: **The comparative genomics of tammar wallaby and Cape fur seal lacation models to examine functions of milk proteins.** In *Milk Proteins: from Expression to Food.* Edited by: Thompson A, Boland M, Singh H. San Diego, CA: Academic Press/ Elsevier; 2009:55-79.
157. Sharp JA, Lefèvre C, Kwek JH, Topcic D, Buscara L, Pooley C, de Iongh R, Familari M, Nicholas KR: **Marsupial milk - identifying signals for regulating mammary function and development of the young.** In *Marsupial Genetics and Genomics.* Edited by: Deakin JE, Waters PD, Graves JAM. Dordrecht, Heidleberg, London, New York: Springer; 2010:317-334.
158. Topcic D, Auguste A, De Leo AA, Lefèvre C, Digby MR, Nicholas KR: **Characterization of the tammar wallaby (*Macropus eugenii)* whey acidic protein gene: new insights into the function of the protein.** *Evol Dev* 2009, **11**:363-375.
159. Wanyonyi SS, Sharp JA, Khalil E, Lefevre C, Nicholas KR: **Tammar wallaby mammary cathelicidins are differentially expressed during lactation and exhibit antimicrobial and cell proliferative activity.** *Comp Biochem Physiol A Mol Integr Physiol* 2011, **160**:431-439.
160. Kwek JH, Iongh RD, Digby MR, Renfree MB, Nicholas KR, Familari M: **Cross-fostering of the tammar wallaby (*Macropus eugenii*) pouch young accelerates fore-stomach maturation.** *Mech Dev* 2009, **126**:449-463.
161. **Tammar Wallaby Resources - Genome assemblies.** [http://bioinf.wehi.edu.au/tammar].
162. Burk A, Springer MS: **Intergeneric relationships among Macropodoidea (Metatheria: Diprotodontia) and the chronicle of kangaroo evolution.** *J Mamm Evol* 2000, **7**:213-237.
163. Flannery TF: **Phylogeny of the Macropodoidae: a study in convergence.** In *Kangaroos, Wallabies and Rat-Kangaroos.* Edited by: Grigg PJG, Hume I. Chipping Norton: Surrey Beatty 1989:1-46.

164. Andersson B, Wentland MA, Ricafrente JY, Liu W, Gibbs RA: **A "double adaptor" method for improved shotgun library construction.** *Anal Biochem* 1996, **236**:107-113.
165. Fujiyama A, Watanabe H, Toyoda A, Taylor TD, Itoh T, Tsai SF, Park HS, Yaspo ML, Lehrach H, Chen Z, Fu G, Saitou N, Osoegawa K, de Jong PJ, Suto Y, Hattori M, Sakaki Y: **Construction and analysis of a human-chimpanzee comparative clone map.** *Science* 2002, **295**:131-134.
166. Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK, Groner Y, Soeda E, Ohki M, Takagi T, Sakaki Y, Taudien S, Blechschmidt K, Polley A, Menzel U, Delabar J, Kumpf K, Lehmann R, Patterson D, Reichwald K, Rump A, Schillhabel M, Schudy A, Zimmermann W, Rosenthal A, Kudoh J, *et al*: **The DNA sequence of human chromosome 21.** *Nature* 2000, **405**:311-319.
167. **Cross_match.** [http://www.phrap.org].
168. **Phrap.** [http://www.phrap.org/phredphrapconsed.html].
169. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.
170. Harris RS: **Improved pairwise alignment of genomic DNA.** *PhD Thesis* The Pennsylvania State University, Department of Computer Science and Engineeering; 2007.
171. **LASTZ.** [http://www.bx.psu.edu/~rsharris/lastz/].
172. Zerbino DR, Birney E: **Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**:821-829.
173. Huson DH, Reinert K, Myers EW: **The greedy path-merging algorithm for contig scaffolding.** *J ACM* 2002, **49**:603-615.
174. Sambrook J, Russell DW: *Molecular Cloning: A Laboratory Manual.* 3 edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2001.
175. Boschman GA, Rens W, van Oven CH, Manders EM, Aten JA: **Bivariate flow karyotyping of human chromosomes: evaluation of variation in Hoechst 33258 fluorescence, chromomycin A3 fluorescence, and relative chromosomal DNA content.** *Cytometry* 1991, **12**:559-569.
176. Langford CF, Fischer PE, Binns MM, Holmes NG, Carter NP: **Chromosome-specific paints from a high-resolution flow karyotype of the dog.** *Chromosome Res* 1996, **4**:115-123.
177. Trask B, van den Engh G, Nussbaum R, Schwartz C, Gray J: **Quantification of the DNA content of structurally abnormal X chromosomes and X chromosome aneuploidy using high resolution bivariate flow karyotyping.** *Cytometry* 1990, **11**:184-195.
178. **Ensembl.** [http://www.ensembl.org/].
179. Pathak S, Ronne M, Brown NM, Furlong CL, VandeBerg JL: **A high-resolution banding pattern idiogram of *Monodelphis domestica* chromosomes (Marsupialia, Mammalia).** *Cytogenet Cell Genet* 1993, **63**:181-184.
180. Lander ES, Waterman MS: **Genomic mapping by fingerprinting random clones: a mathematical analysis.** *Genomics* 1988, **2**:231-239.
181. Hillier LW, Fulton RS, Fulton LA, Graves TA, Pepin KH, Wagner-McPherson C, Layman D, Maas J, Jaeger S, Walker R, Wylie K, Sekhon M, Becker MC, O'Laughlin MD, Schaller ME, Fewell GA, Delehaunty KD, Miner TL, Nash WE, Cordes M, Du H, Sun H, Edwards J, Bradshaw-Cordum H, Ali J, Andrews S, Isak A, Vanbrunt A, Nguyen C, Du F, *et al*: **The DNA sequence of human chromosome 7.** *Nature* 2003, **424**:157-164.
182. Deloukas P, Earthrowl ME, Grafham DV, Rubenfield M, French L, Steward CA, Sims SK, Jones MC, Searle S, Scott C, Howe K, Hunt SE, Andrews TD, Gilbert JG, Swarbreck D, Ashurst JL, Taylor A, Battles J, Bird CP, Ainscough R, Almeida JP, Ashwell RI, Ambrose KD, Babbage AK, Bagguley CL, Bailey J, Banerjee R, Bates K, Beasley H, Bray-Allen S, *et al*: **The DNA sequence and comparative analysis of human chromosome 10.** *Nature* 2004, **429**:375-381.
183. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W: **Human-mouse alignments with BLASTZ.** *Genome Res* 2003, **13**:103-107.
184. Suzuki Y, Sugano S: **Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method.** *Methods Mol Biol* 2003, **221**:73-91.
185. **NCBI RefSeq Database.** [ftp://ftp.ncbi.nlm.nih.gov/blast/db/].
186. Abrusan G, Grundmann N, DeMester L, Makalowski W: **TEclass - a tool for automated classification of unknown eukaryotic transposable elements.** *Bioinformatics* 2009, **25**:1329-1330.
187. McCarthy EM, McDonald JF: **LTR_STRUC: a novel search and identification program for LTR retrotransposons.** *Bioinformatics* 2003, **19**:362-367.
188. Lucier JF, Perreault J, Noel JF, Boire G, Perreault JP: **RTAnalyzer: a web application for finding new retrotransposons and detecting L1 retrotransposition signatures.** *Nucleic Acids Res* 2007, **35**:W269-274.
189. Chen Y, Zhou F, Li G, Xu Y: **MUST: A system for identification of miniature inverted-repeat transposable elements and applications to *Anabaena variabilis* and *Haloquadratum walsbyi*.** *Gene* 2009, **436**:1-7.
190. Xu Z, Wang H: **LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons.** *Nucleic Acids Res* 2007, **35**:W265-268.
191. Grad Y, Aach J, Hayes GD, Reinhart BJ, Church GM, Ruvkun G, Kim J: **Computational and experimental identification of *C. elegans* microRNAs.** *Mol Cell* 2003, **11**:1253-1263.
192. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9**:R137.

doi:10.1186/gb-2011-12-8-r81
Cite this article as: Renfree *et al.*: Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biology* 2011 **12**: R81.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

Additional file 1

Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development

SUPPLEMENTARY MATERIAL

Genome sequencing

DNA purification and sequencing reactions (standard AB fluorescent Sanger sequencing technology) were performed using automated solid phase reversible immobilization (SPRI) technology for processing of the DNA preps (SprintPrep Kit, Agencourt Bioscience Corporation) and the sequencing reactions (CleanSeq Kit, Agencourt Bioscience Corporation) [164].

Construction and end-sequencing of tammar large-insert genomic DNA libraries.

A bacterial artificial chromosome (BAC) library, named as MEB1, was constructed from DNA isolated from a male tammar wallaby cell line. In brief, genomic Sac I-digested DNA fragments were inserted to the BAC vector pKS145 [165]. The X chromosome enriched fosmid library, named as MEFX, was constructed from DNA isolated from flow-sorted X chromosomes of a female tammar wallaby cell line, #2070. The extracted DNA molecules were physically fragmented and inserted to the fosmid vector pKS150 [166]. In total, 74,381 MEB1 clones and 17,706 MEFX clones were end-sequenced. The DDBJ accession numbers were DE842016-DE989327 for MEB1 and DH430751-DH462000 for MEFX.

Genome assembly

The Meug_1.0 assembly was produced by assembling whole genome shotgun reads with the Atlas genome assembly system at the Baylor College of Medicine Human Genome Sequencing Center [47]. About 10 million reads were presented to the assembler and 8.5 million reads were assembled, representing about 6.8 Gb of sequence and about 2x sequence coverage of the (clonable) tammar genome (using a 3.6 Gb estimated genome size: see Tables S1-S3). The clone coverage was estimated to be approximately 4x, based on the sum of insert sizes for paired reads divided by the estimated genome size of 3.6 Gb.

The 9.7 million reads were screened using a Cross_match [167] comparison to known sequencing and cloning vectors with the following parameters (coarse screen parameters of minmatch = 20, penalty = -2, and minscore = 30; and finescreen parameters of minmatch = 12, penalty = 2 and minscore = 20). The reads were trimmed with Atlas-screen-window with minimum of 50 contiguous good bases (-l 50), and 40 bases window (-w 40) with quality scores >= 20 (-q 20).

The overlaps among the 9.6 million trimmed reads were assessed using the Atlas-overlapper [47] with the following parameters (-b 0 -e 30000 -B 10 -R 150 -H 151 -k 15 -Y 150 -m 6 -M 50 -S 199 -I -G 8 -p 103058537). Highly repetitive reads were defined as reads that had more than 22 reads that shared sequence overlaps or reads that overlapped with reads that had more than 90 reads that shared sequence overlaps.

Trimmed reads that clustered into groups of two or more that were not highly repetitive, reads were assembled by group using Phrap with the (-ace -forcelevel 10) parameters [168].

The tammar assembly process incorporated two additional steps that were not in the published methods [47]. These steps include the removal of highly repetitive reads using the criteria described above and the application of Atlas-bettergraph. Atlas-bettergraph attempts to

removing "forking" overlaps at the potential repeat boundaries with these parameters (d=0.4 and p=0.5).

The assembled sequences were linearized using Cross_match (minscore = 100) to detect and merge overlapping sequences between neighboring contigs within scaffolds for overlapping tails shorter than 100 bp.

Reads that could not be clustered by the Atlas Overlapper into groups of 2 or more were not assembled. The N50 of the contigs is 2.5 kb and the N50 of the scaffolds is 16.05 kb for the Sanger Meug_1.0 assembly. The N50 size is the length such that 50% of the assembled genome lies in blocks of the N50 size or longer. The total length of all contigs for the Sanger Meug_1.0 assembly is 2.55 Gb. When the gaps between contigs in scaffolds are included, the total span of the assembly is 2.945 Gb.

Highly repetitive reads were defined as reads that had more than 22 reads that shared sequence overlaps or reads that overlapped with reads that had more than 90 reads that shared sequence overlaps. These reads were assembled separately to generate the repetitive contigs for the genome.

Following the Sanger Meug_1.0 sequence assembly, the genome was further sequenced using the ABI SOLiD technology (Table S4) to upgrade the genome assembly by additional superscaffolding. Only about 15 percent of the data, the mate-pairs where both ends are uniquely mapped, was used. Of these mates, about one third had the potential to bridge between scaffolds and two thirds were within a single scaffold. There were few errors in the existing Sanger Meug_1.0 assembly as indicated by the assessment of consistency of the mate-pairs that mapped within scaffolds. About 0.03 percent of these mates were mis-oriented and about one-tenth of one percent had an inter-mate-pair distance greater than 5kb. Using these data, a large number of scaffolds were merged to form new scaffolds and a small number of contigs were removed due to redundancy. The assembly statistics in Table S5 reflect these changes.

The completeness of the assembly was assessed by comparison to the available cDNA data. Using 758,062 454 FLX cDNAs sequences, 76 percent are found to some extent in the assembly and 30 percent are found with more than 80% of their length represented. See Table S6 for statistics. Compared to 14,878 ESTs from Sanger data, more than 85% are found in the assembly with at least one half their length aligned. See Table S7 for data.

Assembly improvement method

The first step in the assembly improvement pipeline is to map the paired reads against the draft contigs using appropriate mapping tools. Bowtie [169] was chosen as the mapping tool for the Illumina data. It was run with default parameters, except each read was allowed up to 3 mismatches total. The longer 454 reads were mapped with lastz [170, 171] with suggested parameters for mapping 454 data. The main difference between the two strategies is the longer 454 reads are better aligned with a tool that allows gapped alignment, Bowtie does not provide this feature and also has hard coded read size limits. The mapping results were filtered to only retain pairs for which both reads in a pair uniquely mapped, or one mate mapped uniquely and the other did not map at all. The uniqueness criterion ensures that the local re-assembly will not attempt to assemble repeated regions.

Next all unique short read pairs associated with a particular set of contigs, along with the draft contigs were reassembled using PHRAP [168] with default parameters. The minimum length of an overlap between two sequences was 14 base pairs, and the minimum score for an alignment was set to 30. Although there are many assembly tools available, PHRAP was found to fit best into this pipeline because it works with reads of different length. It should be noted that the assembler is unaware it is performing a local assembly, and was not designed to fully utilize paired information of short reads to determine order, orientation and position of the contigs. Therefore if the order of the new contigs differed from the order of original contigs, the offending contigs were replaced with their unaltered counterparts. This quality assurance step ensures the structure of the 1.2 assembly is consistent with the 1.1 assembly. It is our assumption that the draft contigs were ordered and oriented correctly since that process was done with the SOLiD reads which had the highest coverage.

The final step is to estimate the gap sizes based on mate pairs which span two adjacent contigs using an expectation maximization algorithm. There are several methods to estimate the size of a gap between consecutive contigs. Mate-pairs and pair-end reads mapped on the contigs form links between them, which can be generally referred to as edges. These edges have some statistical characteristics, such as insertion length and a standard deviation that are a result of the library preparation kit, method and operator error. This information can be used to estimate the gap size. One scaffolding tool, Bambus [42], brings the contigs to the midpoint of the range defined by the length constraints on the edges that linked the contigs. Another, Velvet [172], followed the method in [173], which first initialize the gap length to the mean of all edge insertion lengths. Starting from this value, it then bundles all compatible edges together and uses the mean insertion size as an estimation of the gap length. It would seem that using the mean is a natural way of estimating the gap lengths, however in a draft genome with many potentially large gaps between short contigs, there will exist a bias in the observed pairs mapped. As a result the mean of the observed pairs is usually not the right basis for gap estimation. The issue is resolved in our method through sampling.

The gap between two contigs is estimated using an expectation maximization algorithm. The max maximization step is to compute the gap estimate x, letting the mean insertion length of N pairs equal u, the initial value is the library average

$$x = \frac{1}{N}\sum_{i=1}^{N}(u - d_{i1} - d_{i2}).$$

Here, d is the distance from the position of the mapped read to the appropriate edge of the contig. The sampling step is as follows; given x, and the length of contigs, sample u from completely mapped reads spanning the gap

$$\sum_{i=1}^{N}(d_{i1} + x + d_{i2} - u)^2 .$$

These steps are repeated until the estimates converge. Additionally since multiple libraries were utilized the gap size estimates of two or more libraries was combined according to the following formula:

$$b(e) = \frac{\frac{n_1}{S^{2_1}}e_1 + \frac{n_2}{S^{2_2}}e_2}{\frac{n_1}{S^{2_1}} + \frac{n_2}{S^{2_2}}}$$

Here, b is the bundled estimate, S is the standard deviation for particular library, and n is the number of pairs from that library.

The results were checked by PCR and a comparison of mapping statistics. In order to choose contigs to verify using PCR, the total pool of altered contigs was filtered to only include altered contigs with a gap size of less than 2 kb. From this set of contigs, 10 were randomly selected, 3 of which were triplets. Primers were designed on the left and right most contigs and it was verified that a complete extension occurred. Additionally a set of cDNA was downloaded from ENSEMBL as was a set of fully and partially sequenced BACs from NCBI for *Macropus eugenii*. These were mapped against the initial and improved assemblies and it was found that more bases and sequences were mapped against the improved assembly.

The paired read coverage is the number of pairs that mapped to contigs within 3 standard deviations of the expected insert size. This is indicative of the integrity of changes made to the contigs during the assembly update. There was almost a 2-fold improvement in paired read coverage between the assemblies (Table S8). It should be noted that the Illumina paired read coverage is very low as the libraries were generated with loose control over insert size selection. Table S9 shows the accuracy of changes made to the contigs between assemblies Meug_1.1 and Meug_2.0. Meug_2.0 recovered more of the high confident BACs and transcripts compared to the previous assembly.

Estimation of genome size

Direct DNA content assessment

Genome size was determined using quantitative PCR, according to the method described by [46]. Human (3 Gb) and *Saccharomyces cerevisiae* (12.1 Mb) genomes were used as controls to assess the accuracy of the assay and percentage error rate in the calculations. Tammar tissue was obtained from the same animal used to generate the genome. High molecular weight human genomic DNA was obtained from Invitrogen. DNA was extracted from the tammar tissue according to standard methods [174]. Real-time PCR was conducted using primers listed in Table S10. The single copy gene target in the tammar was *Dhh* as it has been shown to be single copy by Southern blot, fluorescent *in situ* hybridization, PCR and based on the initial genome assembly (1.0) and sequence trace archives (O'Hara, W.A., Azar, W.J., Behringer, R.R., Renfree, M.B., and Pask, A.J., unpublished results). The assay was performed on undiluted DNA, 1:10 and 1:100 dilutions to ensure accuracy and validate results. Both human (of similar size to the tammar genome) and *S.cerevisiae* (three orders of magnitude smaller) genomes were calculated to be within 8% and 4% of their actual size respectively, validating the technique (Table S11).

The tammar genome size assessment was carried out a total of 14 independent times over three separate real time plates and on two independent DNA extractions. The genome size estimates were extremely consistent across replicates and the tammar haploid genome size was calculated to be 2.7 Gb ± 170 Mb with a weight of approximately 2.96 pg, slightly smaller than that of human and mouse.

Cytometry

Tammar and opossum chromosome sizes were determined by flow cytometry in two different laboratories, RIKEN and Cambridge. At Cambridge, bivariate flow karyotyping was performed as previously described [175-177]. Chromosome preparations of tammar and human, or opossum and human, were measured together and separately but sequentially with the same flow cytometry settings, allowing a direct comparison between the tammar chromosomes and human chromosomes, and opossum chromosomes and human chromosomes. The "DNA-line" in the flow karyotypes was drawn from the origin through human chromosome 4. Each human, tammar, or opossum chromosome peak was projected onto this line, and the distance from the origin to this projection was an estimate of the DNA

content of that particular chromosome. The established human chromosome sizes [178] were used as references for the tammar and opossum chromosomes taking into account both conversion and offset. Tammar and opossum genome size was determined by adding up the chromosome sizes.

The tammar chromosome complement consists of seven autosome pairs and one pair of sex chromosomes, X and Y. Chromosome 1 is relative large, chromosome 2 to 6 are medium sized, and chromosome 7 and X are smaller sized. Chromosome Y is a relative tiny chromosome. The sizes are presented in Table S12. Tammar chromosomes 1-6 are all larger than human chromosomes. For instance, MEU1 is about twice the size of HSA1. Notable is the size of MEUX of 150 Mb, which is a value between HSA8 and HSAX. Despite MEUX having a much smaller gene content than HSAX and being regarded as the ancestral therian X, its DNA content is comparable with HSAX, due to the addition of repetitive sequences of this NOR-bearing chromosome. The opossum chromosome X does not carry the NOR and has a size of 97 Mb (Table S12), which is about the expected 2/3 size of the eutherian chromosome X. By comparison of the DNA content and GC content of the tammar X and opossum X (Table S12) and assuming that the remainder part of the tammar X is conserved between tammar and opossum, one can calculate that the region added to the tammar X consisted of around 10 Mb GC and 43 Mb AT. The tammar genome size from the flow sorted chromosomes is estimated to be 2.457 Gb determined by adding up the chromosome sizes given in Table 1. This physical size is smaller than the human genome size of 3.08 Gb.

The karyotype of the South American *Monodelphis domestica* consists of eight autosome pairs and one pair of X-Y sex chromosomes. The chromosomes in the karyotype are ordered according to [39, 179]. Chromosome 1 is relatively large, chromosome 2 to 8 are all medium sized, only chromosome X is smaller sized. Chromosome 5 and 8 have similar sizes as chromosomes were ordered by centromere position as first parameter [179]. Chromosome Y is a tiny chromosome and too small to be measured by flow cytometry. Apart from MDO7 and MDOX, all MDO chromosomes are larger than human chromosomes. MDOX is much smaller than the human X (about two third), which is due to autosomal genomic regions being translocated to the eutherian X during evolution. The opossum genome size is estimated to be 3.172 Gb determined by adding up the chromosome sizes given in Table S12. This physical size is similar to the human genome size of 3.08 Gb, but bigger than the tammar genome size of 2.457 Gb.

Flow sorting genome size estimation of tammar chromosomes at RIKEN was performed using a standard flow-cytometry protocol (Cycle TEST PLUS DNA Reagent Kit, Becton & Dickinson). Nuclei fractions were isolated from cultured cells derived from three different tammar individuals, including the animal used for the genome sequencing and from two human cell lines, #GM130B, #GM18940, and stained with propidium di-iodide. The estimated tammar genome size was 3.6 Gb in average given that of human genome was 3.0 Gb. The exact values were 3.73, 3.59 and 3.65 Gb for the cell lines #2409♂, #3469♂ and #2070♀, respectively (see Table S13).

Sequencing-based genome size estimation

Finally, the genome size was estimated from the Sanger WGS sequence data using the Atlas-Genometer. This method is based on the Lander-Waterman model [180] and a linear model to relate the number of reads that have no overlaps to the observed number of these reads adjusting for repeats and sequencing errors. The method calculates the genome size based upon sampling subsets of reads from low coverage data samples and uses an Expectation

Maximization algorithm to calculate a convergent genome size. Using a variety of Atlas-overlapper calculations [47] with between 360,000 and 3,600,000 reads, and initial genome sizes of 3.2 to 4.9 Gb, the tammar genome size is estimated to be 3.65 Gb for calculations that converge (smaller starting genome sizes did not converge) (Table S13). The method computes a value of 3.07 Gb for the human genome (3.08 Gb including Ns, 2.85 Gb excluding Ns for NCBI build 36) using 0.5x 454 data. Assuming that the estimate is within 10% of the actual genome size, the range of estimated genome sizes for the tammar would be 3.29 to 4.1 Gb. Based on the available EST data and comparisons to the Meug_1.0 and Meug_1.1 genome assemblies (Ensembl), the genome size is estimated to be between 3.00 and 3.78 Gb. This calculation uses the fraction of the ESTs that are not aligned over at least 50% of their length in the existing assemblies to calculate the missing fraction of the genome. The lower number is based upon comparison to 14,878 Sanger ESTs, and the higher number is based upon comparison to 758,062 454 FLX ESTs.

Tammar and opossum GC-content

The GC content of each tammar or opossum chromosomes was determined using the GC content of human chromosome 7 [181] and human chromosome 10 [182] as a reference. Each peak in the flow karyotypes was projected on the GC axis of the flow karyotype. The distance to the origin of these projections was regarded as a measure for the absolute GC content. The relative GC content of human chromosome 7 and 10 was converted into an absolute number using the DNA content values given by Ensembl. The calculated absolute GC content, taking into account both conversion and offset, of tammar and opossum chromosomes were converted into relative GC contents using the DNA content of these chromosomes calculated above. In tammar, the chromosome GC varies from a relatively low 27% (chr7) to 37% (chr1) with an average of 34% (Table S12). This is substantially lower that the GC for human (42%) and opossum (40%). The tammar X also has a GC content (34%) lower than that of the opossum X (42%). Thus, tammar chromosomes are relatively GC poor.

Sequence conservation

Analyses were performed on tammar assembly version 1.0. Ensembl genesets used for comparative analyses were from release 58. BlastZ-net pairwise alignment [183] was constructed between tammar and opossum.

Transcriptome sequencing and analysis

Construction and analyses of full-length cDNA libraries

Prof. Renfree (University of Melbourne) provided tissues from testis, ovary and hypothalamus cDNA sequencing. The full-length cDNA libraries were constructed by Prof. Sugano's group at University of Tokyo (120,000 clones in total). End sequencing was produced by RIKEN-GSC (Toyoda and Kuroki) and Prof. Kohara's group at National Institute of Genetics. Sequence data from three cDNA libraries were combined and processed by removing vectors, low-quality (average QV<=15) nucleotides and poly(A) sequences. The 5' sequences were clustered at ninety percent identity using the cd-hit-est program. Dr. Tatsumoto (RIKEN-GSC) performed data analysis and clustering.

The oligo-capped full-length cDNA libraries were constructed from RNAs extracted from adult tammar tissues as described previously [184]. The resulting cDNA libraries were named as METC for testis, MEOC for ovary, MEHC for hypothalamus, and MEGC for gravid uterus, respectively. According to the method, we estimated that the range of the insert size of the libraries as ~3 kb. From these libraries, we obtained 360,350 end-sequence reads in total for both 5' and 3' ends of the inserts (DDBJ MEGC: FY469875-FY560833. MEHC:

FY560834-FY602565. MEOC: FY602565-FY644882. METC: FY644883-FY736474).

CDNA sequencing was also performed at BCM-HGSC using the 454 GS FLX platform for two thymus libraries (thoracic and cervical samples). These data can be found in the NCBI Sequence Read Archive under Experiment numbers SRX019249 and SRX019250.

Tissue-specific high-quality full-length cDNA libraries are invaluable for the future biology in the post genome-sequencing era. We constructed such libraries from testis, ovary, hypothalamus and gravid uterus using oligo-capping technology [184]. Clones from each library were subjected to end-sequencing to evaluate composition and complexity of each transcriptome (360,350 sequence reads in total). After clustering, the library from hypothalamus, named MEHC, showed highest complexity, whereas the one from ovary, named MEOC, was lowest, 44.3% and 18.8%, respectively (percentage ratio of the number of clusters against those of sequence-reads). The clusters cover 88.5% (MEHC) to 92.9% (MEOC) of the current genome assembly (Meug_1.1), the others fell in-between, indicating high quality of our libraries. We then looked for representative genes in each library within the Ref-seq database (refseq_rna(blastdb), Date: 2010.10.30, [185]). For example, homolog of *KLH10* (OMIM 608778) and *ODF1/2* (OMIM 182878) genes, both function in spermatogenesis and male fertility were found in the testis library, named METC. The number of the reads of these genes reached 4.3% and 3.6% of the entire reads having E value <= 1e-30, respectively. Similarly, the hypothalamus library, MEHC, was rich in tubulin family genes, 7.9%, and some hormone related genes such as *SST* (somatostatin; OMIM 182450), 1.8% (Tables S18a).

When the transcriptomic reads were assembled and aligned to the tammar genome, we found that of the 141,574 contigs and unassembled reads, 94,890 aligned to 15,043 (84%) Ensembl-predicted genes (Release 57) were supported. Among the 2,834 of Ensembl genes not supported by transcriptomic data, 60% were protein-coding genes, 20% pseudogenes, and 20% non-coding RNAs. Only 25,415 contigs aligned to unannotated regions of the assembly (>2kb from a predicted gene). Of these, 5,545 had high quality matches (E-value<10^{-5}) to proteins in the NCBI Non-Redundant (NR) database.

Repeats

The repeat content of the tammar genome was assessed using RepeatMasker, RepeatModeler and *ab initio* repeat prediction programs. The Repbase database of consensus repeat sequences was used to identify repeats in the genome derived from known classes of elements (Table 2). RepeatModeler uses a variety of *ab initio* tools to identify repetitive sequences irrespective of known classes [58]. Unannotated repeats were further queried in the following specialized repeat class annotation tools; TEclass [186], LTR_STRUC [187], RTAnalyzer [188], MUST [189], LTR-FINDER [190]. The remaining putative *de novo* repeats were aligned to the Repbase repeat annotations using BLAST. Any *de novo* repeat with≥50% identity and coverage was annotated as that specific Repbase element. All of the putative *de novo* repeats that could not be annotated were considered *bona fide, de novo* repeats. The results from the database and *de novo* RepeatMasker annotations were combined, and any overlapping annotations were merged if they were of the same class of repeat element. Overlapping repeats from different classes were reported; therefore each position in the genome may have more than one unique annotation.

Small RNAs

The small RNA sequences were processed to remove sequencing adapters then selected for the appropriate size range. The trimmed and size selected libraries were then mapped to the genome using Bowtie [169]. One mis-match was allowed and all valid alignments for each of the sequences were reported.

Bona fide hairpin structures were identified using the hairpin loop prediction tool SRNALOOP [191]. A 100 base pair window up and downstream of each valid micro RNA alignment was selected as input for each experiment. This process is computationally challenging and a pipeline was developed to perform this task on a computer cluster. Hairpin target locations are simply defined as all valid mapped micro RNA locations.

The miRBase [74] mature sequences were mapped using Bowtie allowing for 3 mismatches and the longer hairpin and sequences were mapped using the local alignment tool Lastz [170, 171], 75% identity was required for each hit.

ChIP-seq mapping was performed as per (Lindsay, J., Carone, D.M., Murchison, E., Hannon, G., Pask, A.J., Renfree, M.B. and O'Neill, R.J., unpublished results) and MACsPeak calls were obtained as per [192]. CrasiRNAs were mapped to the genome using bowtie as above.

SUPPLEMENTARY TABLES

Table S1. Sanger-assembled reads statistics for Meug_1.0

	Reads (million)	Bases (billion)
Total Sequenced	9.7	NA
Trimmed	9.6	7.7
Assembled	8.5	6.8
Unassembled	0.52	0.42
Repeat	0.6	0.48

Table S2. Summary of Sanger sequencing libraries

	Number of Libraries	Number of Reads
3kb	29	3,654,977
4kb	2	2,788,140
6.25 kb	2	3,248,488

Table S3. Sanger sequencing libraries

Library	Number of reads	Insert size*	Std.dev.
2070LUS01	61,488	3,000	1,000
2070LUS03	124	3,000	1,000
2070LUS05	1,839	3,000	1,000
2070LUS07	954	3,000	1,000
2070LUS08	14,509	3,000	1,000
2070LUS09	7,391	3,000	1,000
2070LUS10	1,535	3,000	1,000
2070LUS11	22,286	3,000	1,000
2070LUS12	16,170	3,000	1,000
2070LUS14	641	3,000	1,000
2070LUS15	735	3,000	1,000
2070LUS17	40,730	3,000	1,000
2070LUS18	85,122	3,000	1,000
2070LUS19	122,240	3,000	1,000
2070LUS21	1,396	3,000	1,000
2070LUM20	673	3,000	1,000
2070LUM22	2,031	3,000	1,000
2070LUS23	1,352	3,000	1,000
2070LUS24	1,405	3,000	1,000
2070LUS25	99,869	3,000	1,000
2070LUS26	2,928	3,000	1,000
2070LUS27	52,134	3,000	1,000
2070LUS28	1,394,102	3,000	1,000
2070LUS29	20,716	3,000	1,000
2070LUS30	91,093	3,000	1,000
2070LUS31	249,513	3,000	1,000
2070LUS32	660,312	3,000	1,000
2070LUS33	636,348	3,000	1,000
2070LUS34	65,341	3,000	1,000
AUWAP	2,785,918	4,000	1,000
AUWCP	2,222	4,000	1,000
AUWBP	1,158,027	6,250	1,250
AUWDP	2,090,461	6,250	1,250

*Insert size is the targeted insert size for the library

Table S4. SOLiD read statistics

	F3	R3	Total
Raw reads	356,241,796	357,736,722	713,978,518
Uniquely Mapped Pairs	52,995,899	52,995,899	105,991,798
Percent of Raw Reads	15	15	15
Scaffold Bridge	17,158,908	17,158,908	34,317,816
Percent of Mapped Pairs	32	32	32
Within Scaffold	35,836,991	35,836,991	71,673,982
Percent of Mapped Pairs	68	68	68
Mis-oriented	10,777	10,777	21,554
Percent of Mapped Pairs	0.0301	0.0301	0.0301
Inter-pair distance > 5kb	36,687	36,687	73,374
Percent of Mapped Pairs	0.1024	0.1024	0.1024

Table S5. Genome assembly statistics

		Number	N50(kb)	Bases+Gaps(Mb)	Bases(Mb)
Meug_1.0	Contigs	1,211,471	2.5	2,549	2,549
Meug_1.1	Contigs	1,174,382	2.6	2,536	2,536
Meug_1.0	Scaffolds	616,418	16.05	2,945	2,549
Meug_1.1	Scaffolds	277,711	36.6	3,075	2,536

Table S6. Comparison of genome assembly to 454 FLX cDNA data

	Meug_1.0		Meug_1.1	
Length of aligned cDNA	Number of Reads	Percentage	Number of Reads	Percentage
100%	5,248	0.69%	5,280	0.70%
95%	9,652	1.27%	9,722	**1.28%**
80%	230,242	30.37%	229,378	30.26%
50%	511,128	67.43%	508,791	67.12%
>0%	579,735	76.48%	576,647	76.07%
Total cDNAs	758,062	100.00%	758,062	100.00%

Table S7. Comparison of genome assembly to 14,878 ESTs

	Percent Aligning to Genome	
Length of cDNA aligned	Meug_1.0	Meug_1.1
100%	25.40%	25.50%
95%	50.00%	50.40%
80%	71.30%	72.20%
50%	84.80%	85.80%

Table S8. Summary statistics of the tammar genome assemblies.

type	count	total bases	avg length	coverage (2.7GB)	paired read coverage Meug_1.1	paired read coverage Meug_2	Coverage Change
Sanger	11745817	9,445,367,851	804	3.5	n/a	n/a	n/a
454	1719180	396,757,212	230	0.15	0.41750	0.46629	1.12
SOLiD	710427474	17,760,686,850	25	6.58	0.13556	0.23594	1.74
Illumina	271875064	27,187,506,400	100	10.07	0.00056	0.00097	1.7376

Table S9. Comparison of Meug 1.1 and 2.0 genome assemblies.

pool	assembly	# sequences	# bases	# mapped seqs	# mapped bases
BACs	1.1	169	1,269,736	35	831,033
BACs	2	169	1,269,736	35	836,798
ESTs	1.1	84,718	104,868,227	68,336	60,076,451
ESTs	2	84,718	104,868,227	68,518	60,676,841

Table S10. Primers used for genome size assessment by qPCR.

Target Gene	Name	Sequence 5' →3'	Product Size (bp)
Saccharomyces cerevisiae ribosomal protein S3 gene (rps3), GenBank accession no. U34347.	S.c. RPS3-F1	CGCTGACGGTGTCTTCTAC	382
	S.c. RPS3-R1	CCAACCAAGACCGAAGTTAT	
	S.c. RPS3-F2	CGGAAACAACAACTTCACAA	172
	S.c. RPS3-R2	GACAGCGGACAAACCA	
Homo sapiens sapiens p53 tumor suppressor gene, GenBank accession no. X54156.	H.s. p53-F1	CGGCGCACAGAGGAAGAGAAT	342
	H.s. p53-R1	TTCCTAGCACTGCCCAACA	
	H.s. p53-F2	CAAATGCCCCAATTGCAGGTA	125
	H.s. p53-R2	GACTGGAAACTTTCCACTTG	
Macropus eugenii desert hedgehog gene, Ensembl transcript ID ENSMEUG00000014181	M.e. DHH F1	CCTGGACCGAGACTTACAGC	465
	M.e. DHH R1	CAGCCCAGTAGTTCCTCTGC	
	M.e. DHH F2	GAGCTGGGGACTCGGTACTT	179
	M.e. DHH R2	ATGAGCCCACTGGTGACTCT	

Table S11. Genome sizes estimated by qPCR.

Target	Sample Concentration	Calibration curve	Amplified copies	C (pg)	Γ (bp)
Tammar Dhh	58.415 ng/µl	Y = -3.9329× + 37.314 (R^2 = 0.999)	$1.972 \times 10^4 \pm 6\%$	2.96	2.7×10^9
S.cerevisiae rps3	7.776 ng/µl	Y = -3.4661x + 38.755 (R^2 = 0.999)	$5.87 \times 10^5 \pm 4\%$	0.014	12.7×10^6
H.sapien p53	273 ng/µl	y = -3.6772x + 38.461 (R^2 = 0.996)	$7.88 \times 10^4 \pm 8\%$	3.48	3.1×10^9

Table S12. Tammar and opossum chromosome sizes estimated by flow-sorting.

Tammar chromosome	Size (Mb)	GC	Opossum chromosome	Size (Mb)	GC
Chromosome 1	486	0.37	Chromosome 1	615	0.42
Chromosome 2	367	0.35	Chromosome 2	472	0.40
Chromosome 3	355	0.35	Chromosome 3	472	0.40
Chromosome 4	340	0.34	Chromosome 4	406	0.41
Chromosome 5	340	0.34	Chromosome 5	297	0.39
Chromosome 6	286	0.31	Chromosome 6	272	0.40
Chromosome 7	133	0.27	Chromosome 7	243	0.34
			Chromosome 8	297	0.39
Chromosome X	150	0.34	Chromosome X	97	0.42
Total Genome Size	2457	0.34		3172	0.40

Table S13. Tammar genome sizes estimated by flow-sorting at RIKEN.

Sample name	Estimated genome size (Gb)
Tammar♂ #2409	3.50
Tammar♀ #2070	3.74
Tammar♂ #3469	3.59
Average	3.61

Table S14. Sample size, non-overlap reads and resulting genome size estimates for Sanger WGS comparisons

Total Reads	Non-overlaping reads	Genome size (bp)
360,000	292,354	3,696,864,843
720,000	500,119	3,683,969,866
1,080,000	649,020	3,613,898,638
1,440,000	750,921	3,628,158,993
1,800,000	820,449	3,624,589,468
2,160,000	858,833	3,646,632,509
2,520,000	880,001	3,653,505,638
2,880,000	881,118	3,672,731,370
3,240,000	872,580	3,694,813,340
3,600,000	855,625	3,655,653,421

Table S15. Tammar-specific gene family expansions

Tammar Ensembl Gene ID	Associated Gene Name
ENSMEUG00000008632	Cd9
ENSMEUG00000007545	Cd9
ENSMEUG00000013079	Cd9
ENSMEUG00000011422	Atp12a
ENSMEUG00000003357	Atp12a
ENSMEUG00000006532	Trmt2b
ENSMEUG00000004230	Slc37a2
ENSMEUG00000001778	Slc37a2
ENSMEUG00000014731	Olfr270
ENSMEUG00000002021	Olfr270
ENSMEUG00000002000	Olfr270
ENSMEUG00000005442	Hba-x
ENSMEUG00000005434	Hba-x
ENSMEUG00000003990	Snapc5
ENSMEUG00000008996	Snapc5
ENSMEUG00000005596	H2-Oa
ENSMEUG00000014073	H2-Oa
ENSMEUG00000010599	H2-Oa
ENSMEUG00000003576	H2-Oa
ENSMEUG00000015590	Tmbim6
ENSMEUG00000012022	Tmbim6
ENSMEUG00000014841	Tbl1xr1
ENSMEUG00000007721	Tbl1xr1
ENSMEUG00000005644	Sept8
ENSMEUG00000007024	Sept8
ENSMEUG00000013104	Slc29a1
ENSMEUG00000000898	Slc29a1

Tammar Ensembl Gene ID	Associated Gene Name
ENSMEUG00000010554	Rfng
ENSMEUG00000010012	Rfng
ENSMEUG00000000128	Ccna2
ENSMEUG00000008916	Ccna2
ENSMEUG00000010254	2310046K01Rik
ENSMEUG00000005866	2310046K01Rik
ENSMEUG00000015304	2310046K01Rik
ENSMEUG00000004730	Ncbp2
ENSMEUG00000008126	Ncbp2
ENSMEUG00000009125	Net1
ENSMEUG00000009296	Net1
ENSMEUG00000013967	Ocrl
ENSMEUG00000002996	Ocrl
ENSMEUG00000006516	Mospd1
ENSMEUG00000002880	Mospd1
ENSMEUG00000001573	Rbm22
ENSMEUG00000009348	Rbm22
ENSMEUG00000015569	Olfr15
ENSMEUG00000016279	Olfr15
ENSMEUG00000015565	Olfr15
ENSMEUG00000000279	Olfr15
ENSMEUG00000012279	Smc6
ENSMEUG00000001811	Smc6
ENSMEUG00000004750	Fzr1
ENSMEUG00000015522	Fzr1
ENSMEUG00000006976	Fzr1
ENSMEUG00000006899	Fzr1
ENSMEUG00000008590	Casp3
ENSMEUG00000014666	Casp3
ENSMEUG00000013544	Casp3

Table S16. Tammar-opossum sequence conservation

Chromosome	% opossum sequence aligned to tammar	% opossum gene sequence aligned to tammar	% opossum unannotated sequence aligned to tammar
1	40.6	13.9	26.7
2	38.5	14.7	23.8
3	36.0	11.5	24.4
4	38.2	13.0	25.2
5	36.1	11.4	24.7
6	39.4	14.1	25.2
7	36.2	9.9	26.3
8	39.0	13.3	25.7
X	40.9	12.7	28.2

Table S17. Summary statistics from transcriptome libraries
(a) Overview of all libraries

Tissue	Mammary gland	Ovary	Gravid uterus	Testis	Hypothalamus	Thymus
Number of reads	14,836	84,790	184,781	100,863	119,542	758,062
Length (bp)	608	1135	1125	725	1115	183
GC (%)	45.47	46.40	46.89	49.33	44.50	44.53
Bases Masked (%)[a]	8.7	2.4	3.6	5.02	7.15	31.1
Valid for assembly[b]	13,595	84,154	182,280	96,986	114,688	379,695

[a] by RepeatMasker
[b] after pre-processing with Seqclean

(b) Hybrid assembly with MIRA3

Number of reads input	1,294,241 (491,703 Sanger cDNAs)
Number of reads assembled	880,700 (68.0%)
Number of singlets	20,422
Number of contigs	121,152
Largest contig	4908
N50 contig size	1389
N90 contig size	509
N95 contig size	239
Max coverage (total)	1383
Max coverage (Sanger)	1740
Max coverage (454)	1361
Average consensus quality	16
Strong unresolved repeat positions (SRMc)	58

Table S18. Summary statistics from transcriptome libraries
(a) Sanger end-sequenced cDNA libraries

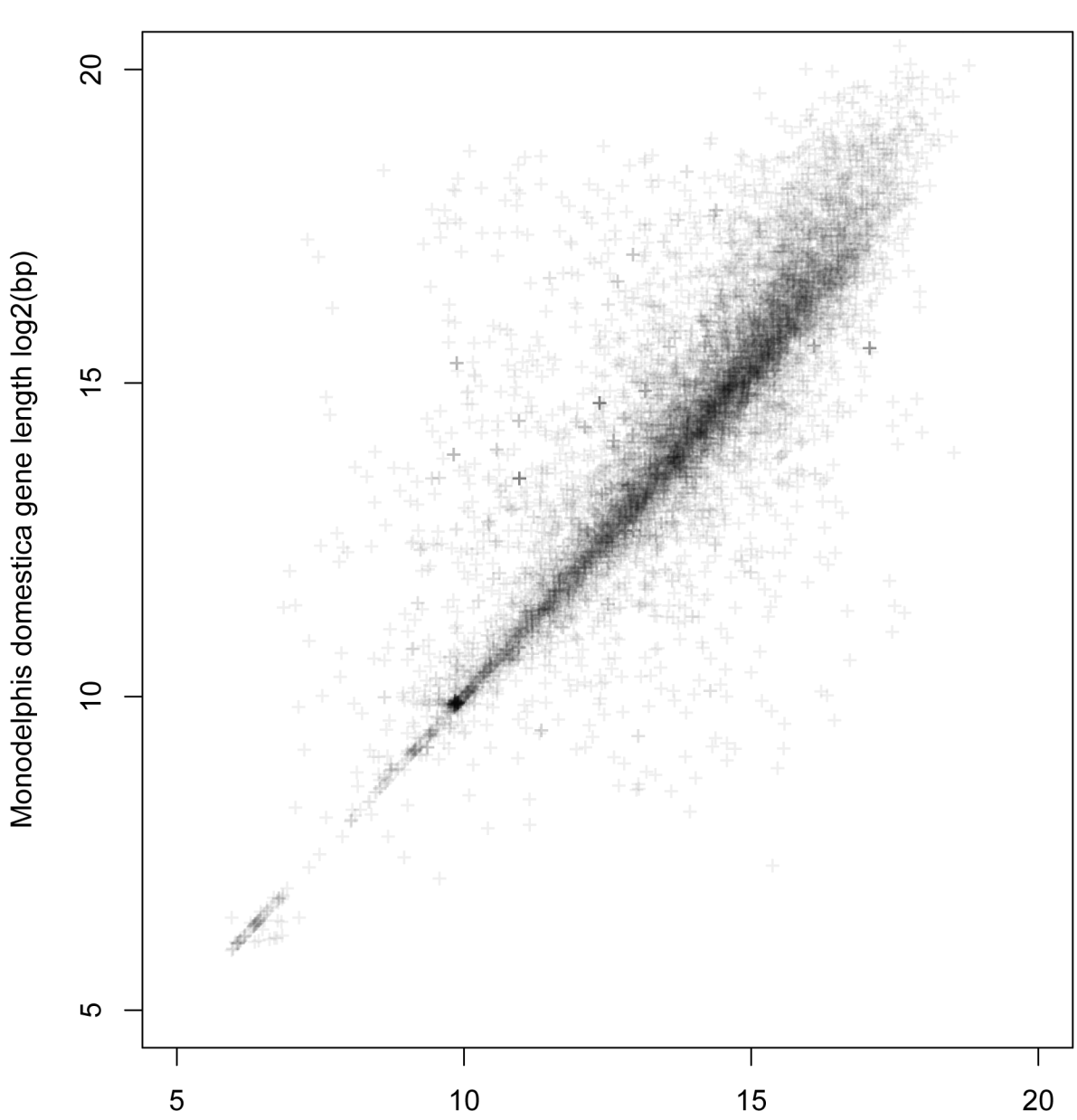
cDNA Library: Number of Reads and Clusters

library	insert	reads	clusters	ratio(%)
MEGC	F (5')	45456	10897	23.97%
	R (3')	44861	9479	21.13%
MEHC	F (5')	48876	21646	44.29%
	R (3')	48531	16832	34.68%
MEOC	F (5')	40475	7618	18.75%
	R (3')	40636	6743	16.66%
METC	F (5')	77924	17377	22.30%
	R (3')	13591	5742	42.25%
TOTAL		360350	96334	26.73%

(b) Hypothalamus top 20 most abundant hits in RefSeq

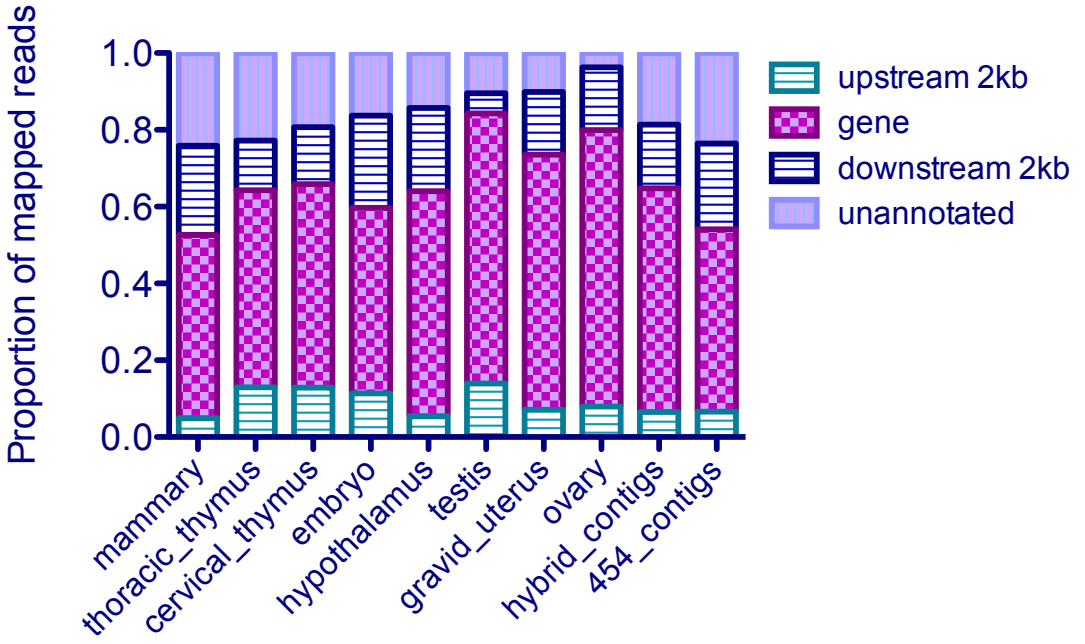
MEHC top 20 hits in RefSeq

ID	number of cDNAs	annotations
1407	279	kelch-like 10 (Drosophila) (KLHL10), mRNA, spermiogenesis and male fertility
6828	269	eukaryotic translation elongation factor 1 alpha 1 (EEF1A1), mRNA
5307	217	actin-like protein (FKSG30), mRNA
4782	207	eukaryotic translation elongation factor 1 alpha 1 (EEF1A1), mRNA
822	176	ubiquitin A-52 residue ribosomal protein fusion product 1 (UBA52), transcript variant 1, mRNA
1469	176	ribosomal protein L4 (RPL4), mRNA
697	147	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide mRNA
1597	146	outer dense fiber of sperm tails 1 (ODF1), mRNA
228	143	nuclear protein localization 4 homolog (S. cerevisiae) (NPLOC4), mRNA
473	140	actin-like protein (FKSG30), mRNA
1700	139	ubiquitin A-52 residue ribosomal protein fusion product 1 (UBA52), transcript variant 1, mRNA
662	135	ubiquitin A-52 residue ribosomal protein fusion product 1 (UBA52), transcript variant 1, mRNA
1128	132	tubulin, beta 2A (TUBB2A), mRNA
1376	126	ubiquitin A-52 residue ribosomal protein fusion product 1 (UBA52), transcript variant 1, mRNA
219	123	zinc finger protein 259 (ZNF259), mRNA
1351	115	hook homolog 1 (Drosophila) (HOOK1), mRNA
570	110	dihydropyrimidinase-like 5 (DPYSL5), mRNA
548	107	chaperonin containing TCP1, subunit 3 (gamma) (CCT3), transcript variant 3, mRNA
3321	99	ubiquitin A-52 residue ribosomal protein fusion product 1 (UBA52), transcript variant 1, mRNA
1176	92	phosphatidylethanolamine binding protein 1 (PEBP1), mRNA

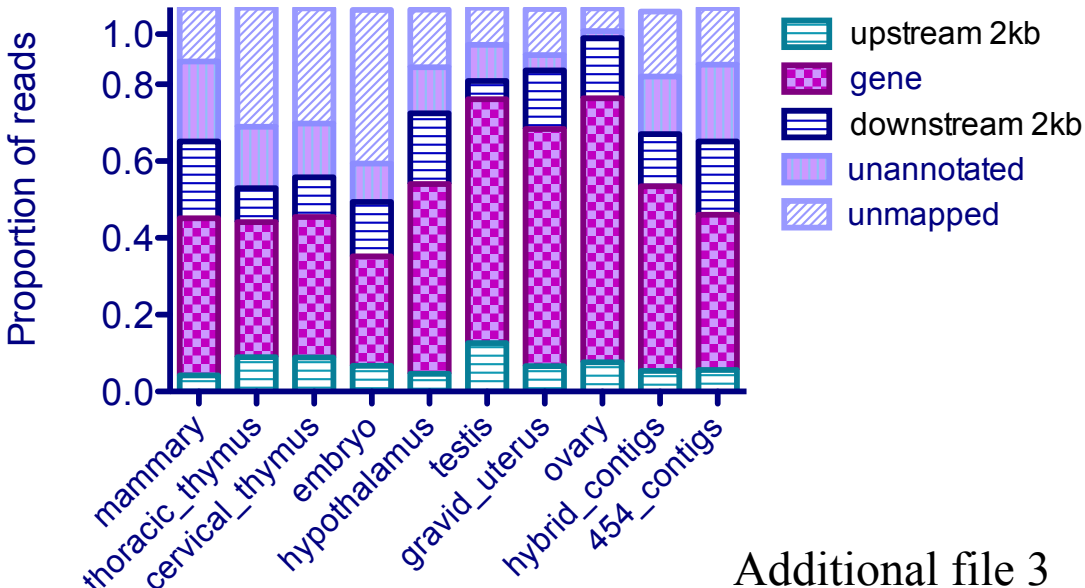


Additional file 2

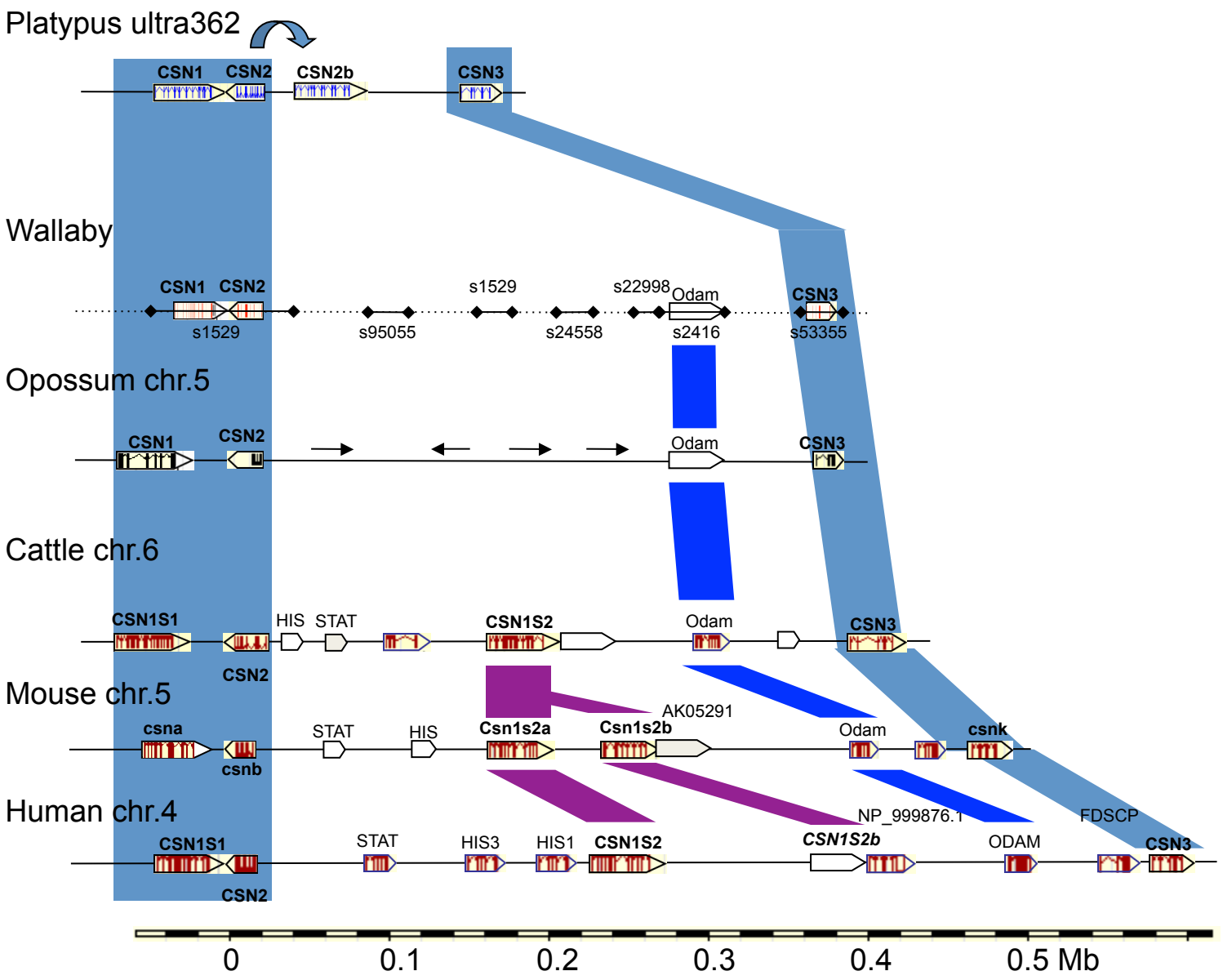
a.



b.



Additional file 3



Additional file 4

Genome Biology

CORRECTION

Correction: Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development

Marilyn B Renfree[1,2*†], Anthony T Papenfuss[1,3,4*†], Janine E Deakin[1,5], James Lindsay[6], Thomas Heider[6], Katherine Belov[1,7], Willem Rens[8], Paul D Waters[1,5], Elizabeth A Pharo[2], Geoff Shaw[1,2], Emily SW Wong[1,7], Christophe M Lefèvre[9], Kevin R Nicholas[9], Yoko Kuroki[10], Matthew J Wakefield[1,3], Kyall R Zenger[1,11,7], Chenwei Wang[1,7], Malcolm Ferguson-Smith[8], Frank W Nicholas[7], Danielle Hickford[1,2], Hongshi Yu[1,2], Kirsty R Short[12], Hannah V Siddle[1,7], Stephen R Frankenberg[1,2], Keng Y Chew[1,2], Brandon R Menzies[1,13,2], Jessica M Stringer[1,2], Shunsuke Suzuki[1,2], Timothy A Hore[1,14], Margaret L Delbridge[1,5], Hardip R Patel[1,5], Amir Mohammadi[1,5], Nanette Y Schneider[1,2,15], Yanqiu Hu[1,2], William O'Hara[6], Shafagh Al Nadaf[1,5], Chen Wu[7], Zhi-Ping Feng[16,3], Benjamin G Cocks[17], Jianghui Wang[17], Paul Flicek[18], Stephen MJ Searle[19], Susan Fairley[19], Kathryn Beal[18], Javier Herrero[18], Dawn M Carone[6,20], Yutaka Suzuki[21], Sumio Sugano[21], Atsushi Toyoda[22], Yoshiyuki Sakaki[10], Shinji Kondo[10], Yuichiro Nishida[10], Shoji Tatsumoto[10], Ion Mandiou[23], Arthur Hsu[16,3], Kaighin A McColl[3], Benjamin Lansdell[3], George Weinstock[24], Elizabeth Kuczek[1,26,25], Annette McGrath[25], Peter Wilson[25], Artem Men[25], Mehlika Hazar-Rethinam[25], Allison Hall[25], John Davis[25], David Wood[25], Sarah Williams[25], Yogi Sundaravadanam[25], Donna M Muzny[24], Shalini N Jhangiani[24], Lora R Lewis[24], Margaret B Morgan[24], Geoffrey O Okwuonu[24], San J Ruiz[24], Jireh Santibanez[24], Lynne Nazareth[24], Andrew Cree[24], Gerald Fowler[24], Christie L Kovar[24], Huyen H Dinh[24], Vandita Joshi[24], Chyn Jing[24], Fremiet Lara[24], Rebecca Thornton[24], Lei Chen[24], Jixin Deng[24], Yue Liu[24], Joshua Y Shen[24], Xing-Zhi Song[24], Janette Edson[25], Carmen Troon[25], Daniel Thomas[25], Amber Stephens[25], Lankesha Yapa[25], Tanya Levchenko[25], Richard A Gibbs[24], Desmond W Cooper[1,28], Terence P Speed[1,3], Asao Fujiyama[22,27], Jennifer A M Graves[1,5], Rachel J O'Neill[6], Andrew J Pask[6,1,2], Susan M Forrest[1,25] and Kim C Worley[24]

Correction

While compiling the article describing the genome sequence of the tammar wallaby [1] one of the authors was inadvertently omitted from the author list. This author, Hardip Patel, has been included in the corrected author list above.

Author details

[1]The Australian Research Council Centre of Excellence in Kangaroo Genomics, Australia. [2]Department of Zoology, The University of Melbourne, Melbourne, Victoria 3010, Australia. [3]Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia. [4]Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Victoria 3010, Australia. [5]Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia. [6]Department of Molecular and Cell Biology, Center for Applied Genetics and Technology, University of Connecticut, Storrs, CT 06269, USA. [7]Faculty of Veterinary Science, University of Sydney, Sydney, NSW 2006, Australia. [8]Department of Veterinary Medicine, University of Cambridge, Madingley Rd, Cambridge, CB3 0ES, UK. [9]Institute for Technology Research and Innovation, Deakin University, Geelong, Victoria, 3214, Australia. [10]RIKEN Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [11]School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia. [12]Department of Microbiology and Immunology, The University of Melbourne, Melbourne, Victoria 3010, Australia. [13]Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, Berlin 10315, Germany. [14]Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge, CB22 3AT, UK. [15]Department of Molecular Genetics, German Institute of Human Nutrition, Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany. [16]Department of Medical Biology, The University of Melbourne, Melbourne, Victoria 3010, Australia. [17]Biosciences Research Division, Department of Primary Industries, Victoria, 1 Park Drive, Bundoora 3083, Australia. [18]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. [19]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. [20]Department of Cell Biology, University of Massachusetts Medical School, Worcester, MA 01655, USA. [21]Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8560, Japan. [22]National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan. [23]Department of Computer Science

†Equal contributors
*Correspondence: m.renfree@unimelb.edu.au; papenfuss@wehi.edu.au
[2]Department of Zoology, The University of Melbourne, Melbourne, Victoria 3010, Australia
[3]Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia
Full list of author information is available at the end of the article

BioMed Central © 2011 BioMed Central Ltd

and Engineering, University of Connecticut, Storrs, CT 06269, USA. [24]Human Genome Sequencing Center, Department of Molecular and Human Genetics Baylor College of Medicine, Houston, TX 77030, USA. [25]Australian Genome Research Facility, Melbourne, Victoria, 3052 and the University of Queensland, St Lucia, Queensland 4072, Australia. [26]Westmead Institute for Cancer Research, University of Sydney, Westmead, New South Wales 2145, Australia. [27]National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. [28]Department of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, NSW 2052, Australia

Published: 30 December 2011

Reference

1. Renfree MB, Papenfuss AT, Deakin JE, Lindsay J, Heider T, Belov K, Rens W, Waters PD, Pharo EA, Shaw G, Wong ESW, Lefevre CM, Nicholas KR, Kuroki Y, Wakefield MJ, Zenger KR, Wang C, Ferguson-Smith M, Nicholas FW, Hickford D, Yu H, Short KR, Siddle HV, Frankenberg SR, Chew KY, Menzies BR, Stringer JM, Suzuki S, Hore TA, Delbridge ML *et al*.: **Genome sequence of an Australian kangaroo, Macropus eugenii, provides insight into the evolution of mammalian reproduction and development.** *Genome Biology* 2011, **12**:R81.

doi:10.1186/gb-2011-12-12-414
Cite this article as: Renfree MB, *et al*.: Correction: Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biology* 2011, **12**:414.