

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Meyer, Braydon Ashley

Title:

The epigenetic landscape of paediatric acute myeloid leukaemia

Date:

2019

Persistent Link:

<https://hdl.handle.net/11343/238756>

Terms and Conditions:

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

The Epigenetic Landscape of Paediatric Acute Myeloid Leukaemia

Braydon Ashley Meyer

ORCID: 0000-0003-2029-888X

Submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Paediatrics

Faculty of Medicine, Dentistry and Health Sciences

The University of Melbourne

September, 2019

Abstract

Paediatric acute myeloid leukaemia (AML) is a cancer of the blood and bone marrow. It is currently one of the leading causes of cancer-related mortality in children. While induction therapy is largely successful in achieving patient remission, the relatively high mortality rate is driven by the large genetic heterogeneity of AML and recurrence of disease. Disease relapse rate is higher than other childhood leukaemias, is fast acting and often chemotherapy resistant.

While much of the genetic contribution to disease has been described, there is still a component of AML pathogenesis that has yet to be discovered. Many of the genetic lesions found in adult AML directly affect epigenetic modifying genes, however this is not the case in children. Despite this, previous research has shown vast epigenetic alteration in paediatric AML. As such it is possible that some of the unexplained pathogenesis in childhood AML can be elucidated by modulation of gene activity via aberrant changes in the most widely studied epigenetic process, DNA methylation (DNAm).

Few studies have comprehensively interrogated the DNA methylome of paediatric AML, nor has the prognostic utility or biomarker potential of DNAm been explored. In this study, we explored the global methylation profile of paediatric AML in comparison to non-leukaemic controls and subtype-dependant and independent biomarkers of disease that may have functional relevance. Furthermore, we described DNAm signatures with potential prognostic utility, to accurately identify predisposition to relapse at diagnosis.

Genome-wide DNAm was interrogated via the HumanMethylation450 BeadChip Array (HM450K) on a cohort comprising of 128 archival and fresh bone marrow tissue sourced from multiple hospitals around Australia. This data was then combined with the TARGET AML cohort comprising of a further 231 bone marrow samples. Targeted replication and validation of findings was undertaken on a reduced cohort using SEQUENOM MassArray EpiTYPER. Bioinformatic and machine learning analyses were undertaken in R.

The findings revealed subtype-independent genome-wide average methylation (GWAM) to be increased in diagnostic samples compared to non-leukaemic controls. This was further verified by differences in the global methylation proxy genes known as *LINE1* and *A/u*. Deeper interrogation of these differences demonstrated wide-spread differential methylation in previously implicated genes in AML pathogenesis including *WT1* and *DGKG*, both of which were validated in an independent cohort. Other genes identified to be differentially methylated included *ZSCAN1*, *REC8* and *IRX1*. Subtype analysis validated previous studies showing inv(16)-specific differential methylation in *MNI* and *MEIS1*. Finally, DNAm was used as the primary feature for a machine learning model designed to predict patient relapse at diagnosis. The final model achieved an area under the curve (AUC) of 94% with correct identification of 91% of all cases involved (F-measure=0.914).

To date, this study represents the largest and most comprehensive insight into aberrant DNAm in paediatric AML. Results have increased our understanding of genes that are differentially methylated and highlight the potential utility of DNAm as a future prognostic biomarker. It is anticipated that these findings will serve as a foundation for future functional studies aimed at delivering truly personalised treatment regimens for children with AML.

Declaration

This is to certify that:

1. The thesis comprises only my original work towards the Doctor of Philosophy except where indicated in the Preface;
2. Due acknowledgement has been made in the text to all other material used;
3. The thesis is less than 100,000 words in length, exclusive of tables, figures, bibliographies and appendices.

Braydon Ashley Meyer
Department of Paediatrics
The University of Melbourne
Parkville, Victoria 3052
Australia

Preface

Contributions

The work conducted in this thesis was completed with the assistance of the following:

Dr Louise Ludlow assisted with writing of the application to the Children's Cancer Centre Tissue Bank (CCCTB) allowing us to access additional paediatric AML samples diagnosed at the Royal Children's Hospital.

A/Prof Daniel Catchpoole & Li Zhou of the Children's Hospital of Westmead, Dr Andrew Moore & Caedyn Stinson of the University of Queensland Diamantina Institute, and Dr Catherine Cole of the Princess Margaret Hospital assisted by supplying additional AML samples to supplement our discovery cohort.

Dr Françoise Mechinaud was the oncologist responsible for clinical input and initial curation of all AML samples diagnosed at the Royal Children's Hospital. Without the meticulous storage of these samples this project would not be possible.

Dr Soheil Meshinchi & Dr Robert Arceci for their curation and leading roles in the TARGET AML Study. Also for making this data publicly available and easily accessible by the greater scientific community.

Dr Benjamin Ong (MCRI) assisted with the spectra acquisition in all MALDI-TOF mass spectrometry experiments used in this study.

Patricia Holtmann (Germany) & Carlotta Migliolo (Italy) both assisted with creation of primers, analysis and replication of findings pertaining to subtype specific findings and ongoing experiments.

Presentations

Poster presentations:

Braydon Meyer, Leah Morenos, Louise Ludlow, Daniel Catchpoole, Andrew Moore, Andrew Elefanty, Paul G. Ekert & Richard Saffery (2015), *The epigenetic landscape of paediatric acute myeloid leukaemia*, Australian Epigenetics Alliance Conference, Hobart, Australia.

Braydon Meyer, Leah Morenos, Louise Ludlow, Daniel Catchpoole, Andrew Moore, Andrew Elefanty, Paul G. Ekert & Richard Saffery (2017), *The epigenetic landscape of paediatric acute myeloid leukaemia*, EMBL Cancer Genomics Conference, Heidelberg, Germany.

Braydon Meyer, Leah Morenos, Louise Ludlow, Daniel Catchpoole, Andrew Moore, Andrew Elefanty, Paul G. Ekert & Richard Saffery (2017), *The epigenetic landscape of paediatric acute myeloid leukaemia*, Australian Epigenetics Alliance Conference, Brisbane, Australia. Awarded 1st place.

Braydon Meyer, Leah Morenos, Louise Ludlow, Daniel Catchpoole, Andrew Moore, Andrew Elefanty, Paul G. Ekert & Richard Saffery (2015), *The application of DNA methylation and machine learning in paediatric acute myeloid leukaemia*, Lorne Cancer Conference, Lorne, Australia.

Oral presentations:

Invited speaker (2016): *The epigenetic landscape of paediatric acute myeloid leukaemia*, Australian Society for Medical research (ASMR) Student Symposium, Melbourne, Australia. Awarded 2nd place and student choice.

Invited speaker (2016): *The epigenetic landscape of paediatric acute myeloid leukaemia*, Murdoch Children's Research Institute Research Students Association Symposium, Melbourne, Australia. Awarded 2nd place.

Invited speaker (2016): *Can DNA methylation and machine learning predict paediatric AML patient outcome?* Murdoch Children's Research Institute, inaugural Work in Progress seminar, Melbourne Australia.

Invited speaker (2017): *Insights into DNA methylation induced pathogenesis of paediatric acute myeloid leukaemia*, Murdoch Children's Research Institute Research Students Association Symposium, Melbourne, Australia. Awarded 1st place.

3-minute thesis oral presentation (2018): *Prediction and pathogenesis - The epigenetic landscape of paediatric acute myeloid leukaemia*, Murdoch Children's Research Institute, Melbourne, Australia.

Short Courses

February 2016 – November 2016, Graduate Certificate in Advanced Learning and Leadership (GCALL), Centre for the Study of Higher Education, Melbourne Graduate School of Research, The University of Melbourne, Melbourne, Australia.

Funding

During my PhD I was supported by the following competitive funding schemes:

- NHMRC project grant APP1045161 (2013) – 5 years
- Victorian Cancer Agency, (TRP16007- 2015)
- MCRI internal project grant, (2017)
- RCH Children’s Cancer Centre Foundation, (2017)
- EMBL Travel Grant, EMBL Heidelberg, (2017)

Ethics Clearance

This study was approved by the royal Children’s Hospital Human Research Ethics Committee (HREC 29140C/D/E/F and HREC 27138F)

Acknowledgements

This has been an absolute journey. Without doubt, the most arduous, time-consuming, demanding journey I have ever been on. Yet, I wouldn't change a single thing about it. Okay, maybe a few things... But it's made me more than just the scientist I am today, but the person I am too. The personal growth I've been through these last four years has been something I didn't think was possible, nor did I expect this would be the environment for it to happen. I guess you could say it altered my epigenetics. But none of this would have been possible without the people I had around me.

Firstly, I would like to thank my supervisors. To Richard, words cannot explain how grateful I am for the opportunity you gave me. But beyond opportunity, you believed in me in times when I wasn't even sure I believed in myself. I would undoubtedly not have made it here without your guidance. It's genuinely something I will *never* forget. To Elizabeth, Andrew, Paul and Don, your contributions to the study at hand, your guidance and philosophy have been a beacon these past few years. It might seem little to you, but it's had a major impact on me, and I cannot thank you enough.

The Epigenetics lab was probably one of best places to have landed. You all made me actually want to come to work. Thank you all for the daily trivia and solid sharn sessions. A special thank you goes out to Sarah and Toby for always being there. From asking me to sit with you on day 1, to the board games, the tram rides home together and the drunken Christmas parties, you have been incredible. To the friends I've made along the way, Becks, Pedro, Jeremy, Blake, Mel, Alex, Rita, Ingrid, Lyra, Jackie, Moni, Camanda, Alena, Kater, Patricia, Carlotta, Reg, Pete, Jake, Dilys, Talmel, Emma, TayTay, Sam, Amy, Rick, Sachi, Ashani, Anna, Bowon & Boris, you have made this place, and my time here, something I'll never forget.

Finally, to my family and my wonderful partner Kiara, I could not have achieved any of this without your support. Thank you for dealing with me when I come home completely over science and writing, for looking after me when things go wrong, but most importantly, for being there to share some of the best times I've had in my life. I couldn't have done this without you.

Table of Contents

Abstract	II
Declaration	IV
Preface	V
Acknowledgements	VIII
Table of Contents	IX
List of Figures	XV
List of Tables	XXI
Abbreviations	XXIII
Chapter 1 Introduction	1
1.1 Cancer	2
1.1.1 Hallmarks of Cancer	2
1.1.2 Comparison of Paediatric and Adult Cancers	3
1.2 Leukaemia	5
1.2.1 Aetiology and Types of Leukaemia	5
1.2.2 Acute Myeloid Leukaemia	6
1.3 Epigenetics & DNA Methylation	24
1.3.1 Introduction	24
1.3.2 DNA Methylation in Cancer: Implications for Prognosis	28
1.3.3 DNA Methylation in Acute Myeloid Leukaemia and its Implication in Patient Prognosis	30
1.4 The Present Study	34
1.4.1 Hypotheses and Aims	35
Chapter 2 Materials and Methods	37
2.1 Materials	38

2.1.1	Equipment	38
2.1.2	Reagents	39
2.1.3	Buffers and Solutions	40
2.2	Study Cohorts & Sample Collection	41
2.2.1	Australian Paediatric Acute Myeloid Leukaemia Cohort	41
2.2.2	Therapeutically Applicable Research to Generate Effective Treatments (TARGET) Study	42
2.2.3	Healthy Bone Marrow Data	42
2.2.4	Combined Analyses & Data Harmonisation	43
2.3	Cohort Demographics	44
2.4	Isolation of Nucleic Acids	46
2.4.1	Patient Sample Pre-Processing	46
2.4.2	Extraction of DNA from Archival Bone Marrow	46
2.4.3	Extraction of DNA and RNA from Fresh/Frozen Bone Marrow	47
2.4.4	Measurement of Yield and Quality	48
2.5	Measurement of DNA Methylation	50
2.5.1	Bisulphite Conversion	50
2.5.2	Illumina Infinium HumanMethylation450 BeadChip Array (HM450k)	51
2.5.3	Primer Design for Locus Specific Methylation Analysis	53
2.5.4	MassArray EpiTYPER (SEQUENOM) Analysis	54
2.6	Data Cleaning and Analysis	59
2.5.5	Overview	59
2.5.6	Illumina Infinium HumanMethylation450 BeadChip Array (HM450k) Analysis	59
2.5.7	SEQUENOM MassArray EpiTYPER (SEQUENOM) Data Cleaning and Analysis	61
2.5.8	Pathway Analysis	61
2.5.9	Additional Statistical Methods	62
2.5.10	Graphical Methods	63

Chapter 3	Interrogation of Global Methylation in Paediatric Acute Myeloid Leukaemia	65
3.1	Rationale	66
3.2	Study Cohort	68
3.3	Genome-wide Average Methylation in Paediatric Acute Myeloid Leukaemia	70
3.3.1	Comparison between Diagnostic, Remission and Healthy Bone Marrow Samples - Discovery & Pooled-Analyses	70
3.3.2	Global Methylation Levels Assessed Using <i>LINE1</i> and <i>Alu</i> between Diagnostic and Remission Bone Marrow Samples in Paediatric AML	78
3.4	Global Methylation Levels in Association with Specific Genetic Lesions	84
3.4.1	Comparison of Global Methylation between AML Subtypes	84
3.4.2	Effects of Loss of Heterozygosity upon Global Methylation in Paediatric Acute Myeloid Leukaemia	87
3.5	Discussion	91
3.5.1	Interrogation of Global DNA Methylation Reveals Differences between Diagnostic and Remission AML Bone Marrow Samples	91
3.5.2	Varying Chromosomal Aberrations Reveal Unique Global Methylation Profiles	94
3.5.3	Limitations & Future Directions	96
Chapter 4	DNA Methylation Profiling of Paediatric Acute Myeloid Leukaemia	100
4.1	Rationale	101
4.2	Study Cohort	104
4.3	Genome-wide DNA Methylation of Diagnostic, Remission & Healthy Samples - Australian Cohort	105
4.3.1	Categorisation of Samples According to Disease State	105
4.3.2	Genes Differentially Methylated in Diagnostic AML Samples	108
4.3.3	Pathway Analysis	114
4.4	Locus Specific Replication of DNA Methylation	116

4.4.1	Validation and Replication of Methylation between HM450K Array and SEQUENOM Assayed CpG Sites	116
4.5	Replication of Differential Methylation in a Combined AML Cohort: A Pooled-Analysis	129
4.5.1	DNA Methylation Profiling to Identify AML Signatures	129
4.5.2	Genes Differentially Methylated in Diagnostic AML Samples - A Pooled-Analysis	132
4.5.3	Pathway Analysis	139
4.6	Identification of Differential DNA Methylation Associated with Patient Relapse	141
4.6.1	Methylation Profiling of Diagnostic and Matched Relapsed Bone Marrow Samples	141
4.6.2	Pathway Analyses	145
4.7	Discussion	147
4.7.1	Interrogation of Genome-Wide DNA Methylation Reveals Differential Signature Associated with Paediatric Acute Myeloid Leukaemia	147
4.7.2	Pathways Showing Epigenetic Disruption Associated with Paediatric Acute Myeloid Leukaemia	150
4.7.3	Limitations and Future Directions	154
Chapter 5	Defining DNA Methylation Signatures of Cytogenetically Distinct Subtypes of Paediatric Acute Myeloid Leukaemia	157
5.1	Rationale	158
5.2	Study Cohort	160
5.3	Distinct Methylation Signatures of AML Cytogenetic Subtypes	162
5.3.1	Defining Patient Subtype through Unsupervised Analysis	162
5.4	Differential DNA Methylation Associated with Chromosome 16 Rearranged Paediatric AML	170
5.4.1	Principal Components Analysis	170
5.4.2	Least Squares Regression Analysis	171

5.4.3	Probe and Gene Specific Differential DNA Methylation Associated with Chromosome 16 Rearranged Paediatric AML	174
5.5	Locus Specific Analysis of Genes Implicated in Chromosome 16 Rearranged AML: <i>MEIS1</i> & <i>MNI</i>	178
5.5.1	Validation and Replication of Methylation between HM450k Array and SEQUENOM Assayed CpG Sites	178
5.5.2	Unsupervised Clustering Analysis of <i>MEIS1</i> & <i>MNI</i> Methylation to Distinguish Chromosome 16 Rearranged Paediatric AML	185
5.6	A Combined Analysis of DNA Methylation Signatures in Chromosome 16 Rearranged Paediatric AML across Two Independent Cohorts	188
5.6.1	Principal Components Analysis	188
5.6.2	Least Squares Regression Analysis	189
5.6.3	Probe and Gene Specific Differential DNA Methylation between Chromosome 16 Rearranged Paediatric AML and Other Paediatric AML Subtypes	192
5.7	Discussion	196
5.7.1	Improved Prognostication in AML by Genome-Wide DNA Methylation Profiling	196
5.7.2	The Methylation Signature of Chromosome 16 Rearranged Acute Myeloid Leukaemia	198
5.7.3	Pathways Showing Epigenetic Disruption in Chromosome 16 Rearranged AML	199
5.7.4	Future Directions & Limitations	201
Chapter 6	The Potential of DNA Methylation as a Predictor of Outcome in Paediatric Acute Myeloid Leukaemia	205
6.1	Rationale	206
6.2	Study Cohort	209
6.3	DNA Methylation Profiling to Predict Patient Outcome in an Australian AML Cohort	211

6.3.1	Using DNA Methylation to Define Patient Outcome	211
6.3.2	Probe & Gene Specific Differential DNA Methylation at Diagnosis Associated with Poor Patient Outcome	215
6.3.3	Using the Top Differentially Methylated Probes to Predict Patient Outcome	217
6.4	Whole-genome DNA Methylation Analysis to Predict Patient Outcome: A Pooled-Analysis	221
6.4.1	DNA Methylation to Define Patient Outcome in a Combined AML Cohort	221
6.4.2	Probe & Gene Specific Differential DNA Methylation Underpinning Poor Patient Outcome: A Pooled-Analysis	224
6.4.3	Comparing the Top Differentially Methylated Probes to Predict Patient Outcome: A Pooled-Analysis	229
6.5	Machine-Based Learning Approaches to Patient Outcome Prediction	234
6.5.1	Finding the Appropriate Machine-Based Learning Algorithm	234
6.5.2	Using Random Forest Analysis to Predict Patient Outcome	237
6.6	Discussion	240
6.6.1	Utility of DNA Methylation Profiling to Predict Patient Outcome in Paediatric AML	240
6.6.2	Machine-Based Learning Approaches to Define Paediatric AML Outcome	242
6.6.3	Epigenetic Disruption in Pathways Associated with Paediatric AML Patient Outcome	243
6.6.4	Future Directions and Limitations	245
Chapter 7	General Discussion	249
	References	257
	Appendices	295

List of Figures

Figure 1.1	Hierarchical view of the haematopoietic cell lineage	6
Figure 1.2	Estimated frequencies of mutations in paediatric acute myeloid leukaemia.....	10
Figure 1.3	Age related differences in mutations and structural alterations in AML.....	12
Figure 1.4	Event free and overall survival of patients with mutually exclusive genetic alterations	17
Figure 1.5	Timeline of clonal evolution from primary to relapsed AML during therapy..	23
Figure 1.6	Waddington's classic epigenetic landscape (Updated).....	24
Figure 1.7	Chromatin structure and location of epigenetic modifications	26
Figure 1.8	A complete pathway for dynamic modifications of cytosine.....	28
Figure 1.9	The aberrant DNA methylation profile found in cancer cells	29
Figure 2.1	Effect of bisulphite conversion upon methylated and unmethylated DNA.....	50
Figure 2.2	Description of the probes on the HM450K DNA methylation array	53
Figure 2.3	Analysis of methylation by base-specific cleavage and MALDI-TOF MS.....	56
Figure 3.1	Inspection of global methylation differences between diagnostic, remission and healthy bone marrow samples through density and scatter analysis	71
Figure 3.2	Inspection of global methylation differences between diagnostic, remission and healthy bone marrow samples in a pooled cohort through density and scatter analysis.....	74
Figure 3.3	Inspection of global methylation through density analysis between diagnostic and relapsed paediatric AML samples.....	76
Figure 3.4	Boxplot of global methylation differences between diagnostic and remission sample restricted to genomic region in a pooled AML cohort.....	77
Figure 3.5	Density plot of average imputed methylation of 2462 Alu CpG sites across the whole genome in diagnostic, remission and healthy bone marrow samples	79

Figure 3.6	Boxplot of differential methylation between patient time points at the global methylation proxy <i>LINE1</i>	80
Figure 3.7	Boxplot of differential methylation between patient time points at the global methylation proxy <i>Alu</i>	81
Figure 3.8	Correlation of average <i>LINE1/Alu</i> locus-specific methylation with global methylation from the HM450k array.....	82
Figure 3.9	Boxplot of average differential methylation between diagnostic and matched relapsed AML samples at the global methylation proxies <i>LINE1</i> and <i>Alu</i>	83
Figure 3.10	Interrogation of global methylation between AML subtypes through density analysis	85
Figure 3.11	GISTIC deletion plot of all diagnostic pooled AML samples.....	88
Figure 3.12	Venn diagram of genes that were both identified in the GISTIC analysis and had hypermethylated HM450k data available	89
Figure 3.13	Inspection of global methylation differences between patients with and without allelic deletion and differences in patient cohort.....	90
Figure 4.1	Unsupervised principal components analysis plot defining diagnostic from remission and healthy adult bone marrow samples	106
Figure 4.2	Heatmap of methylation levels across the 4219 highest ranked DMPs associated with AML status	107
Figure 4.3	Manhattan plot showing significance of all probes after least squares regression analysis between diagnostic and control bone marrow samples	108
Figure 4.4	Circular heatmaps showing DNA methylation levels within specific probes for diagnostic, remission and control samples.....	110
Figure 4.5	Boxplots of the most significantly differentially methylated probes between diagnostic and non-AML bone marrow samples	113
Figure 4.6	REVIGO gene ontology network associated with 4219 differentially methylated probes	114

Figure 4.7	Validation of HM450K methylation using MALDI-TOF MS - <i>WT1</i>	119
Figure 4.8	Regional differential methylation in <i>WT1</i> between diagnostic and remission AML samples.....	121
Figure 4.9	Validation of HM450K methylation using MALDI-TOF MS - <i>DGKG</i>	123
Figure 4.10	Regional differential methylation in <i>DGKG</i> between diagnostic and remission AML samples.....	125
Figure 4.11	Validation of HM450K methylation using MALDI-TOF MS - <i>ZNF876P</i> ..	127
Figure 4.12	Regional differential methylation in <i>ZNF876P</i> between diagnostic and remission AML samples.....	128
Figure 4.13	Unsupervised principal components analysis of diagnostic and remission AML and healthy adult bone marrow samples in a combined cohort.....	130
Figure 4.14	Heatmap of beta values across the 1622 highest ranked differentially methylated probes associated with sample time point in a pooled cohort.....	131
Figure 4.15	Circular heatmaps indicating DNA methylation levels in diagnostic, remission and control samples of all probes within identified genes of interest after pooled-analysis..	134
Figure 4.16	Boxplots of the most significantly differentially methylated probes between diagnostic and non-AML bone marrow samples after pooled-analysis.....	138
Figure 4.17	REViGO gene ontology network associated with 1622 differentially methylated probes identified in a diagnostic vs remission regression analysis	140
Figure 4.18	Unsupervised principal components analysis plot defining diagnostic from relapsed bone marrow samples in a pooled cohort	142
Figure 4.19	Heatmap of beta values across the 50 highest ranked differentially methylated probes between diagnostic and relapse samples by p-value	143
Figure 4.20	Boxplots of the most significantly differentially methylated probes between diagnostic and relapsed bone marrow samples in a pooled-analysis	144
Figure 4.21	REViGO gene ontology network associated with 342 differentially methylated probes identified in a diagnostic vs relapse regression analysis.....	146

Figure 5.1	Heatmap of beta values across the top 250 most variably methylated probes in paediatric AML: Australian and Pooled-Analyses	164
Figure 5.2	t-Distributed Stochastic Neighbour plot of the 250 most variable probes in a pooled cohort to define sample subtype.....	166
Figure 5.3	Variable importance plot of top 100 probes contributing to random forest algorithm determining paediatric AML patient subtype	168
Figure 5.4	Heatmap of beta values across the top 100 probes as defined by variable importance in a random forest algorithm in paediatric AML.....	169
Figure 5.5	Unsupervised principal components analysis plot defining inv(16) paediatric AML from other subtypes in diagnostic bone marrow samples	171
Figure 5.6	Heatmap of beta values across the top 100 differentially methylated probes associated with the inv(16) subtype of paediatric AML.....	172
Figure 5.7	REVIGO gene ontology network associated with 12003 differentially methylated probes	173
Figure 5.8	Circular heatmaps of differential DNA methylation within genes of interest in inv(16) relative to all other subtypes of paediatric AML.....	175
Figure 5.9	Boxplots of the most significantly differentially methylated probes between inv(16) and other subtype diagnostic bone marrow samples	177
Figure 5.10	Validation of HM450K methylation using MALDI-TOF MS - <i>MEIS1</i>	180
Figure 5.11	Methylation of all CpG units measured by MALDI-TOF MS - <i>MEIS1</i>	181
Figure 5.12	Validation of HM450K methylation using MALDI-TOF MS - <i>MNI</i>	183
Figure 5.13	Methylation of all CpG units measured by MALDI-TOF MS - <i>MNI</i>	184
Figure 5.14	Multi-dimensional scaling plot for both <i>MNI</i> , <i>MEIS1</i> and combination of genes to separate inv(16) diagnostic samples from other subtypes of AML.....	186
Figure 5.15	Unsupervised hierarchical clustering and heatmap analysis of 19 CpG sites across <i>MEIS1</i> and <i>MNI</i> to define inv(16) diagnostic samples from other subtypes of AML	187

Figure 5.16	Unsupervised principal components analysis plot defining inv(16) paediatric AML from other subtypes in diagnostic bone marrow samples: A pooled-analysis	189
Figure 5.17	Heatmap of beta values across the top 50 differentially methylated probes associated with the inv(16) subtype of paediatric AML.....	190
Figure 5.18	REVIGO gene ontology network associated with 1744 differentially methylated probes replicated in both discovery and pooled-analyses.....	191
Figure 5.19	Circular heatmaps of differential DNA methylation within genes of interest in inv(16) relative to all other subtypes of paediatric AML after pooled-analysis	193
Figure 5.20	Boxplots of the most significantly differentially methylated probes between inv(16) and other subtypes diagnostic bone marrow samples in a pooled-analysis.....	195
Figure 6.1	Survival and relapse curve analysis based on patient outcome and cytogenetic abnormalities.....	212
Figure 6.2	Unsupervised principal component analysis plot of poor and good outcome cases	213
Figure 6.3	Overlaid density plot of methylation distribution between outcome groups	214
Figure 6.4	Heatmap of beta values for 73 AML patients across the 100 highest ranked probes associated with patient outcome	216
Figure 6.5	Correlation matrix of candidate probes based on disease progression	218
Figure 6.6	Receiver operator characteristic curve of optimum combination of probes to predict patient outcome.....	220
Figure 6.7	Unsupervised principal components analysis plot to define patient outcome in a pooled cohort.....	222
Figure 6.8	Overlaid density plot to interrogate global methylation differences between outcome groups in a pooled cohort.....	223
Figure 6.9	Venn diagram of significantly differentially methylated probes between the discovery and combined cohorts.....	224

Figure 6.10	Heatmap of beta values for 158 AML patients across the 500 highest ranked probes associated with patient outcome	225
Figure 6.11	Correlation matrix and boxplot/line graph defining differential methylation between pooled patient outcome groups in MIRLET7A , PTPRN2 and PAX6	227
Figure 6.12	Correlation matrix of candidate probes after pooled linear regression based on disease progression.....	231
Figure 6.13	Receiver operator characteristic curve of optimum combination of probes to predict patient outcome from pooled linear regression model	233
Figure 6.14	Evaluation of random forest model using 4036 differentially methylated probes through confusion matrix and ROC curve	236
Figure 6.15	Hyperparameter random forest model tuning of ntree and mtry.....	237
Figure 6.16	Density plots of top ten highest ranked probes by mean decreased Gini in the random forest model	239

List of Tables

Table 1.1	Summary and estimated frequency of translocations and genotypes in childhood acute myeloid leukaemia	16
Table 1.2	European LeukaemiaNet (ELN) standardised system for correlation of cytogenetic and molecular genetic data in AML with clinical data	17
Table 2.1	Study equipment.....	38
Table 2.2	Study reagents	39
Table 2.3	Study buffers and solutions	40
Table 2.4	Clinical characteristics and demographic data of all samples used in this study	44
Table 2.5	Composition of full master mix for bisulphite converted DNA amplification PCR.....	57
Table 2.6	Cycling conditions for bisulphite converted DNA amplification PCR.....	57
Table 2.7	Composition of T-cleavage RNase cocktail.....	58
Table 2.8	R packages used for statistical tests in this study.....	62
Table 2.9	R packages used for graphs presented in this study	63
Table 3.1	Clinical characteristics of diagnostic, matched remission and cell sorted, healthy bone marrow samples from the Australian and TARGET AML cohorts.....	69
Table 3.2	Comparison of GWAM between diagnostic and remission AML bone marrow samples restricted by genomic feature	77
Table 3.3	Comparison of GWAM of patient subtypes against control bone marrow samples	86
Table 4.1	Top 100 probes as defined by least squares regression analysis comparing AML to non-AML bone marrow samples.....	111
Table 4.2	Top 100 probes as defined by least squares regression analysis comparing AML to non-AML bone marrow samples in a pooled-analysis.....	137
Table 5.1	Clinical characteristics of diagnostic samples included in analyses.....	161

Table 5.2	Final model evaluation of random forest algorithm using 5-fold, 3-times repeated cross validation to predict paediatric AML subtype at diagnosis (n=171)	167
Table 6.1	Clinical characteristics of good and poor outcome patients from both the Australian (discovery) and TARGET (replication) AML cohorts.....	210
Table 6.2	Fisher's exact test and ROC analysis predictive utility of clinical characteristics for patient outcome	211
Table 6.3	Results of ROC analysis of individual candidate probes.....	219
Table 6.4	Results of Fishers exact test and ROC analysis for clinical characteristics upon patient outcome in a pooled cohort.....	221
Table 6.5	ROC analysis in the combined sample of probes from Australian only ROC analysis	230
Table 6.6	Results of ROC analysis of candidate probes based on disease progression after pooled-analysis	232
Table 6.7	Weighted average model evaluation of machine-based learning algorithms upon TARGET AML dataset using 2046 differentially methylated probes.....	235
Table 6.8	Weighted average model evaluation of machine-based learning algorithms using 5-fold validation of 4036 differentially methylated probes.....	235
Table 6.9	Final model evaluation of random forest MBL algorithm using 5-fold, 3-times repeated cross validation of varying differentially methylated probes (n = 158)	239

Abbreviations

Term	Definition
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
A	Adenine
ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
AUC	Area under the curve
AYA	Adolescent/Young Adult
BMMC	Bone marrow mononuclear cell
BMT	Bone marrow transplant
C	Cytosine
CCCTB	Children's Cancer Centre Tissue Bank
CIMP	CpG island methylator phenotype
CLL	Chronic lymphoblastic leukaemia
CML	Chronic myeloid leukaemia
CMP	Common myeloid progenitor
CN	Cytogenetically normal
CNS	Central nervous system
CNV	Copy number variation
COG	Children's Oncology Group
CpG	Cytosine-phosphodiester bond-guanine
CR	Complete remission
DHS	DNase hypersensitivity
DMP	Differentially methylated probe
DMR	Differentially methylated region
DNA	deoxyribonucleic acid
DNAm	Deoxyribonucleic acid methylation
EFS	Event free survival
ELN	European LeukaemiaNet
FAB	French American British
FGD	Functional genomic distribution
FISH	Fluorescence <i>in situ</i> hybridisation
FPR	False positive rate
G	Guanine
GEO	Gene expression omnibus
GISTIC	Genomic Identification of Significant Targets in Cancer
GMP	Granulocyte macrophage progenitor
GO	Gene ontology

GvHD	Graft vs. host disease
GWAM	Genome-wide average methylation
HM450K	Illumina Infinium HumanMethylation 450 BeadChip Array
HPLC	High performance liquid chromatography
HREC	Human research ethics committee
HSC	Haematopoietic stem cell
inv	Inversion
LOH	Loss of heterozygosity
LSC	Leukaemic stem cell
MALDI-TOF MS	Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry
MBL	Machine-based learning
MDS	Myelodysplastic syndrome
MEP	Megakaryocyte-erythroid progenitor
MPP	Multipotent progenitor
MRD	Minimal residual disease
MSI	Microsatellite instability
MSS	Microsatellite stable
NK	Natural killer
OS	Overall survival
PCA	Principal components analysis
PCR	Polymerase chain reaction
RCH	Royal Children's Hospital
REViGO	Reduce + Visualise Gene Ontology
RF	Random Forest
RIC	Reduced intensity conditioning
RUV	Remove unwanted variance
SAP	Shrimp alkaline phosphatase
SNP	Single nucleotide polymorphism
T	Thymine
t	Translocation
TARGET	Therapeutically applicable research to generate effective treatments
TCGA	The Cancer Genome Atlas
TSG	Tumour suppressor gene
TSS	Transcription start site
t-SNE	t-distributed Stochastic neighbour embedding
U	Uracil
UCSC	University of California Santa Cruz
VAF	Variant allele frequency
Weka	Waikato Environment for Knowledge Analysis
WHO	World Health Organisation

Chapter 1 Introduction

1.1 Cancer

1.1.1 Hallmarks of Cancer

The world's most comprehensive cancer database, GLOBOCAN, reported that there were 14.1 million new incidences, and 8.2 million deaths caused by cancer in 2012 alone (Ferlay, Soerjomataram et al. 2015).

Cancer is thought to arise from a single cell. It is defined by cells that can proliferate indefinitely and circumvent DNA repair mechanisms and apoptosis (Hanahan and Weinberg 2000). Cancer is a disease that is heavily driven by genetic alteration, particularly in adults. However, many other influences such as epigenetics and/or environmental factors play a role in disease aetiology. The main genes affected during carcinogenesis include oncogenes, tumour suppressors and DNA repair genes (Laura, Parsons et al. 2007)

The term 'cancer' is collective for multiple diseases that share the same fundamental abnormalities, presenting as different forms depending on the genes affected, tumour location, age of onset and even environmental exposures. As a result of disruption to the normal cell cycle, tumours can manifest and metastasise which have potential to disturb organ function as well as spread to other areas of the body (Lobo, Shimono et al. 2007) causing significant damage.

There are several types of cancers that can be separated into broad categories (Hoadley, Yau et al. 2014). These include:

- Carcinoma; a cancer of the skin and organ lining
- Sarcoma; a cancer of bone, fat, muscle or other connective tissues
- Lymphoma and myeloma; a cancer that starts in the immune system
- Leukaemia; a cancer of the blood and bone marrow
- Central nervous system (CNS) tumours; which are cancers of the brain and spinal cord.

Hanahan *et al.* have postulated that there are ten ‘hallmarks of cancer’. These hallmarks are associated with sustaining proliferative signalling, resisting cell death, evading growth suppressors, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis (Hanahan and Weinberg 2000). The list has recently been altered to include deregulating cellular energetics, tumour-promoting inflammation, genome instability and mutation & avoiding immune destruction (Hanahan and Weinberg 2011). This is consistent with the idea that cancers require several mutations to fulfil the roles of majority of these hallmarks, rather than a single driver to cause disease (Loeb, Loeb et al. 2003). Despite knowing the fundamental causes of carcinogenesis, there are still aspects of pathogenesis, recurrence and the differences between adult and paediatric cancers that elude us.

1.1.2 Comparison of Paediatric and Adult Cancers

In addition to the fundamental molecular disruptions found across tumour types, there is an added layer of complexity when comparing adult and paediatric cancers. While the phenotypes may be similar according to location, the burden of disease and the resulting treatment on the child has additional negative developmental effects along with standard chemotherapy-associated side-effects compared to that of an adult (Hutter 2010).

It is thought that childhood cancers more commonly develop during an event in embryogenesis or arise due to genetic predisposition (Plon and Nathanson 2005). Studies from The Cancer Genome Atlas (TCGA) have suggested that as few as 2-6 driver mutations may lead to cause cancer in adults (Kandoth, McLellan et al. 2013), however, this analysis generally excludes large-scale structural aberrations that are clearly important in cancer pathogenesis. Even fewer events are implicated in the paediatric setting (Downing, Wilson et al. 2012). Granted, adult cancers are often a result of accumulated genetic alteration due to environmental exposures such as smoking (Hashim and Boffetta 2014) or sun exposure (Green, Whiteman et al. 1999). This is in line with the current paradigm of adults having, on average, more than twice the mutational genetic load compared to children (Tomlinson, Sasieni et al. 2002). It is thought that, due to this 2.4 fold increase in mutational burden found in adult cancers (Chalmers, Connelly et al. 2017), appropriate treatments become more difficult to administer for adults. While treatment response is more variable in adults, children that undergo therapy often suffer

from long term side effects to a much greater extent than adults (LeBaron, Zeltzer et al. 1988). Furthermore, several examples of age-associated genetic variation can be found in many cancers (Jones, Perryman et al. 2012), however this is highly prevalent in leukaemia where cytogenetics plays a key role in pathogenesis and treatment.

1.2 Leukaemia

1.2.1 Aetiology and Types of Leukaemia

Leukaemia is a cancer of the blood and bone marrow, primarily characterised by abnormal proliferation of naïve haematopoietic cells (Hutter 2010). There are several different types of leukaemia, largely defined by the cell lineage affected (myelogenous & lymphoblastic) but also by the rate at which the leukaemia proliferates (chronic & acute).

Chronic leukaemias are characterised by their slow progression and occur more commonly in adults (Pui, Carroll et al. 2011). Chronic myeloid/lymphoblastic leukaemia (CML/CLL) is known to overcrowd areas of the body such as the bone marrow and blood stream, largely through sustained proliferation. Conversely, acute myeloid/lymphoblastic leukaemia (AML/ALL), defined by a rapid expansion of naïve cells and are primarily found in children (Pui, Carroll et al. 2011). They are known to move large quantities of naïve cells into the blood stream causing oxygen and immune cell starvation. This creates symptoms such as an inability to clot and higher susceptibility to infection (Faderl, Talpaz et al. 1999). Furthermore, leukaemia can present with a combination of the above phenotypes and are known as mixed lineage/phenotype leukaemia (Muntean and Hess 2012).

There are several stages of differentiation during haematopoiesis in both myeloid and lymphoblastic lineages (Figure 1.1). Distinct leukaemia subtypes are characterised whereby a block in differentiation occurs, in addition to underlying cytogenetic abnormalities (Faderl, Talpaz et al. 1999).

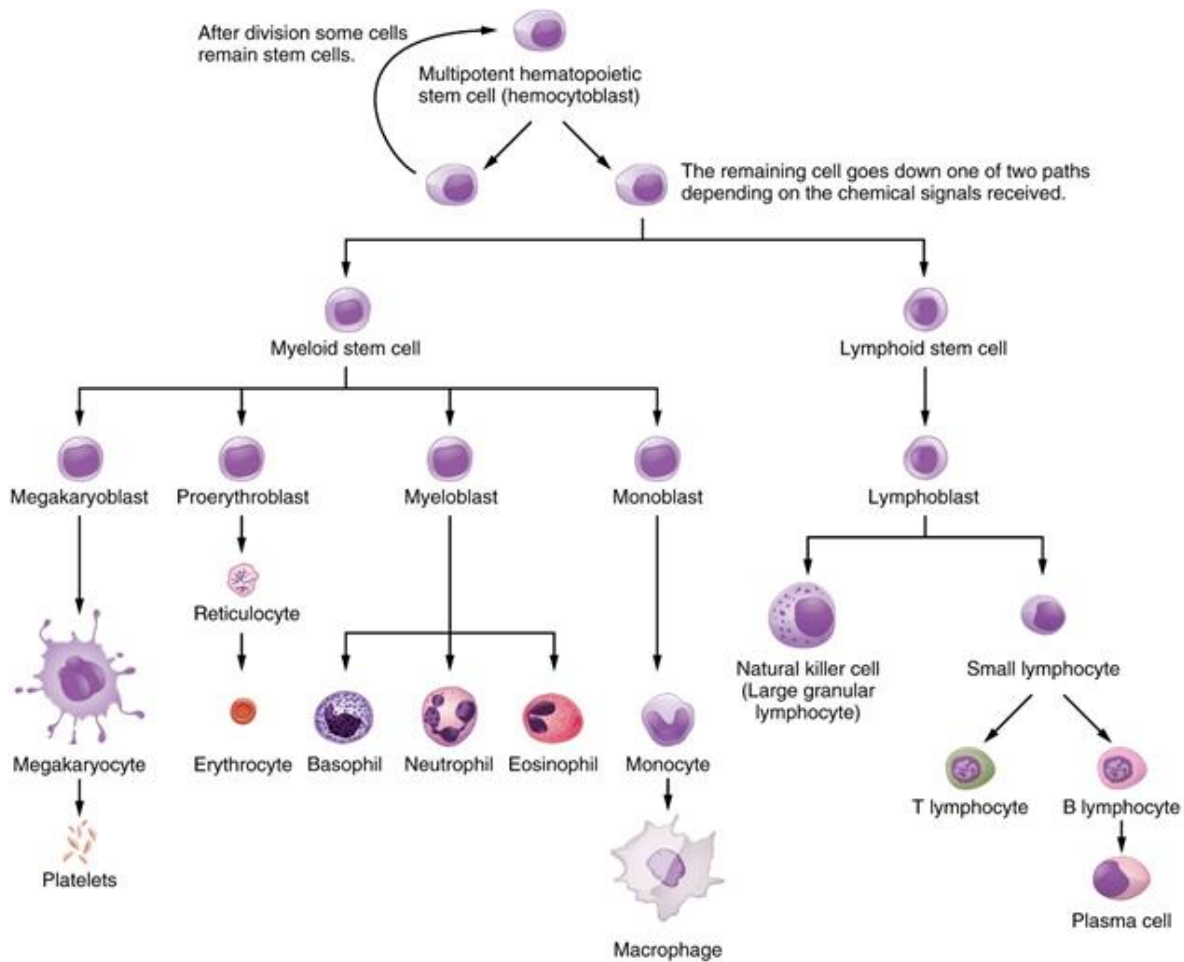


Figure 1.1 Hierarchical view of the haematopoietic cell lineage

Haematopoiesis is the process of multipotent stem cell differentiation into blood and immune cells. Several stages of differentiation, as outlined above, occur in the bone marrow before being released into the blood. Myeloid cell differentiation is to the left, and lymphoid differentiation to the right. The block in differentiation associated with leukaemia is most commonly found at the common myeloid/lymphoid progenitor cell. Figure adapted from (OpenStax 2016).

1.2.2 Acute Myeloid Leukaemia

1.2.2.1 Introduction & Epidemiology

Acute myeloid leukaemia (AML) is a cancer of the blood and bone marrow. It is defined by rapid expansion of abnormal and naïve myeloid stem cells known as leukaemic blasts. This interferes with the production of matured haematopoietic cells required for a functional immune system (Estey 2014). With the loss of haematopoietic function, sufferers of AML can experience fatal infections, excessive bleeding with the inability to clot, or even organ infiltration by AML blasts, particularly in the spleen and liver (Estey and Dohner 2006). Diagnosis of

AML can be made upon immunohistochemical, molecular and cytogenetic testing of the patient's bone marrow. These tests help define AML subtype, but have limited prognostic value, rarely impacting initial treatment approaches (Estey and Dohner 2006).

It has been estimated that paediatric AML makes up 25% of all cases of paediatric leukaemia (Tarlock and Meshinchi 2015), however incidence of mortality remains the highest of all leukaemias at about 40% (Dohner, Estey et al. 2010). Global registries have estimated an incidence rate between 5-8 per million with peak incidence in children aged under one year (Puumala, Ross et al. 2013). Furthermore, GLOBOCAN reports that worldwide, the incidence of AML in 2012 was 4.7 per 100,000 people (age standardised) with an estimated 351,965 new cases (WHO 2014). Males are 1.4 times more likely to be diagnosed compared to females. Unfortunately, these statistics do not take into account the difference in paediatric and adult AML cases, although it should be noted that AML is far more prevalent in adults (WHO 2014). Worldwide mortality rates were estimated at 265,461 cases per year, however, this again may not be reflected in the paediatric setting as there is a higher mortality rate in adults. Puumala *et al.* found that in paediatric AML, 5-year event free survival (EFS) was estimated at 64.3%, but varied greatly between subtypes of disease, ranging anywhere from 22% to 90% EFS.

In the paediatric setting, AML has been identified as the sixth most prevalent cancer in Australian children under 18 years with an incidence of 10 in 100,000 Victorian children in 2012 (Thursfield, Staines et al. 2013). However, more recent statistics from the Cancer Australia Children's Cancer division estimate that leukaemia in general will be the most prevalent paediatric cancer in Australia with over 200 children being diagnosed yearly (Queensland 2014). This suggests that close to 50 of these diagnoses will be AML, making it the fourth most prevalent paediatric cancer in 2017. While numbers pertaining to children are not specified in the report, 5-year EFS is stated to be 55%. It has also been found that rate of relapse in paediatric AML patients varies by trial and ranges between 20% and 41%, with overall survival rates around 75% (Tsukimoto, Tawa et al. 2009; Abrahamsson, Forestier et al. 2010; Rubnitz, Inaba et al. 2010; Cooper, Franklin et al. 2012; Hasle, Abrahamsson et al. 2012; Creutzig, Zimmermann et al. 2013; Pession, Masetti et al. 2013). Studies have found that 5-year EFS following relapse ranges from 16-34% (Stahnke, Boos et al. 1998; Abrahamsson,

Clausen et al. 2007; Rubnitz, Razzouk et al. 2007; Goemans, Tamminga et al. 2008; Gorman, Ji et al. 2010; Sander, Zimmermann et al. 2010; Kaspers 2014). While treatments and survival rates are being continuously improved, the biology behind AML pathogenesis, along with the differences between paediatric and adult AML, is still being discovered.

1.2.2.2 Pathogenesis

Genetic heterogeneity is the main driver of AML pathogenesis. This is primarily the result of hundreds of fusion proteins arising from chromosomal translocations, in combination with point mutations (Alcalay, Orleth et al. 2001). It is thought that most mutations that occur in AML are actually pre-leukaemic (i.e. insufficient in isolation to cause disease), and therefore cooperating mutations/rearrangements are necessary for full leukaemic transformation (Welch, Ley et al. 2012). Most genes related to AML translocations/mutations are associated with one or multiple pathways involved in the aforementioned hallmarks of cancer. The proteins formed from leukaemic translocations are often still functional but act as aberrant transcriptional regulators which interfere with myeloid differentiation and enhance cell survival. However, the loss of function of many different genes also plays a substantial role in disease progression.

The most common variants of AML, making up 40% of all cases combined, include t(8;21), t(15;17), inv(16) and t(9;11) (Figure 1.2). These translocations encode for the *AML1-ETO*, *PML-RAR α* , *CBF β -MYH11* and *KMT2A-AF9* proteins respectively (Alcalay, Orleth et al. 2001). While many other translocations occur, they often still affect one of the genes involved in the above proteins, but with a range of fusion partners. Therefore, it is often believed that the lack of a fully functioning protein is responsible for driving AML, rather than the aberrant functionality of the fusion protein itself (Alcalay, Orleth et al. 2001).

KMT2A

The Lysine Methyltransferase 2A (*KMT2A*, formerly known as *MLL*) encodes a transcriptional coactivator that is essential in the regulation of gene expression during haematopoiesis and early development. This gene is often used as the main example for fusion

protein functionality as there are several different binding partners that can make up this translocation, suggesting alternative binding partners contribute little to disease pathogenesis. Detailed studies in mice have revealed that *KMT2A* fusion proteins arrest differentiation of myeloid precursor cells and may also play a role in immortalisation of haematopoietic precursor cells (Orlovsky, Kalinkovich et al. 2011), however the mechanisms remain largely unknown. Although *KMT2A* translocations are found in all types of leukaemia, it is often believed that the accompanying translocations or mutations define the variability in disease manifestation and subsequent treatment (Marschalek 2015). Recent studies have begun to show that particular breakpoints within the *KMT2A* genes are associated with different types of leukaemia at different ages. This suggests that different functional domains within the gene can evoke different pathways toward carcinogenesis, and ultimately require different treatment regimens (Meyer, Burmeister et al. 2018). Over 135 different breakpoint partners have been identified with *KMT2A* to date, highlighting the complexities associated with understanding AML pathogenesis and informing treatment approach (Meyer, Burmeister et al. 2018).

AML1/CBF β

Translocations involving Runt Related Transcription Factor 1 (*AML1/RUNX1*) and Core-Binding Factor Subunit Beta (*CBF β*) have also been linked with multiple binding partners giving rise to different forms of leukaemia including AML and ALL (Friedman 1999). These two genes are subunits of a core-binding transcription factor that is thought to be the master regulator of genes specific to haematopoiesis (Ho, Alonzo et al. 2009). In contrast to many other poorly understood translocations, the biology of *AML1* and *CBF β* pathogenesis is relatively well understood. Studies in mice show that knock-outs of *AML1* and *CBF β* cause embryonic death (Okuda, Cai et al. 1998), whereas TET-controlled expression of *AML1-ETO* in mice demonstrated some ability to block myeloid differentiation (Rhoades, Hetherington et al. 2000). However, translocations involving *AML1* and *CBF β* are insufficient to drive leukaemia alone and must be coupled with other mutations or chromosomal rearrangements to undergo malignant transformation (Rhoades, Hetherington et al. 2000; Schnittger, Bacher et al. 2007).

PML-RAR α

The *PML/RAR α* fusion gene is also one of the most commonly found fusions in AML. Retinoic Acid Receptor Alpha (*RAR α*) is a member of the steroid-thyroid receptor superfamily of nuclear hormone receptors and is preferentially expressed in myeloid cells (Cole, Verdoni et al. 2015). Promyelocytic Leukaemia (*PML*) is known for its regulation of senescence and apoptosis, as well as functionality as a growth suppressor. *In vitro* studies have suggested that this translocation is sufficient to induce leukaemic transformation; however, penetrance is low and quite delayed (Grisolano, Wesselschmidt et al. 1997). It is known that these fusions are often accompanied by other mutations to genes such as *p21*, *DNMT3A* or *p51* (Viale, De Franco et al. 2009; Alessandra, Angelo et al. 2013; Cole, Verdoni et al. 2015), however the fusion alone is rare and believed to be insufficient to induce leukaemia in humans.

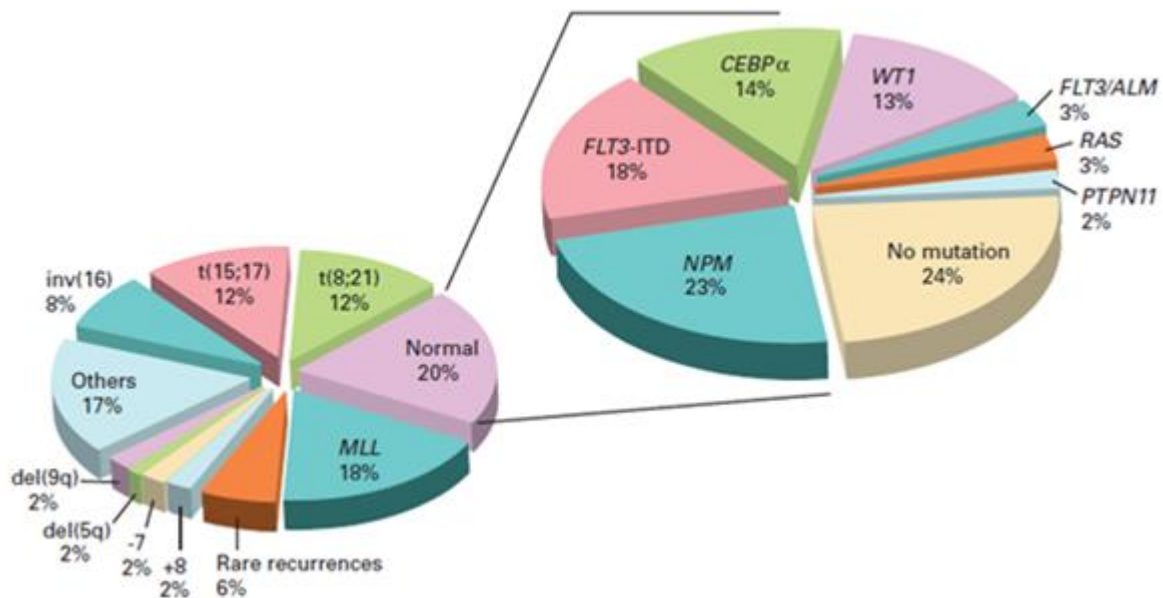


Figure 1.2 Estimated frequencies of mutations in paediatric acute myeloid leukaemia
Genetic abnormalities in childhood AML. The left diagram demonstrates the most common chromosomal alterations. Mutational profile of children for those without chromosomal abnormalities is shown in the right diagram. 80% of children have an AML associated abnormal karyotype and ~15% have at least one known mutation. Figure adapted from (Pui, Carroll et al. 2011)

Mutations in AML

There are many other abnormalities that are associated with AML. These include genomic additions and deletions, neocentromeric chromosomes (de Figueiredo, Mkrtchyan et al. 2009), polysomy, including trisomies and monosomies (particularly trisomy 8 and monosomy 7) as well as an array of rarer and novel translocations. Children with Down syndrome (trisomy 21) are also at high risk of developing AML, with the mechanisms unclear (Crispino 2005). However, some of the more perplexing cases are those that do not appear have any chromosomal abnormalities at all.

Cases of AML are often accompanied by individual point mutations. As previously mentioned, increase in mutational load is vastly different between children and adults, however this is not the case when comparing infants to children and young adults (Figure 1.3). As a result, cytogenetically normal AML is more common as age (and mutational load) increases (Bolouri, Farrar et al. 2017). It is not uncommon for mutations to drive malignancy alone. Genes that are often affected include; *FLT3/ITD*, *WT1*, *CEBP α* , *NPM1*, *RAS*, *KIT*, *GATA1*, *PTPN11*, *DNMT3A*, *IDH1/2* or *TET2* (Pui, Carroll et al. 2011). Mutations in the first four genes of the above list are most commonly found in children, accounting for 70% of all mutations found. Loss of function of the *CEBP α* protein often helps block cell differentiation and has been proven to be highly expressed in myeloid progenitor cells (Wouters, Jorda et al. 2007). *WT1*, originally described as a key oncogene, has been shown to be highly expressed in AML and is also necessary for cell differentiation (Sinha 2014). *FLT3/ITD* mutations are primarily associated with altered signal transduction and increased cell proliferation. Poor prognosis is often related to this mutation (Garg, Nagata et al. 2015). Finally, *NPM1* mutations are still relatively unknown in terms of AML pathogenesis, but they are believed to be passenger mutations that are often acquired in early life (Hefazi, Siddiqui et al. 2015).

Despite being relatively common in adult AML, *DNMT3A*, *IDH1/2* and *TET2* mutations are rarely found in children, appearing closely associated with advanced age, rather than disease (Maegawa, Gough et al. 2014). Unfortunately, there are a small percentage of children (<5%) that have none of the common fusion partners or mutations, suggesting that other mechanisms contributing to disease remain to be uncovered (Pui, Carroll et al. 2011). Considering the

heterogeneity within AML, it is paramount that additional approaches for stratifying patients are developed to help with prognostic approaches and personalised treatment.

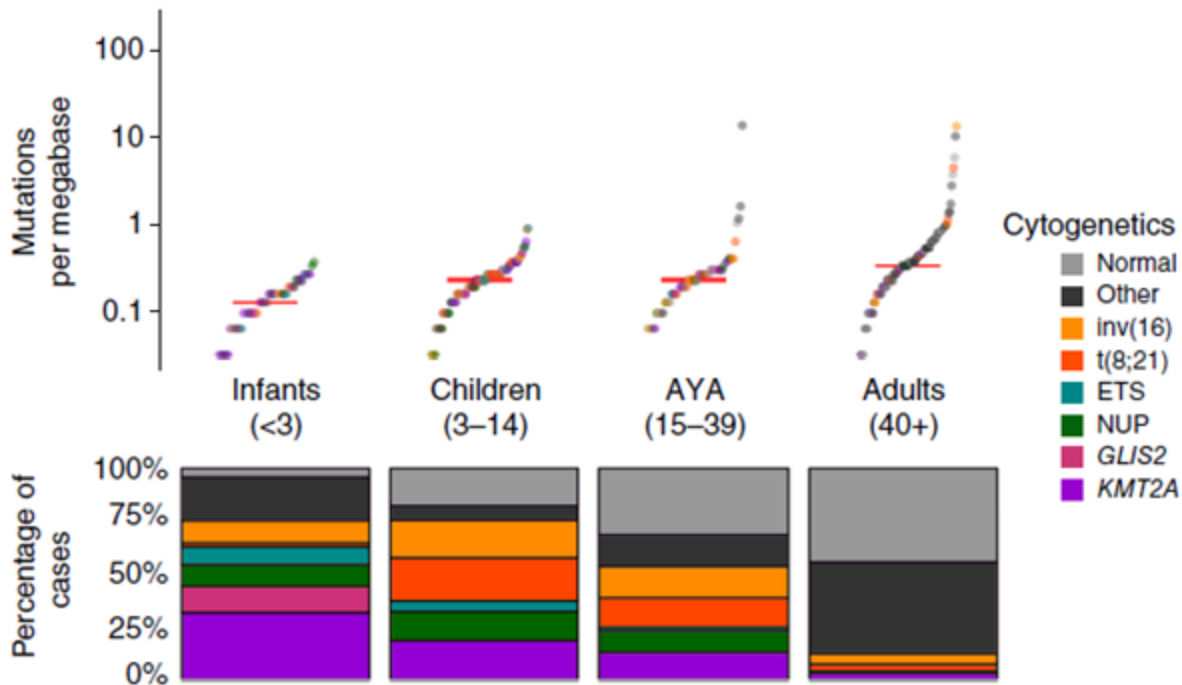


Figure 1.3 Age related differences in mutations and structural alterations in AML

Childhood AML has a low somatic mutational burden that increases with age as defined by the increasing mutations per megabase demonstrated in the top diagram. Structural variation shows an overall shift from *KMT2A* dominant rearrangements in infants towards fewer rearrangements overall and an even spread of structural abnormalities in adult/young adolescents (AYA). This is substantially different compared to adult AML that is dominated by the CN subtype and other genetic alterations. Figure adapted from (Bolouri, Farrar et al. 2017).

1.2.2.3 Classification of AML

Accurate classification of disease is critical to elucidate the most appropriate methods of care, treatment and prognosis. From 1970-2001 AML was stratified based on the cell type from which the leukaemia developed as well as the maturity of the cell. This is known as the French-American-British (FAB) classification (Bennett, Catovsky et al. 1976). However, this classification was of limited value in prognosis and selection of treatment. From 2001, and then later revised in 2008 & 2016, the World Health Organisation (WHO) released their own classification model (with the FAB system integrated into it) based on patient outcome, which has since become the global standard of AML classification (Vardiman, Harris et al. 2002; Vardiman, Thiele et al. 2009; Arber, Orazi et al. 2016). The classification is as follows:

AML with recurrent genetic abnormalities:

- AML with t(8;21)(q22;q22.1); *RUNX1-RUNX1T1*
- AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFβ-MYH11*
- AML with t(9;11)(p21.3;q23.3); *MLLT3-KMT2A*
- APL (M3) *PML-RARα*
- AML with t(6;9)(p23;q34.1); *DEK-NUP214*
- AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3q26.2); *GATA2, MECOM*
- AML (megakaryoblastic) with t(1;22)([13.3;q13.3); *RBM15-MKL1*
- *Provisional entry*: AML with *BCR-ABL1*
- AML with mutated *NPM1*
- AML with biallelic mutations of *CEBPα*
- *Provisional entry*: AML with mutated *RUNX1*

AML with myelodysplasia-related changes

Therapy-related myeloid neoplasms

Myeloid sarcoma

Myeloid proliferations related to Down syndrome

- Transient abnormal myelopoiesis (TAM)
- Myeloid leukaemia associated with Down syndrome

AML not otherwise specified (This includes cases of AML that don't fall into one of the above groups, and is similar to the FAB classification)

- AML with minimal differentiation (M0)
- AML without maturation (M1)
- AML with maturation (M2)
- Acute myelomonocytic leukaemia (M4)
- Acute monoblastic/monocytic leukaemia (M5)
- Pure erythroid leukaemia (M6)
- Acute megakaryoblastic leukaemia (M7)
- Acute basophilic leukaemia
- Acute panmyelosis with myelofibrosis

Acute leukaemias of ambiguous lineage

1.2.2.4 Prognosis

The WHO classification is defined according to prognosis. This allows for treatment to be tailored in regards to dosage and variable extent of patient monitoring post-chemotherapy. However, there are many other strategies implemented to define prognosis for any given patient. Overall, patients aged over 60 years do not fare as well as younger patients, but this may be due to older patients having a higher mutational load, leading to a more complex disease. Having a white blood cell count >100,000/L at diagnosis is linked to a poorer outcome, as is having prior blood disorders or cancer (Amin, Yang et al. 2005).

As stated earlier, there are several different pathogenic pathways that can be affected during malignant transformation. Each of these pathways are affected differently, and therefore have varying effects on patients and require tailored treatments. Pui *et al.* has reviewed several studies and created one of the most comprehensive subtype-specific demographics table (Table 1.1) associated with paediatric AML (Pui, Carroll et al. 2011). In similar fashion, Metzeler *et al.* identified subgroups defined by patient mutation that could distinguish three tiers of patient outcome with increasing severity (Metzeler, Herold et al. 2016). Using statistics such as 5-year EFS and overall survival (OS) in combination with the aforementioned prognostic makers, clinicians can gain a clearer view of the current severity of disease, likelihood of recurrence and potential treatment outcomes (Figure 1.4).

A study performed by Mrozek *et al.* and recently updated by Estey *et al.* show data which further stratified patients into four outcome risk categories based purely on cytogenetic subtyping and mutational status of *NPM1*, *CEBPA*, and *FLT3-ITD*: Favourable, Intermediate I, Intermediate II or Adverse (Table 1.2). This is designed to complement the classification system and has since been integrated into the WHO guidelines (Mrozek, Marcucci et al. 2012; Estey 2016). While this helps in classification of patients, it is noted that uncharacterised genetic lesions may greatly impact outcome. For example, patients with *WT1* mutations, in conjunction with *FLT3-ITD* have a worse EFS than those with *WT1* in isolation (Summers, Stevens et al. 2007). As defined by the European LeukaemiaNet (ELN) standardised system (Table 1.2, (Estey 2016)) this would therefore suggest that a standard Intermediate I defined patient should actually be considered under Intermediate II or Adverse outcome.

Furthermore, Bolouri *et al.* have found that while this classification is generally applicable in the paediatric setting, *FLT3-ITD* and *NPM1* co-mutations only present an unfavourable outcome when also combined with a *DNMT3A* mutation (Bolouri, Farrar et al. 2017). Considering this mutation is rarely found in childhood AML and co-mutation alters patient prognosis, this would suggest that the classification outlined by Estey *et al.* is currently insufficient and could be improved or altered to encompass paediatric AML.

Table 1.1 Summary and estimated frequency of translocations and genotypes in childhood acute myeloid leukaemia

Subtype	Affected Genes	Frequency (%)	5-Year EFS (%)	5-Year OS (%)
Rearrangements				
t(8;21)(q22;q22)	<i>ETO-AML1</i>	12	55-71	75-85
inv(16)(p13;q22)	<i>CBFB-MYH11</i>	8	72-88	75-85
t(8;16)	<i>MOZ-CBP</i>	1	ID	ID
t(15;17)(q22;q12)	<i>PML-RAR</i>	12	71	90
t(1;17)(q23;q12)	<i>PLZF-RARA</i>	Rare	ID	ID
t(1;22)	<i>RBM15-MKL1</i>	2-3	ID	ID
t(6;9)(p23;q34)	<i>DEK-NUP214</i>	Rare	ID	ID
KMT2A	<i>KMT2A</i> (partnered genes)	18		
t(1;11)(q21;q23)	<i>AF1q</i>	3	92	100
t(4;11)(q21;q23)	<i>AF4</i>	2	29	27
t(6;11)(q27;q23)	<i>AF9</i>	5	11	22
t(9;11)(p22;q23)	<i>AF10</i>	43	50	63
t(10;11)(p11.2;q23)	<i>AF10</i>	2	17	27
t(10;11)(p12;q23)	<i>AF10</i>	13	31	45
t(11;19)(q23;p13)	<i>ELL</i> or <i>ENL</i>	4	49	49
t(11;19)(q23;p13.1)	<i>ELL</i> or <i>ENL</i>	4	46	61
t(11;19)(q23;p13.3)	<i>ELL</i> or <i>ENL</i>	3	11	47
t(11;17)(q23;q21)	<i>AF17</i>	2	39	22
Other rearrangements		19	39	54
Gene mutations (NK)		20		
<i>NPM</i>	Nucleophosmin	23	65-80	75-85
<i>CEBPα</i>	Transcription factor	14	70	83
<i>FLT3/ALM</i>	Receptor for <i>FLT3</i>	3	50-60	60-70
<i>FLT3-ITD</i>	Receptor for <i>FLT3</i>	18	< 35	< 35
<i>WT1</i>	Transcription factor	13	22-35	35-56
<i>RAS</i>	Signal transduction	3	ID	ID
<i>PTPN11</i>	Tyrosine phosphatase	2	ID	ID
No known mutation		24	40	50
Poor-risk cytogenetics		< 15	< 40	< 40
Del 5q		1		
-7		2		

Abbreviations: EFS - Event free survival; ID - Insufficient data;

Data extracted from (Creutzig, Zimmermann et al. 2005; Gibson, Wheatley et al. 2005; Ravindranath, Chang et al. 2005; Smith, Alonzo et al. 2005; Meshinchi, Alonzo et al. 2006; Brown, McIntyre et al. 2007; Rubnitz, Razzouk et al. 2007; Franklin, Alonzo et al. 2008; Lange, Smith et al. 2008; Ho, Alonzo et al. 2009; Hollink, van den Heuvel-Eibrink et al. 2009; Ho, Zeng et al. 2010).

Table adapted from (Pui, Carroll et al. 2011)

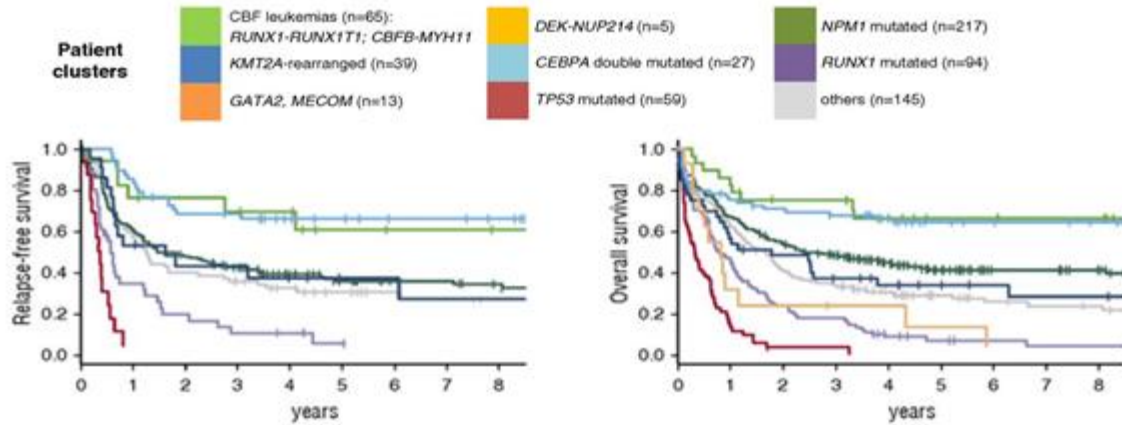


Figure 1.4 Event free and overall survival of patients with mutually exclusive genetic alterations

Kaplan-Meier analysis demonstrating prognostic implications of genetic alterations in patients with adult AML. *TP53* mutations show the most adverse outcome while *CBF* and *CEBPA* rearrangements continue to show favourable outcomes. Figure adapted from (Metzeler, Herold et al. 2016).

Table 1.2 European LeukaemiaNet (ELN) standardised system for correlation of cytogenetic and molecular genetic data in AML with clinical data

Outcome	Subsets
Favorable	t(8;21)(q22;22); <i>RUNX1-RUNX1T1</i> inv(16)(p13.1q22) or t(16;16)(p13.1;q22); <i>CBFβ-MYH11</i> Mutated <i>NPM1</i> without <i>FLT3-ITD</i> (normal karyotype) Double mutated <i>CEBPα</i> (normal karyotype)
Intermediate I	Wild-type <i>NPM1</i> and <i>FLT3-ITD</i> or double mutated <i>CEBPα</i> (normal karyotype)
Intermediate II	Cytogenetic abnormalities not classified as favorable or adverse Mutated <i>FLT3-ITD</i>
Adverse	t(6;9)(p23;q34); <i>DEK-NUP214</i> -5 or del(5q) -7 abnl(17p), abnl(13q) or abnl(3q) Complex karyotype* Mutations in <i>TP53</i> , <i>ASXL1</i> , or <i>RUNX1</i>

*Complex karyotype is defined as three or more chromosome abnormalities in the absence of one of the WHO designated recurring translocations or inversions stated above. Table adapted from (Estey 2016)

Many additional molecular markers have been suggested in order to refine prognostication for specific subtypes or individual genes/karyotypes. The array of biomarkers proposed relate to

different aspects of AML and differ greatly based on the time points at which they show potential utility. Some are more helpful in detecting the presence of minimal residual disease (MRD), while others for predicting relapse or overall survival.

At the forefront of alternative prognostic markers is gene expression. Niavarani *et al.* suggests that using an expression panel of 4 genes (*CD109*, *KIAA0125*, *NGFRAP1* & *ZC3H12C*) associated with high *WT1* expression can predict EFS (Niavarani, Herold et al. 2015). Most importantly this held true in two other adult cohorts and over multiple expression platforms. Individual genes have also shown utility in predicting patient outcome for particular subtypes of AML, usually in association with downstream cellular effects of chromosomal rearrangements. For example, elevated expression of microRNA-261b (Zhang, Wu et al. 2017), microRNA-210 (Tang, Chen et al. 2015), microRNA-19b (Zhang, Lin et al. 2017), *WT1* (Candoni, De Marchi et al. 2017), *DOCK1* (Lee, Chiu et al. 2017), *CD4* (Guo, Atenafu et al. 2017), *CXCR4* (Matsuo, Nakamura et al. 2016), *DUSP5P1* (Zhou, Yin et al. 2015), and reduced expression of *MRP1* (Paprocka, Bielawska-Pohl et al. 2017) and microRNA-215 (Wang, Zhang et al. 2016) all predict an adverse outcome in specific AML subtypes. Conversely, a favourable outcome is conferred through overexpression of *BMPI1P1* (Zhou, Zhai et al. 2016), microRNA-181 (Guo, Luan et al. 2017), microRNA-370 (Lin, Wang et al. 2015), *NEDD9* (Pallares, Hoyos et al. 2017) and *ABR* (Namasu, Katzerke et al. 2017).

Genetic polymorphisms can also be utilised to help predict a patient's outcome. For example, Niavarani *et al.* showed that the *WT1* variant rs2234593 is a prognostic marker in normal karyotype leukaemia for EFS and OS (Niavarani, Horswell et al. 2015). Similarly, Zheng *et al.* and Hou *et al.* showed that AML patients harbouring an *SF3B1*, *U2AF1* or *SRSF2* mutation have a much shortened EFS and OS (Hou, Liu et al. 2016). Furthermore, myelodysplastic syndrome (MDS) patients have an increased risk of AML transformation (Zheng, Zhan et al. 2017). Stengel *et al.* presented research that suggests loss of wt-*RUNX1* showed a significantly shorter OS compared to multiple *RUNX1* mutations and even more so for only a single *RUNX1* mutation. Furthermore they suggested that wt-*RUNX1* loss was also associated with increases in other mutations such as *ASXL1* (Stengel, Kern et al. 2017). Recent studies by Papaemmanuil *et al.* have shown that there are several other subgroups of patients that can be defined by driver mutation analysis. Specifically, chromatin-spliceosome and *TP53* mutations

conferred a poor outcome, especially when coupled with previously aforementioned, class-defining, driver mutations (Papaemmanuil, Gerstung et al. 2016).

Other measurable factors can also be used to infer prognosis. Large cohort studies including that of Rubnitz *et al.* showed that combinations of guidelines set by the FAB and WHO systems, like karyotype stratification, subtype, white blood cell count as well as other factors such as length of remission and sex are all associated with EFS and complete remission (Rubnitz, Razzouk et al. 2007). Recent advances in sequencing technologies have allowed us to interrogate the genome in ways that were once unimaginable. Nibourel *et al.* recently used a SNP array to find copy-number variations (CNVs) that can predict patient outcome and drug resistance while delineating novel adult AML risk groups (Nibourel, Guihard et al. 2016). Body mass index (Li, Chen et al. 2017), rate of blast cell clearance during early induction chemotherapy (Nakano, Fujiwara et al. 2015), and bone marrow T-cell percentage (Ismail and Abdulateef 2016) have all been shown to have prognostic implications in AML. However, research shows that returning an MRD-negative result during and after consolidation therapy is still the most predictive method for predicting EFS and OS (Ravandi, Jorgensen et al. 2016).

Given the heterogeneity of the condition, it is unlikely that a single prognostic test will show utility across every permutation of every subtype in AML. As it currently stands, our best forms of prognostication come from knowing the core driver alterations that cause the leukaemia along with accompanying mutations and clinical variables. However, with the introduction of cost-effective high throughput sequencing technologies, coupled with machine-based learning (MBL) approaches, we may soon be able to more accurately predict outcome. Most importantly, these approaches will help us better understand the pathogenesis of each individual patient and allow us to treat them accordingly; often referred to as a personalised or precision approach to treatment.

1.2.2.5 Treatment

While the prognosis for children with AML has improved significantly in recent decades, treatment options have remained fairly static. The majority of patients under 60 years undergo two main cycles (Rasche, Zimmermann et al. 2018); the first being induction therapy and the

second, consolidation (intensification) therapy. Induction consists of cytarabine for 7 days and either daunorubicin or idarubicin for 3 days, but this can change based on disease severity. Induction is often repeated 2-3 more times to ensure that all blast cells have been removed from the bone marrow. At this point, most children will have reached clinical remission, that is <5% blast count within the bone marrow. Consolidation therapy consists of between 2-4 cycles of high dosage cytarabine, but this may also vary based on clinical advice (Rasche, Zimmermann et al. 2018).

Several other treatments are available as alternatives if patients are unresponsive or have become immune to standard therapy. These are the first wave of personalised treatments in AML as they are very specific to particular subtypes or mutations associated with the disease. In 2017, several drugs were newly approved to treat AML. Examples include midostaurin, a tyrosine kinase inhibitor used primarily in *FLT3-ITD* mutated patients. This has been very effective in younger patients and has been shown to extend EFS and OS dramatically (Stone, Mandrekar et al. 2015). Similarly, crenolanib also targets *FLT3-ITD* (used to treat refractory or relapsed leukaemia) and also helps alleviate myelosuppressive side effects (Randhawa, Kantarjian et al. 2014). CPX-351 is a formulation of cytarabine and daunorubicin designed to circumvent chemotherapy-resistant cases, recently approved for high risk patients (Lancet, Cortes et al. 2014; Stein and Tallman 2015). Immunotherapeutic options such as gemtuzumab has shown to be effective at improving OS across ages, although it does seem to be ineffective against high risk leukaemia (Löwenberg, Beck et al. 2010). Hypomethylating agents such as azacytidine have already been approved with alternative options for take home use and potential as a maintenance therapy are being explored (Garcia-Manero, Gore et al. 2011). The recently developed venetoclax is a BCL-2 inhibitor which, when combined with hypomethylating agents, achieves a 75% complete remission rate in patients that are deemed unfit for conventional induction therapy (Konopleva, Pollyea et al. 2016).

Stem cell or bone marrow transplants (BMT) are often used post-treatment of AML. There are two main types; allogeneic and autologous. Studies have suggested that allogeneic transplantation is far superior in terms of overall survival than autologous transplantation or intensified chemotherapy, but autologous transplants work occasionally and should be considered if allogeneic transplants are not an option (Woods, Neudorf et al. 2001). The

decision to transplant is usually at the discretion of the clinical team and varies between institutions. Some hospitals choose to transplant patients after completion of therapy as an added measure to prevent relapse as well as an immediate boost to immune function (Locatelli, Masetti et al. 2015). However, this decision is weighted heavily on factors such as karyotype and white blood cell count, as the karyotype may give a favourable prognosis to the patient, and thus the transplant may be unnecessary (Yanada 2015). Whether to transplant is still a fairly heavily debated topic primarily due to the risk of graft versus host disease (GvHD). This may occur after transplantation due to the donated cells recognising the host's cells as foreign and subsequently attacking them. GvHD effects range from mild acute cases of skin rash to life threatening liver disease (Ferrara, Levine et al. 2009). Bone marrow transplants are often used as a last resort should conventional therapy prove ineffective.

While treatment is similar for children and adults, long-lasting effects are usually more severe for children. Toxicity of chemotherapy is often exacerbated when exposure occurs during development, often in association with later life complications and relapse. Chemotherapy from others cancers has been known to cause secondary, or treatment-related AML (Hulegardh, Nilsson et al. 2015). Designing treatments that avoid these problems is paramount for better personalised medicine given that prognosis post-relapse remains poor.

1.2.2.6 Relapse

Relapse is the major cause of death in AML. While those in the paediatric setting generally respond better to treatment than adults, the rate of relapse differs little, suggesting treatment regimens may play a role in recurrence. The most prevalent and well-studied concept for relapse is that of clonal evolution (Ding, Ley et al. 2012). Over time, subpopulations of cells with additional mutations arise, often in association with enhanced stemness or even chemotherapy resistance (Figure 1.5). While these fractions may be a minority of the primary tumour, they are selected for during induction therapy and become the primary clone upon relapse.

It is believed that the leukaemic stem cells (LSCs) reside at the apex of the cellular hierarchy, can self-renew and stop division in association with chemotherapy resistance. Such cells are

more likely to be maintained through chemotherapy and are the likely source of relapse in AML (Thomas and Majeti 2017). Additionally, recent studies have shown that in particular subtypes of AML pre-leukaemic events often stay dormant many years after complete remission, maintaining the ability to re-initiate the process of malignancy through genomic instability and/or additional mutations (Hirsch, Zhang et al. 2016). This is evidenced in the clinical presentation of relapsed AML where patients show accumulated mutational load and often show a lack of response to chemotherapy that was originally used to achieve remission.

This is consistent with studies in *FLT3/ITD* patients where the predominant subclone found at relapse was originally present at diagnosis in much smaller numbers (Bachas, Schuurhuis et al. 2015). However, it is important to note no evidence of a relapse clone was found in the original diagnostic sample in 30% of cases, suggesting that secondary malignancies can arise spontaneously. Furthermore, exome sequencing of the Children's Oncology Group (COG) TARGET-AML cohort showed that while there was still variety in mutations between diagnosis and relapse, variants with high variant allele fraction (VAF) were likely shared across a clonal structure, making them key cells for therapeutic targeting (Farrar, Schuback et al. 2016).

Unlike diagnostic samples, the additional mutational load present at relapse often limits the utility of prognostic biomarkers (Kim, Seol et al. 2016). Considering the devastating effects of relapse, especially in children, the need for better methods to monitor MRD, predict relapse and treat stemness in AML are urgently required (Shlush, Mitchell et al. 2017).

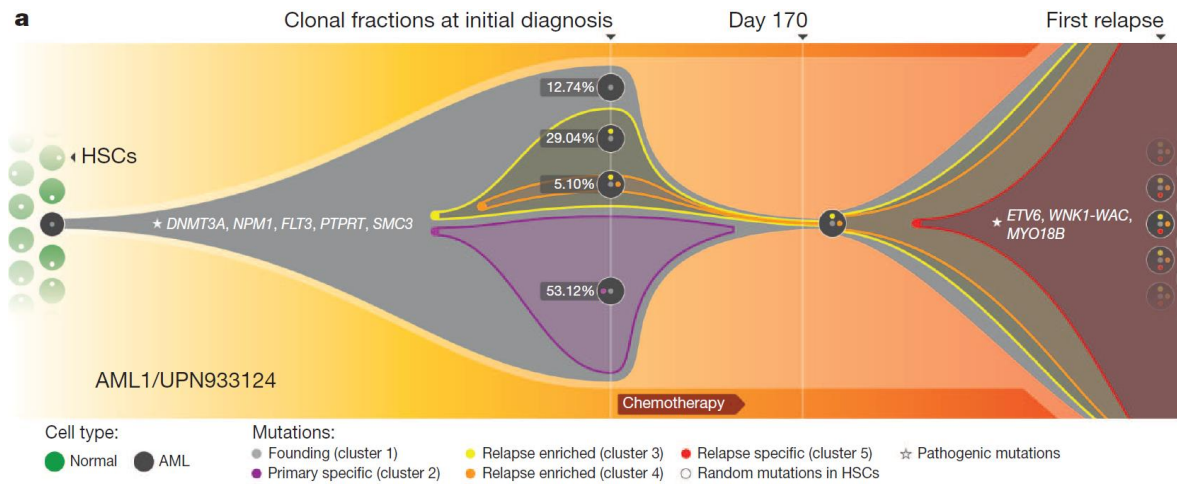


Figure 1.5 Timeline of clonal evolution from primary to relapsed AML during therapy

Primary tumour containing driver mutations, followed by several subpopulations with additional mutations which later become the dominant clone by relapse, acquiring further mutations post relapse. Figure adapted from (Ding, Ley et al. 2012)

1.3 Epigenetics & DNA Methylation

1.3.1 Introduction

The term ‘epigenetics’ was initially coined in 1942 by Conrad Waddington (Goldberg, Allis et al. 2007) to explain the complex relationship between genes, their products, and how they interact to bring about the human phenotype. This could be extended to how an entire cell functions and grows throughout the process of differentiation (Figure 1.6). However, the definition of ‘epigenetics’ has now evolved and varies across areas of research (Bird 2007). Generally speaking, epigenetics (literally meaning ‘above DNA’) can be thought of as changes in gene function that cannot be explained by changes in the underlying DNA sequence in isolation (Greer and McCombe 2012).

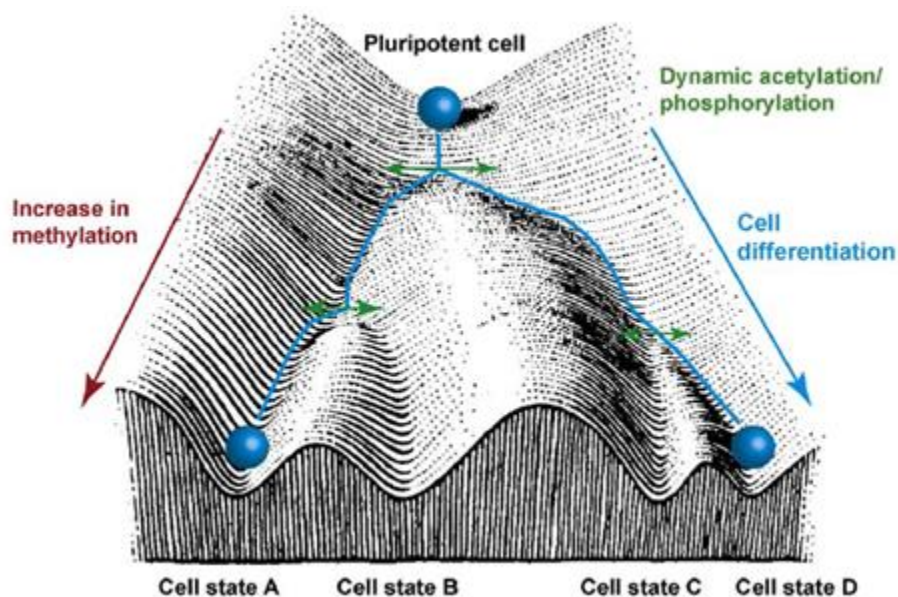


Figure 1.6 Waddington's classic epigenetic landscape (Updated)

This model was proposed around the same time as the coining of the term ‘epigenetics’. The concept being that pluripotent stem cells can be differentiated into any mature cell as deemed necessary by epigenetic modification during a cell's lifetime (Goldberg, Allis et al. 2007). Figure adapted from (Barth and Imhof 2010)

A complex interplay of epigenetics, genetic variation and the environment is believed to be the main determinant of phenotype (Tammen, Friso et al. 2013). It is well-established that there are links between environmental exposures, (e.g. place of birth, diet and lifestyle) and overall health (Flores, Wolschin et al. 2013). As a result, it is believed that epigenetic modification is

the intermediary of environmental exposure and regulation of gene expression (Tammen, Friso et al. 2013). Changes in gene function might be accredited to an array of different epigenetic mechanisms such as histone modification, small non-coding RNAs, and the most widely studied mechanism of epigenetics, DNA methylation (Eccleston, DeWitt et al. 2007). All have been shown to play a role in disease pathogenesis via altered gene expression.

The diploid human genome consists of over six billion nucleotides and needs to be packaged in such a manner that can be contained within a single cell. Epigenetic mechanisms are an integral part of the packaging process, not only to be able to access the DNA, but to store it efficiently.

1.3.1.1 Histone Modification

Nucleosomes are made up of two copies of the core histones H2A, H2B, H3 and H4 to create an octamer. The DNA double helix is then wrapped around these octamers twice. A collective of nucleosomes is known as chromatin and can be found in two forms; open and closed (Figure 1.7). The closed form of nucleosomes are packaged together tightly (known as heterochromatin) to block transcription of DNA. However, the open form (known as euchromatin) can be readily transcribed, repaired or replicated in its accessible state (Kouzarides 2007).

There are many different post-translational modifications of histones, primarily on the amino terminal tail, and many of their functions are still being defined. Most widely studied modifications include acetylation, phosphorylation and methylation. These modifications are variously designed to weaken the bonds between the DNA and the histone, or weaken the bond between linked nucleosomes resulting in changes to the conformation of chromatin (Bannister and Kouzarides 2011). Many studies have shown specific cell types have unique histone modification patterns defining particular chromatin states at promoters to give rise to cell specific gene expression patterns (Heintzman, Hon et al. 2009).

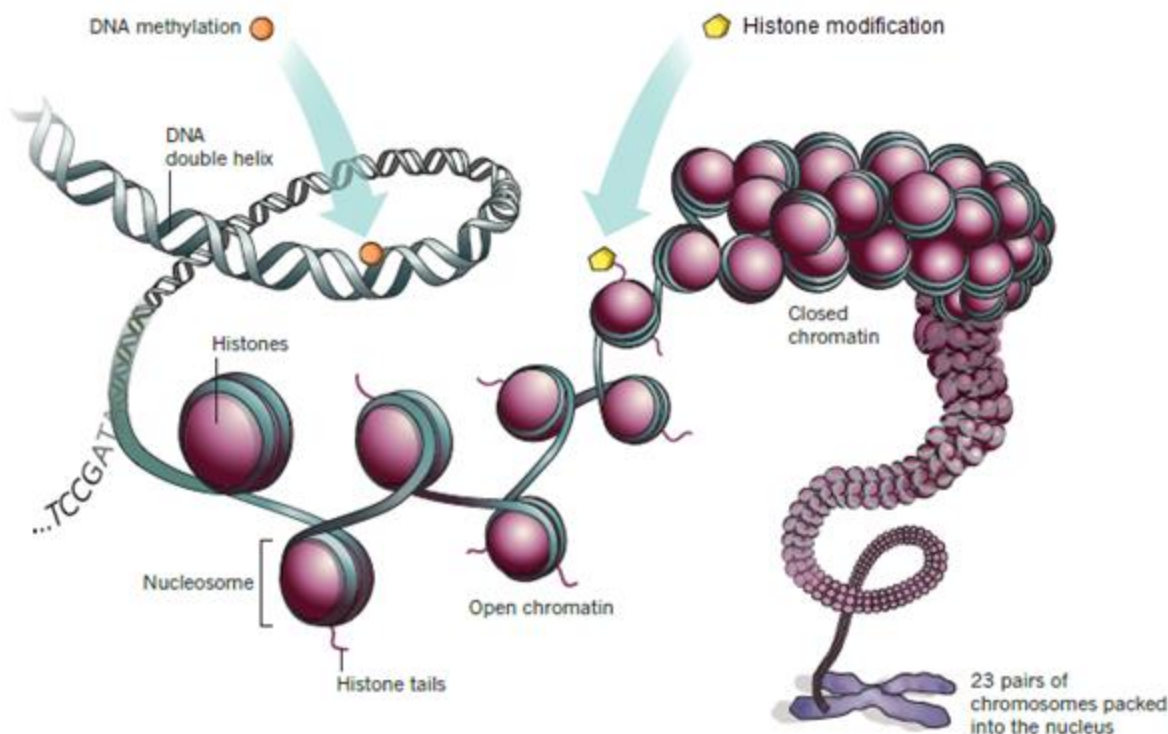


Figure 1.7 Chromatin structure and location of epigenetic modifications

Histone modification is used to control conformation of chromatin complexes, while DNA methylation can be used to modulate gene expression. Figure adapted from (Marx 2012)

1.3.1.2 DNA Methylation

DNA methylation (DNAm) the most-widely studied epigenetic mechanism. This covalent modification occurs primarily at CpG dinucleotides or CpG sites, consisting of a cytosine (C) and guanine (G) base joined by a phosphodiester bond (Bjornsson, Fallin et al. 2004). Methyl groups are covalently added to the fifth carbon of the cytosine base to produce 5-methylcytosine (5mC). CpG sites are underrepresented across the genome as a whole, with around 20% less than the expected occurrence by chance (Consortium 2001). In contrast, they are found in high abundance within the promoter region (approximately 75%) of many genes (Ehrlich, Gama-Sosa et al. 1982). These clusters are known as CpG islands and are often hypomethylated, a state that is generally associated with open chromatin and increased transcription. Conversely, hypermethylated CpG sites found within the regulatory regions of a gene are often associated with tightly packed chromatin and decreased transcription. This is often referred to as methylation-induced gene silencing and is a feature of many cancers,

including AML (Mill and Heijmans 2013). However, in some cases DNAm is a consequence of silenced gene expression, rather than the cause (Daura-Oller, Cabre et al. 2009).

Methylation of CpG sites may prevent transcription factors from binding to the underlying DNA both in a physical sense, and by promoting histone modification through proteins binding to the methyl groups (Jaenisch and Bird 2003). In order to do this, these alterations need to be dynamic. *De novo* DNAm is performed by DNA methyltransferase 3a and 3b (*DNMT3a/DNMT3b*) (Okano, Bell et al. 1999) and is maintained through cell division by DNA methyltransferase 1 (*DNMT1*) (Law and Jacobsen 2010). Demethylation occurs through the recruitment of the ten-eleven translocation (*TET*) family of enzymes that convert 5mC into 5-hydroxymethylcytosine (5hmC) which can subsequently be lost during cell division, thymine DNA glycosylase (TDG)-mediated base excision repair or iterative oxidation (Figure 1.8). Through the cyclical use of these enzymes, DNAm is able to dynamically respond to signals which mediate optimal levels of gene transcription (Kohli and Zhang 2013).

While DNAm plays an important role in regulation of transcription, aberrant methylation is implicated in disease development and maintenance. Examples of aberrant DNAm maintenance include embryonic development, chromatin structure, X chromosome inactivation, genomic imprinting and chromosomal stability (Robertson 2005). Should the genes involved in these processes be disturbed (i.e. mutation or suppression), disease such as Beckwith-Wiedemann, Prader-Willi and Angelman syndromes can arise (Jin and Liu 2018). Furthermore, aberrant DNAm is intrinsic to the development of many cancers (Ehrlich 2002).

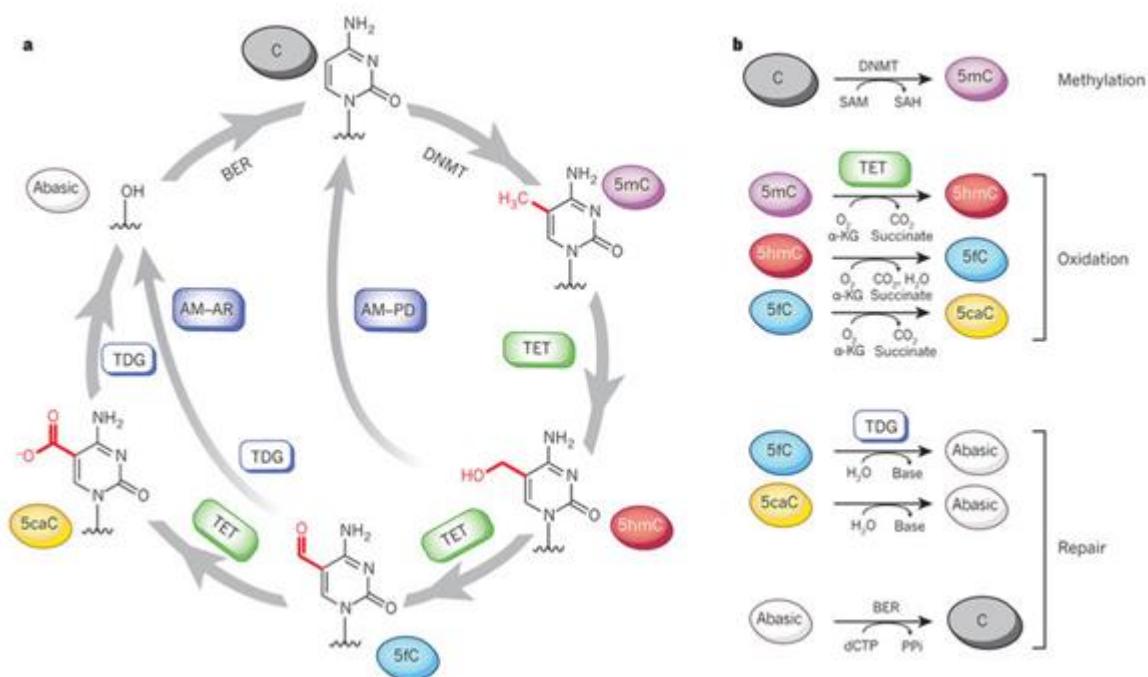


Figure 1.8 A complete pathway for dynamic modifications of cytosine

5mC bases are introduced through the DNA methyltransferase (*DNMT*) enzymes and can be iteratively oxidised through the ten-eleven translocation (*TET*) enzyme family. Furthermore, utilising Thymine DNA Glycosylase (*TDG*), 5-hydroxymethylcytosine (5hmC) can be excised and the cytosine can return to an unmethylated state. Figure adapted from (Kohli and Zhang 2013). Abbreviations; AM - Active modification; AR - Active restoration; PD - Passive dilution; 5fC - 5-formylcytosine; 5caC - 5-carboxylcytosine

1.3.2 DNA Methylation in Cancer: Implications for Prognosis

The DNAm profile in cancer is substantially altered relative to disease-free cells. This includes genome-wide hypomethylation (reduced global 5mC) in addition to promoter-specific hypermethylation (Portela and Esteller 2010). The areas of induced hypomethylation, usually within the body of a gene, intronic and/or within repeat sequences, have been shown to cause chromosomal instability leading to translocations and gene disruption. In contrast, promoter hypermethylation, often in conjunction with other aforementioned epigenetic modifiers, can contribute to carcinogenesis through gene silencing (Feinberg, Ohlsson et al. 2006) (Figure 1.9). These promoter regions subject to hypermethylation are highly variable between different forms of cancers and are enriched for tumour suppressor genes with a role in DNA repair, apoptosis, angiogenesis and more (Esteller 2005; Esteller 2008; Sandoval and Esteller 2012). It is therefore hypothesised that the resulting aberrant pathway signalling cascade can contribute to the pathogenesis of cancer by providing an unregulated environment for their development

and proliferation (Feinberg, Ohlsson et al. 2006). While this disruption to the methylome is found in established cancers, it can also be found prior to carcinogenesis (Sandanger, Nøst et al. 2018). This suggests that DNAm plays a potential causal role in cancer development, with potential utility as a biomarker for prognostic approaches (Belinsky, Nikula et al. 1998).

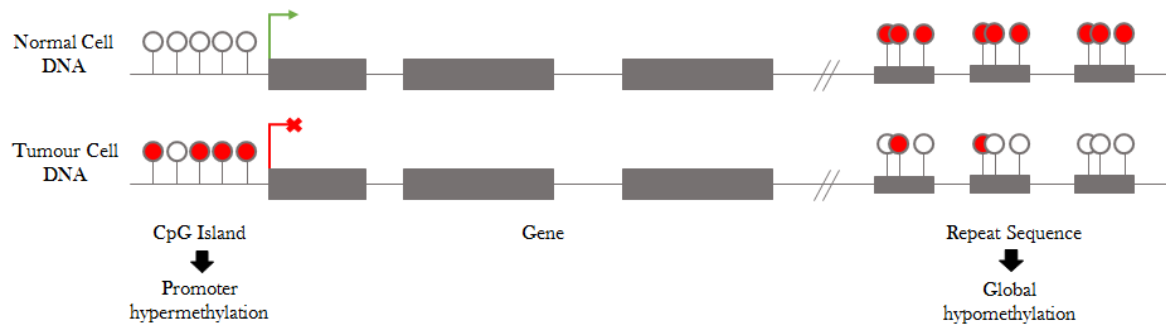


Figure 1.9 The aberrant DNA methylation profile found in cancer cells

Hypermethylation is found in high abundance at CpG sites near promoter regions of tumour suppressor genes while global hypomethylation is found at repeat sequences. White lollipop is non-methylated; red lollipop is methylated.

Interrogation of DNAm has been used in a variety of cancer studies, revealing that aberrant DNAm is a key contributor to all forms of carcinogenesis. Studies have recently shown functional effects in terms of gene silencing, such as in increased activation of the WNT/ β -catenin signalling pathway by silencing of *SFRP1* in colon cancer (Jones and Baylin 2007) as well as activation of the Jak/STAT pathway by silencing *SOCS1* and *SOCS3* in liver and lung cancer (Esteller 2005). More recent, large scale studies have suggested that *EPHA7*, *MGMT* and *MLH1* are all hypermethylated in colorectal patients, *RARB* in prostate tumours and *CASP8* in glioblastoma (Vidal, Sayols et al. 2017). Furthermore, Vidal *et al.* suggest that stem cell clones harbouring larger scale hypermethylation are positively selected for in clonal evolution of tumours (Vidal, Sayols et al. 2017). The fact that aberrant DNAm is found in all cancers raises the possibility of developing clinically relevant methylation biomarkers.

Using DNAm for prognostication is a relatively new field. Acute lymphoblastic leukaemia patients with undefined chromosomal abnormalities can be appropriately subtyped using DNAm as a biomarker (Nordlund, Bäcklin et al. 2013). Using a separate sample set the same authors also found thousands of CpG sites to be hypermethylated when comparing diagnostic to relapsed samples, and even more sites when comparing diagnostic to healthy, control patient

blood. There were also several sites associated with enhancer regions which were directly correlated with gene expression, giving further utility to this test (Nordlund, Bäcklin et al. 2013)

The same principles have been applied in other cancers. In what is perhaps the most iconic use of DNAm being translated in a clinical setting, Capper *et al.* published DNAm array data to predict types of brain tumours based on specific methylation profiles. By releasing this classifier to the public, they obtain further clinical data which retrains the model, making it more robust over time (Capper, Jones et al. 2018). This model has been used to inform clinicians of patient brain tumour subtype in instances where pathology techniques have failed. Similar studies have been undertaken by Schwalbe *et al.* and Danielsson *et al.* who have used a cancer panel methylation array to predict patient outcome as well as further sub-classify medulloblastoma and paediatric brain tumours respectively (Schwalbe, Williamson et al. 2013; Danielsson, Nemes et al. 2015). Furthermore, Sturm *et al.* found that subtypes of the primitive neuroectodermal tumours of the central nervous system (CNS-PNETS) could be further stratified into four different groups all with their own unique recurrent genetic alterations and distinct histopathological and clinical features (Sturm, Orr et al. 2016). There is still much to learn in terms of its true prognostic and pathogenic impact of DNAm in many cancers, especially paediatric AML where few DNAm studies have been published.

1.3.3 DNA Methylation in Acute Myeloid Leukaemia and its Implication in Patient Prognosis

In a notable research conducted by Li *et al.* in AML, they postulate that epigenetic and genetic heterogeneity occur with distinct kinetics, suggesting that both methods can co-create their own independent means of disease progression and clonal evolution (Li, Garrett-Bakelman et al. 2016). Several groups of patients were defined; some having a more epigenetically-driven disease, while others had a more genetic or mixed progression. With each component contributing individually to disease, this creates an added layer of complexity in regards to how to treat a patient that develops disease through different processes (Li, Garrett-Bakelman et al. 2016).

1.3.3.1 Methylation in Adult AML

Most large scale epigenetic studies have been undertaken in adults (See review; (Schoofs, Berdel et al. 2014)). Many of these studies found differential methylation in genes that are associated with both methylation pathways and AML pathogenesis, suggesting widespread methylation dysregulation. Several other studies have associated different patterns of methylation, primarily global hypomethylation, with specific cytogenetic subtypes (Schoofs, Berdel et al. 2014). Furthermore, Kelly *et al.* have shown that a CpG island methylator phenotype (CIMP, which suggests heightened levels of promoter methylation), confers a longer OS to those that test positive (Kelly, Kroeger et al. 2017). They discovered two demethylator phenotypes which were associated with both good (inv(16), t(8;21), t(15;17)) and poor (*NPM1*, *FLT3* and *DNMT3A* mutations) OS respectively, suggesting that epigenetic modification is distinct between, but similar within, cytogenetic subgroups (Kelly, Madzo et al. 2018). Overexpression of *BCAT1* has recently been implicated in AML as downstream effects cause global hypermethylation events through altered *TET* function, akin to harbouring the common AML mutation of *IDH*. Furthermore, high levels of *BCAT1* was associated with stem cell signatures and ultimately a higher risk of relapse (Raffel, Falcone et al. 2017).

Smaller studies have also associated aberrant DNAm in genes such as *MGMT* (Sobieszkoda, Czech et al. 2017), *GPX3* (Zhou, Lin et al. 2016), *C1R* (Bozic, Lin et al. 2015), *ID4* (Zhou, Zhang et al. 2017), *NKD2* (Li, Zhou et al. 2017), *PD-L1* (Goltz, Gevensleben et al. 2016) *CTNNA1* (Li, Gao et al. 2016), *DLX4* (Zhou, Zhang et al. 2016), *SCIN* (Zhang, Zhang et al. 2018), and *SFRPs* (Guo, Zhang et al. 2017), with various effects upon prognosis, disease pathogenesis and subtype classification. These genes are just a few of the many that are regulated by DNAm and serve as an example of the complexity of a disease as heterogeneous as AML.

As methylation patterns for individual genes and complex methylation signatures are associated with AML subtypes, there is the potential for large scale classifier tools to improved accuracy of diagnosis across cytogenetically ambiguous cases. Su-In Lee *et al.* has shown using expression data and *in vitro* drug sensitivity screening that MBL models can be created to determine treatment response to particular levels of expression (Lee, Celik et al. 2018). A novel

method designed for fast clinical turnaround, utilising multiplex evaluation of DNAm combined with a MBL random forest algorithm has shown improved capacity to determine patient subgrouping and ultimately mortality risk (Wertheim, Smith et al. 2015). Furthermore, Luskin *et al.* used the same multiplex methodology and used a ‘methylation statistic’ to predict failure to achieve complete remission (Luskin, Gimotty et al. 2016). These strategies may be applied to DNAm array data for novel approaches to prognostication in both adults and children.

1.3.3.2 Methylation in Paediatric AML

DNAm has also been interrogated in paediatric AML, albeit not to the extent of other cancers. This most likely reflects the rarity of paediatric AML. However, several studies have shown similar utility for DNAm in this field as in adult AML. The flagship study recently completed by Bolouri *et al.* shows that DNAm can be used to define patient subtypes even when both children and adults are combined (Bolouri, Farrar et al. 2017). A hypersilenced group, found to primarily consist of children with *WT1* mutations, along with *CBF*, *KM2TA* and *IDH* mutated groups could be accurately defined. Furthermore, Bolouri *et al.* discovered that many genes in the Wnt- β -catenin signalling pathway, which is necessary for development of leukaemic stem cells, were deleted or aberrantly methylated. Activating ligands for natural killer (NK) cells seemed to be actively repressed in children as well (Bolouri, Farrar et al. 2017).

Smaller studies, such as Tao *et al.* have shown that *EBF3* (Tao, Xu et al. 2015) and *ZFP382* (Tao, Hu et al. 2014) promoter hypermethylation is associated with a decrease in gene expression and that overexpression inhibited cell proliferation and increased apoptosis. Similarly, *GATA4* promoter hypermethylation has been observed in paediatric AML samples, and has been directly correlated to patient white blood cell count and MRD. Those patients showing *GATA4* hypermethylation were also noted to have markedly shorter survival time than those without (Tao, Fang et al. 2015). *MNI* has been implicated in inv(16) paediatric AML through overexpression when *DNMT3B* is unable to methylate the promoter region of the gene (Larmonie, Arentsen-Peters et al. 2017). *MNI* is a known oncogene but its utility in AML pathogenesis is yet to be discovered. Morenos *et al.* was the first example of genome-wide epigenetic analysis in paediatric AML, revealing an increase in methylation at the

DLUE2/Alt1 transcriptional start site leading to repressed gene expression (Morenos 2014). Stratification of cytogenetics also showed that patients with chromosome 11 abnormalities or trisomy 8 were not affected by this aberrant methylation, suggesting that this finding may be karyotype specific.

As established earlier, the more severe effects of AML generally occur at relapse. DNAm is being considered for biomarkers of relapse at multiple time points throughout disease progression. Tan *et al.* has shown that mutations in DNAm regulators, namely *DNMT3A*, cause carcinogenic transformation from haematopoietic stem cells to the pre-leukaemic phase (Tan, Liu et al. 2015). This information can be used both as a therapeutic target as well as a marker for potential relapse. Agrawal *et al.* suggest the use of *ERα/p15* methylation as a biomarker of MRD as many patients that achieved remission but retained methylation in those loci were associated with a poorer 5-year EFS and OS post-relapse (Agrawal, Unterberg et al. 2007). Looking at *RUNX1-RUN1T1* patients only, Zampini *et al.* postulates that a 337 base pair differentially methylated region (DMR) signature can accurately distinguish patients that relapsed. This signature suggests cell-to-cell adhesion and cell-motility are the core pathways involved in the relapse process for this subtype (Zampini, Tregnago et al. 2018)

Many of the epigenetic studies in paediatric AML consist of subtype and loci specific analyses, and therefore lack the scale many adult AML genome-wide studies achieve. Paediatric-specific studies are important as there are few epigenetic comparisons of adult and paediatric leukaemias, despite polymorphisms in epigenetic-modifying genes being virtually absent in the paediatric setting. Furthermore, we need to elucidate the how the differences in chromosomal and genetic heterogeneity affect methylation and ultimately disease in the paediatric setting. Finally, the prognostic use of large scale, DNAm arrays have been relatively unexplored, thus, further studies designed around the paediatric setting will need to be implemented to further elucidate its utility within paediatric AML.

1.4 The Present Study

In recent years, DNAm and other epigenetic modifiers have been gaining attention in cancer pathogenesis, particularly in AML. With newly approved treatments, both using epigenetic and genetic targeted approaches, research into the pathogenesis of AML is slowly bearing fruit. However, it is becoming increasingly clear that both the genetic and epigenetic landscape of paediatric AML varies substantially from its adult counterparts. Subsequently, paediatric AML should be treated as a separate disease. Furthermore, with current risk stratification based only off adult AML genomic data (Mrozek, Marcucci et al. 2012), and recent studies suggesting that co-operative mutation can influence disease risk, an updated risk stratification protocol is necessary to appropriately define paediatric AML prognosis. While these advances in disease pathogenesis continue to inform future treatment, the sheer heterogeneity of paediatric AML leaves much to be discovered, especially in the field of DNAm-defined regulation.

In the present study we interrogated the DNA methylome of paediatric AML to improve our understanding of the biology of disease, to uncover potentially novel pathogenic mechanisms defined by regulation of DNAm, and to identify prognostic approaches to risk stratification. Through this effort we hope to assist in defining patient risk for better long-term outcomes, while defining novel pathogenic mechanisms that may help in future personalised treatments.

1.4.1 Hypotheses and Aims

Aim 1: To interrogate global DNAm levels of paediatric AML diagnostic bone marrow samples relative to matched remission and healthy control samples.

Hypothesis 1a: AML is associated with disruption of global 5mC levels

Hypothesis 1b: Global 5mC levels are associated with individual subtypes of AML.

Aim 2: To identify genes with aberrant DNAm patterns associated with AML pathogenesis, and to distinguish AML diagnostic bone marrow samples from matched, remission samples.

Hypothesis 2a: Genes implicated in AML pathogenesis will be associated with aberrant DNAm.

Hypothesis 2b: Specific patterns of DNAm will distinguish AML bone marrow from healthy, remission bone marrow samples.

Aim 3: To distinguish individual AML subtypes using DNAm signatures, and to identify and replicate gene-specific aberrant DNAm associated with the *CBFB-MYH11* subtype of AML.

Hypothesis 3a: AML subtypes will be defined by DNAm signatures, separating them into distinct groups.

Hypothesis 3b: Genes previously implicated in *CBFB-MYH11* pathogenesis will be associated with aberrant DNAm.

Aim 4: To use DNAm profiling as a tool to predict patient outcome at diagnosis through machine-based learning techniques.

Hypothesis 4: 5-year EFS can be predicted at diagnosis through machine-based learning techniques utilising DNAm array data.

This page is intentionally left blank.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Equipment

Table 2.1 Study equipment

Equipment	Supplier	City
1.5mL Microcentrifuge tubes	Eppendorf AG	Hamburg, Germany
Safelock 2mL Microcentrifuge Tubes	Eppendorf AG	Hamburg, Germany
384-well Microplates	Axygen	Union City, USA
96-well Microplates	Axygen	Union City, USA
Centrifuge 5424 R	Eppendorf AG	Hamburg, Germany
Centrifuge 5804	Eppendorf AG	Hamburg, Germany
Centrifuge 5810 R	Eppendorf AG	Hamburg, Germany
ChemiDoc™ XRS+ System	Bio-Rad	Hercules, USA
Dry Block Heater	Ratek Instruments	Melbourne, Australia
Falcon Conical Tubes	BD Biosciences	New Jersey, USA
Gene Amp PCR System 9700	Applied Biosystems	Melbourne, Australia
Kimwipes	Kimtech Science	Milsons Point, Australia
Manual Multichannel Pipette Pipet-Lite XLS+	Mettler Toledo	Port Melbourne, Australia
MassARRAY Compact Analyzer	SEQUENOM	San Diego, USA
MassARRAY Nanodispenser	SEQUENOM	San Diego, USA
MicroAmp® Clear Adhesive Film	Applied Biosystems	Melbourne, Australia
Micro Centrifuge	Tomy Kogyo	Tokyo, Japan
Midi Gel Tank Kit	Plaztek Scientific	Melbourne, Australia
NanoDrop™ 1000 Spectrophotometer	Thermo Fisher Scientific	Washington, USA
Nunc Cell Scrapers	Thermo Fisher Scientific	Washington, USA
Pipette Aid	Drummond Scientific Company	Broomall, USA
Pipette Tips	Axygen	Union City, USA
Multichannel Pipette Tips	Mettler Toledo	Port Melbourne, Australia
Pipettes	Gilson	Middleton, USA
Qubit® 2.0 Fluorometer	Life Technologies	Singapore, Singapore
SpectroCHIP® II-G384	Agena Bioscience	San Diego, USA
Suspension Mixer	Ratek Instruments	Melbourne, Australia
Thin Wall Strip Cap Tubes	Pacific Laboratory Products	Blackburn, Australia
Veriti 96-well Thermal Cycler	Applied Biosystems	Melbourne, Australia
Vortex	Ratek Instruments	Melbourne, Australia

2.1.2 Reagents

Table 2.2 Study reagents

Reagent	Supplier	City
1 kb Plus DNA Ladder	Invitrogen	Mulgrave, Australia
10 mM dNTP Mix	Life Technologies	Oregon, USA
10x PCR Buffer	Qiagen	Dusseldorf, Germany
2x FastStart™ PCR Master Mix	Sigma-Aldrich	Castle Hill, Australia
5x T7 Polymerase Buffer	Agena Bioscience	San Diego, USA
Absolute Ethanol	Merck Millipore	Kilsyth, Australia
Agarose Powder	Bioline	London, UK
AllPrep DNA/RNA Mini Kit	Qiagen	Dusseldorf, Germany
Chloroform	Sigma-Aldrich	Castle Hill, Australia
CpGenome Universal Methylated DNA	Merck Millipore	Kilsyth, Australia
CpGenome Universal Unmethylated DNA	Merck Millipore	Kilsyth, Australia
Dithiothreitol (DTT)	Agena Bioscience	San Diego, USA
Ethylenediaminetetraacetic Acid (EDTA)	Merck Millipore	Kilsyth, Australia
EZ-96 DNA Methylation-Lightning MagPrep Kit	Zymo Research	Irvine, USA
FlexiGene DNA Kit	Qiagen	Dusseldorf, Germany
GelRed™	Biotium	Hayward, USA
Heat Inactivated Foetal Bovine Serum	Scientifix	South Yarra, Australia
Magnesium Chloride (MgCl ₂)	Sigma-Aldrich	Castle Hill, Australia
Methyl Easy™ Xceed Kit	Human Genetic Signatures	Randwick, Australia
Milli-Q Purified Water	Merk Millipore	Kilsyth, Australia
Nuclease-Free Water	Ambion	Austin, USA
Phenol	Sigma-Aldrich	Castle Hill, Australia
Proteinase K	Qiagen	Dusseldorf, Germany
Qubit® dsDNA BR Assay Kit	Life Technologies	Oregon, USA
Resin	Agena Bioscience	Sand Diego, USA
RNase A	Agena Bioscience	San Diego, USA
RNase-Free ddH ₂ O	Agena Bioscience	San Diego, USA
Shrimp Alkaline Phosphatase (SAP)	Agena Bioscience	San Diego, USA
Sodium Dodecyl Sulphate (SDS)	Roche	Castle Hill, Australia
Sodium Hydroxide Pellets	Sigma-Aldrich	Castle Hill, Australia
T7 RNA and DNA Polymerase	Agena Bioscience	San Diego, USA
T-Cleavage Mix	Agena Bioscience	San Diego, USA
Tris-Hydrochloride	Roche	Castle Hill, Australia
Xylene	Sigma-Aldrich	Castle Hill, Australia

2.1.3 Buffers and Solutions

Table 2.3 Study buffers and solutions

Buffer/Solution	Contents
1x TE Buffer	10mM tris-hydrochloride pH 8.0, 1mM EDTA, stored at room temperature
1x TBE Buffer	90mM tris-hydrochloride pH 8.0, 90mM boric acid, 2mM EDTA, stored at room temperature
6x Loading Dye	40% (v/v) glycerol, 0.05% (v/v) bromophenol blue, 0.05% (v/v) xylene cyanol, stored at room temperature
Extraction Buffer	100mM sodium chloride, 10mM tris-hydrochloride, 25mM EDTA pH 8.0, 0.5% SDS, milli-Q purified water, stored at room temperature
Phenol Chloroform Solution	Chloroform, phenol, and isoamyl-alcohol prepared in a 25:24:1 ratio, stored at 4°C in aluminium foil covered 50mL Falcon tubes as solution is light sensitive

2.2 Study Cohorts & Sample Collection

2.2.1 Australian Paediatric Acute Myeloid Leukaemia Cohort

The Australian paediatric AML sample was generated by combining archival (slide smears) and fresh/frozen bone marrow aspirates from the Royal Children's Hospitals (RCH) in Melbourne and Brisbane, The Children's Hospital of Westmead, Sydney and The Princess Margaret Hospital for Children in Perth.

All samples were obtained following appropriate ethics approvals. Samples from the RCH Melbourne were obtained through several means. Stored archival (room temperature) and frozen bone marrow were obtained from the Children's Cancer Centre. Fresh bone marrow was obtained from the Children's Cancer Centre Tissue Bank (CCCTB). Samples from the Children's Hospital at Westmead Tumour Bank primarily consisted of archival bone marrow aspirates. Fresh/frozen bone marrow samples obtained from RCH Brisbane tissue bank were provided by Dr Andrew Moore. Fresh bone marrow from the Princess Margaret Hospital for Children were provided by Dr Catherine Cole and were exclusively used for replication purposes.

Tissue used in this study were sourced to form a retrospective cohort of samples consisting of diagnostic, remission, and relapse bone marrow aspirates. At the time of collection, sample age ranged from <1 to >15 years of age. A summary of patient demographic data is detailed in Section 2.3

Collectively, this cohort included 202 bone marrow samples across multiple time points derived from 114 patients. Samples were selected according to the following criteria:

- A diagnostic sample where sufficient patient cytogenetic, follow-up, outcome, and standard epidemiological data was available;
- Follow-up samples for each patient were available depending on their outcome status. Patients that did not relapse were to have an available remission sample, while patients that did were to have a relapse sample.

- Diagnostic samples from as many different cytogenetic abnormalities as possible to ensure that findings were not driven solely by a dominant AML subtype.

2.2.2 Therapeutically Applicable Research to Generate Effective Treatments (TARGET) Study

The Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative used high resolution genomic approaches to comprehensively characterise and determine the changes that drive progression and initiation in high-risk childhood cancers. Patients admitted to drug trials being overseen by the Children's Oncology Group (COG) were recruited for the TARGET study. Characterisation of patient samples was undertaken by whole genome sequencing, whole exome sequencing, RNA-seq, mRNA-seq, miRNA-seq, bisulphite sequencing, HM450k array and ChIP sequencing across an array of different cancer types.

We obtained publicly available genome-wide DNAm array files (IDAT files) and clinical data from acute myeloid leukaemia samples used in the TARGET study. These samples will be referred to as the TARGET study/cohort and were used exclusively for pooled-analyses with the Australian discovery cohort. All sample collection, handling and information collection was carried out by the respective research institutes and hospitals throughout North America. AML data was originally published by Bolouri *et al* (Bolouri, Farrar et al. 2017).

2.2.3 Healthy Bone Marrow Data

We obtained additional publicly available genome-wide DNAm array files (IDAT files) from adult, healthy bone marrow donors (GSE63409). This dataset comprised of a total of 24 cell sorted bone marrow samples derived from five patients. Samples included five normal haematopoietic stem cell samples (HSC), four multipotent progenitor cell samples (MPP), five common myeloid progenitor cell samples (CMP), five megakaryocyte-erythroid progenitor samples (MEP) and five granulocyte-macrophage progenitor cell samples (GMP). This data was publicly available from NCBI Gene Expression Omnibus (GEO) and was originally generated by Jung *et al* (Jung, Dai et al. 2015).

2.2.4 Combined Analyses & Data Harmonisation

Data from the three aforementioned cohorts were pooled and re-analysed in order to strengthen statistical power and validate discovery cohort findings. Data between the Australian discovery cohort and the North American TARGET cohort were harmonised where possible. Harmonisation of meta-data was achieved including technical aspects such as what chip and position each sample were run on, as well as biological aspects such as sex, age, sample time points, 5-year patient outcome, overall survival and cytogenetic data. This harmonised meta-data allowed for all analyses in this study to be conducted with all cohorts included while also allowing for technical variation to be discovered and removed accordingly as further discussed in Section 2.6.

2.3 Cohort Demographics

Table 2.4 Clinical characteristics and demographic data of all samples used in this study

	Australian Cohort (n = 112)				TARGET Cohort	
Clinical Characteristics	RCH Melbourne (n = 44 patients)	RCH Brisbane (n = 21 patients)	Children's Hospital of Westmead (n = 38 patients)	Princess Margaret Children's Hospital (n = 9 patients)	North American TARGET AML Cohort (n = 89 patients)	Total (n = 201 patients)
Sample time point						
Diagnostic	41 (65.1%)	21 (48.8%)	38 (53.2%)	9 (50.0%)	89 (38.5%)	198 (47.5%)
Remission	12 (19.0%)	13 (30.3%)	15 (24.2%)	0	103 (44.6%)	143 (34.3%)
Relapse <i>[Secondary]</i>	10 (15.9%)	9 (20.9%)	9/3/ (14.6%)	9 (50.0%)	39 (16.9%)	76 (18.3%)
Sample type						
Archival	41 (65.1%)	0	54 (87.1%)	0	N/A	95 (51.1%)
Frozen	14 (22.2%)	0	0	0	N/A	14 (7.5%)
Fresh	8 (12.7%)	43 (100%)	8 (12.9%)	18 (100%)	N/A	77 (41.4%)
Sample use						
450K	62 (98.4%)	7 (16.3%)	59 (95.2%)	0	232 (100%)	360 (84.9%)
Replication	1 (1.6%)	36 (83.7%)	3 (4.8%)	18 (100%)	0	64 (15.1%)
Age at diagnosis						
Infants (<3 y/o)	14 (31.8%)	8 (38.1%)	13 (34.2%)	4 (44.5%)	25 (28.1%)	64 (31.9%)
Children (3-15 y/o)	22 (50.0%)	11 (52.4%)	20 (52.6%)	5 (55.5%)	46 (51.7%)	104 (51.7%)
AYA (>15 y/o)	8 (18.2%)	2 (9.5%)	5 (13.2%)	0	18 (20.2%)	33 (16.4%)
Sex						
Male	23 (52.3%)	12 (57.1%)	23 (60.5%)	2 (22.2%)	45 (50.6%)	105 (52.2%)
Female	21 (47.8%)	9 (42.9%)	15 (39.5%)	7 (87.8%)	44 (49.4%)	96 (47.8%)
Status						
Alive	26 (59.1%)	14 (66.7%)	24 (63.1%)	6 (66.7%)	40 (44.9%)	109 (54.2%)
Deceased	18 (40.9%)	7 (33.3%)	12 (31.6%)	3 (33.3%)	49 (55.1%)	90 (44.8%)
Unknown [†]	0	0	2 (5.3%)	0	0	2 (1.0%)

Table continued on next page

	Australian Cohort				TARGET Cohort	
Clinical Characteristics	RCH Melbourne (n = 44 patients)	RCH Brisbane (n = 21 patients)	Children's Hospital of Westmead (n = 38 patients)	Princess Margaret Children's Hospital (n = 9 patients)	North American TARGET AML Cohort (n = 89 patients)	Total (n = 201 patients)
Outcome[‡]						
Good	13 (29.5%)	4 (19.0%)	21 (55.3%)	0	10 (11.2%)	48 (23.9%)
Poor	23 (52.3%)	16 (76.2%)	17 (44.7%)	9 (100%)	79 (88.8%)	144 (71.6%)
Unknown	8 (18.2%)	1 (4.8%)	0	0	0	9 (4.5%)
Cytogenetics						
inv(16)/t(16;16)	6 (13.6%)	0	4 (10.5%)	0	6 (6.7%)	16 (8.0%)
t(8;21)	1 (2.3%)	1 (4.8%)	1 (2.6%)	1 (11.1%)	7 (7.9%)	11 (5.5%)
<i>KMT2A</i> rearranged	11 (25%)	6 (28.6%)	7 (18.4%)	3 (33.3%)	14 (15.7%)	41 (20.4%)
Trisomy 8	1 (2.3%)	2 (9.4%)	5 (13.2%)	1 (11.1%)	0	9 (4.5%)
Trisomy 21	0	0	1 (2.6%)	0	0	1 (<1%)
CN	11 (25%)	3 (14.3%)	8 (21.1%)	0	25 (28.1%)	47 (23.4%)
Complex karyotype [*]	7 (15.9%)	3 (14.3%)	4 (10.5%)	0	0	14 (7.0%)
Other [‡]	7 (15.9%)	6 (28.6%)	8 (21.1%)	1 (11.1%)	26 (29.2%)	48 (23.9%)
Unknown	0	0	0	3 (33.3%)	11 (12.4%)	14 (7.0%)

Abbreviations: RCH - Royal Children's Hospital; TARGET AML - Therapeutically Applicable Research To Generate Effective Treatments in Acute Myeloid Leukaemia; y/o - Years old; AYA - Adolescent/Young adult; inv - Inversion; t - Translocation; CN - Cytogenetically normal.

^{*}Complex karyotypes are defined by patients having two or more cytogenetic abnormalities or abnormalities that are otherwise unrelated to AML pathogenesis. [‡]Other cytogenetic abnormalities are defined as del5/7/9/13/20q, inv(5q), -7, t(5;6), t(6;9), t(7;12), t(7;14), t(9;22), t(15;17), t(16;21). Unknown genetic abnormalities are undefined in patient records.

[†]Patients achieved a 5-year EFS however recent data is unavailable to ensure that they are currently still alive.

[‡]Patient outcome is defined as good if the patient achieved a 5-year EFS. Poor outcome is defined as a patient that relapsed or died prior to achieving 5-year EFS

2.4 Isolation of Nucleic Acids

2.4.1 Patient Sample Pre-Processing

Fresh bone marrow aspirates taken from the CCCTB were processed within two hours of being taken from the patient. The bone marrow aspirate was first diluted in an equal volume of RPMI in a 15mL falcon tube. The diluted bone marrow was layered onto 5mL of Ficoll-Paque and centrifuged for 30 minutes at 400 x g. The bone marrow mononuclear cell (BMMC) layer was then removed, placed into a new 15mL falcon tube and was topped up to 14mL with RPMI before being centrifuged at 500 x g for three minutes. Residual RPMI was removed and the pellet was resuspended in 1mL of RPMI before cell counting on a haemocytometer. Three aliquots of 1.0×10^6 cells were stored in 1mL of HI FBS with 10% DMSO cryopreserved by being stored in a -80°C freezer in a controlled freezing container. Remaining cells were aliquoted at a 5-10 x 10^6 /mL density and stored in RNA later or as a dry pellet before being moved to a -80°C freezer. Frozen bone marrow donated from the VCGS was initially frozen at -30°C in a sodium heparin tube before DNA extraction.

Archived patient slides were available to the MCRI from the RCH Melbourne Haematology department. Samples were archived following routine diagnostic testing and stored in two different manners. Unstained bone marrow aspirates comprised whole bone marrow, smeared and air dried on a glass slide. Stained bone marrow aspirates were whole bone marrow slides dyed with Giemsa, mounted with DPX media under a cover slip. Prior to extracting DNA from stained patient sample slides, they were soaked in 100% Xylene for a minimum of 2 days to safely remove the mounting media and coverslip without damaging the sample. The bone marrow aspirate was directly dissolved on the glass slide with tissue lysis buffer, and scraped using a cell scraper (Nunc International, NY) into a 1.5mL microcentrifuge tube for further downstream application.

2.4.2 Extraction of DNA from Archival Bone Marrow

DNA extraction was performed using a standard phenol/chloroform method. This method was adapted to be used with archival bone marrow samples as previously described (Morenos, Saffery et al. 2012). The phenol/chloroform method involved adding 2 x 250µL of lysis buffer

to the bone marrow aspirate slide and scraping it off into a 1.5µL microcentrifuge tube. We added 20µL of proteinase K (10mg/mL) to the sample and heated to 55°C for a minimum of two hours, but most were completed overnight. Once removed from the heat block 500µL of phenol chloroform is added and shaken vigorously. Microcentrifuge tubes were then centrifuged at 12,000 x g for 10 minutes. The aqueous, upper layer of the solution is then removed and placed into a new tube, with the remainder being discarded. We then add 350µL of isopropanol and mixed by vortex to precipitate the DNA. The microcentrifuge tube was again centrifuged at 12,000 x g for 5 minutes. The supernatant was removed with care not to disturb the DNA pellet. The DNA pellet was then left to air dry for 5 minutes before being resuspended in 50µL of 1 x TE buffer. All DNA was tested for quality and quantity before any further downstream applications and is discussed in section 2.2.3.

2.4.3 Extraction of DNA and RNA from Fresh/Frozen Bone Marrow

Extraction of DNA and RNA from fresh bone marrow was achieved through the use of the AllPrep DNA/RNA Mini Kit. Thawed, viably stored samples were centrifuged at 300 x g for 5 minutes to pellet the cells. Supernatant was removed from the tube and 350µL of RLT buffer was added to resuspend and lyse the cells. Once resuspended, the lysate was homogenised by passing it through a 20-gauge needle multiple times and then the remaining lysate was transferred to an AllPrep DNA spin column, placed in a 2mL collection tube and centrifuged for 30 seconds at 8000 x g. This step was repeated to ensure all DNA was captured in the matrix. The DNA column was then transferred to a new collection tube and stored at 4°C while the RNA was extracted. 350µL of 70% ethanol was added to the collection tube with the RNA lysate and mixed by pipetting. The entire contents of the collection tube were then transferred to an RNeasy spin column placed inside a 2mL collection tube and centrifuged for 15 seconds at 8000 x g. Flowthrough was rerun through the column to ensure optimal RNA binding before being discarded. Following this, 700µL of RW1 buffer was added to the column and centrifuged for 15 seconds at 8000 x g. This was repeated once for 2 minutes instead of 15 seconds before being discarded. Placing the column in a new collection tube, the column was then spun for an additional minute at full speed to dry the membrane and ensure no carry over of the buffer. The column was once again moved to a new collection tube, had 50µL of RNase

free water added to it and centrifuged for 1 minute at 8000 x g. The DNA spin column was taken out of 4°C and 500µL of AW1 buffer was added to it. This column was then centrifuged for 15 seconds at 8000 x g and repeated once to ensure a thorough washing. 500µL of AW2 buffer was added to the column and centrifuged at full speed for 2 minutes. Finally, the DNA column was transferred to a 1.5mL collection tube and had 100µL of RNase free water added to elute the DNA. This column was spun at 8000 x g for 1 minute to elute DNA. Flowthrough was rerun through the column again to ensure all DNA was eluted. DNA and RNA were measured using the NanoDrop ND-1000 Spectrophotometer which is described in detail in section 2.4.4.

The FlexiGene DNA Kit was used to extract DNA from frozen bone marrow aspirates. Frozen bone marrow was thawed at 37°C in a water bath. To process 300µL of bone marrow, 750µL of FG1 buffer was added and mixed by inversion in a 1.5mL microcentrifuge tube. This sample was then spun for 20 seconds at 10000 x g. Supernatant was discarded and the tube and pellet were left to air dry for 2 minutes. 150µL of FG2 buffer mixed with 1.5µL of QIAGEN protease was added to the tube and vortexed to homogenise the pellet. The sample was briefly centrifuged and placed in a heating block for 5 minutes at 65°C. Afterwards, 150µL of isopropanol was mixed by inversion into the solution to precipitate the DNA. The sample was then centrifuged for 3 minutes at 10000 x g, and the supernatant was discarded and left to dry on an absorbent paper towel. The sample was then washed with 150µL of 70% ethanol and vortexed for 5 seconds. Following this, the sample was centrifuged for 3 minutes at 10000 x g, had the supernatant removed and left to dry again on a paper towel for 5 minutes. Finally, the remaining DNA was dissolved in 200µL of RNase free water and incubated in a heating block for 1 hour at 65°C before concentration being measured on the NanoDrop ND-1000 Spectrophotometer.

2.4.4 Measurement of Yield and Quality

Successful extraction of DNA from both fresh/frozen and archival bone marrow was assessed by two different methods; spectrophotometry and fluorometry. The NanoDrop ND-1000 Spectrophotometer uses the full spectrum (220-750nm) to quantify and evaluate the purity of samples, including DNA, RNA and proteins. The quality of a sample is determined by the

both the 260nm/280nm and 260nm/230nm ratios. Knowing that peak absorbance wavelength of DNA is at 260nm, phenol at 230nm and protein at 280nm allows us to determine if other contaminants are found in the sample. An optimal ratio of 1.8 and 2.0 respectively suggests a contaminant free sample for DNA. Values lower than 1.8 (260nm/280nm ratio) suggest there is protein contamination while values lower than 2.0 (230nm/260nm ratio) indicates phenol contamination. Following calibration, using 1 μ L of elution buffer (1 x TE buffer), 1 μ L of sample was placed on the device and measured. Concentration, 230nm/260nm ratio and 260nm/280nm ratios were recorded.

The DNA yield was re-measured using the Qubit 2.0 Fluorometer as a more accurate measure of DNA yield, however qualitative data is unavailable through this method. This device allows for a much more accurate determination of yield as it calculates concentrations based off a relative scale. Using the Qubit dsDNA BR Assay Kit, we created a 1:200 working solution of Qubit reagent to buffer. Two standards were created to ensure relative yield was possible to calculate. We used 190 μ L of working solution with 10 μ L of DNA standard. For actual samples, we used 199 μ L and 1 μ L of sample. Tubes were briefly vortexed and left to incubate for 2 minutes at room temperature. A standardised linear range of fluorescent intensity was generated by the two standards (measuring in at 0 ng/ μ L for low range and 1000ng/ μ L at high range) in the Qubit 2.0 Fluorometer. Tubes containing samples to be measured were then placed into the fluorometer and measured relative to the standards. Measurement of DNA was multiplied by 10 to account for the dilution factor between standards and sample and recorded.

2.5 Measurement of DNA Methylation

Many methods are available to analyse DNAm, each with their own advantages and disadvantages, primarily focused upon coverage of the genome and price point. In this study, we have used Illumina Infinium HumanMethylation450 BeadChip Array (HM450K) as our platform of choice for genome wide methylation data generation. We have also verified findings and assayed sites of methylation not covered by the HM450K using the SEQUENOM MassArray EpiTYPER (Ehrich, Nelson et al. 2005). Each of these platforms are described in more detail below.

2.5.1 Bisulphite Conversion

Bisulphite conversion is pivotal to distinguish methylated from unmethylated CpG sites. The process of bisulphite conversion is centred around the concept that unmethylated cytosine bases can be converted to uracil, a process known as deamination, whereas 5-methylcytosines (5mC) are protected from this conversion (Figure 2.1). After this conversion occurs, the sample is PCR amplified and converted uracil is replaced with thymine. Essentially, this process induces a C>T single nucleotide polymorphism (SNP) that is dependent upon methylation status. This SNP can then be detected via mass spectrometry (as used with SEQUENOM) or by sequencing. Genomic DNA used in this study was bisulphite treated at ServiceXS (Netherlands) as part of the service provided with their Illumina platform.



Figure 2.1 Effect of bisulphite conversion upon methylated and unmethylated DNA

Unmethylated cytosine bases (blue) undergo transformation to uracil and subsequently to thymine upon PCR amplification. 5-methylcytosine (5-mC, green) is protected from this reaction and is therefore unchanged.

In-house validation of DNAm was necessary to ensure true-positive findings and therefore samples were also bisulphite converted using the Methyl Easy Xceed Bisulphite Conversion Kit as per manufacturer's instructions. 20ng of genomic DNA was placed in a 1.5mL microcentrifuge tube, diluted with RNase free water to a total of 20µL. Denaturation of DNA was undertaken through the addition of 2.2µL of 3M NaOH solution which was later incubated for 15 minutes at 37°C. Afterwards, 220µL of combined Reagent 1 & 2 were added and mixed by pipetting. The solution is light sensitive, therefore upon addition the tube was covered with foil and incubated at 80°C for 45 minutes in a heating block. Tubes were centrifuged to ensure condensation was removed and subsequently had 240µL of Reagent 3 added to it. The contents were then transferred to 1.5mL purification columns within a collection tube. Columns were centrifuged at maximum speed for 1 minute and flowthrough was re-run through the column to ensure optimal binding. Flowthrough was then discarded and 300µL of Reagent 4 was added and again centrifuged at maximum speed for 1 minute. This step was repeated once to thoroughly wash the sample. Flowthrough was again discarded, and the remaining spin column was centrifuged for 4 minutes at maximum speed to dry the sample. Columns were placed in 1.5mL microcentrifuge tube and were eluted using 20µL of pre-warmed Reagent 5. Columns were incubated briefly at room temperature before being centrifuged at maximum speed for 1 minute. Microcentrifuge tubes were then incubated for 20 minutes at 95°C to ensure converted DNA was appropriately dissolved.

2.5.2 Illumina Infinium HumanMethylation450 BeadChip Array (HM450k)

The Illumina Infinium HumanMethylation450 BeadChip Array (HM450k) is a microarray-based platform designed to interrogate genome-wide DNAm. This platform analyses the methylation status of 485,764 CpG dinucleotides (probes), covering 99% of all RefSeq genes with an average of 17 probes per gene. These probes are distributed primarily throughout promoter regions of genes but can be found anywhere throughout the genome. A breakdown of probes across functional genomic regions, chromosomes and neighbourhood context can be found in Figure 2.2.

After sample processing was completed, an optimal amount of 1µg of DNA (but as low as 300ng) was sent to ServiceXS (Netherlands) for quality control, bisulphite conversion and array processing. Quality control consisted of a picogreen assay to ensure minimum requirements were met, while fragmentation analysis was undertaken to ensure quality standards. DNA was then bisulphite converted, amplified, and hybridised to the HM450K bead chip according to the manufacturer's instructions. This array has two methods in which it measures DNAm. Infinium probe type I has two silica beads representing methylated and unmethylated. Each of these beads has a corresponding sequence attached which allows it to bind to its respective DNA strands. Single base extension allows for the addition of a labelled nucleotide that defines methylation status based on bisulphite conversion status. The second method uses the Infinium probe type II. This probe uses only one type of bead that binds its sequence to the DNA and single base extension (dependent upon bisulphite conversion) attached either a G to methylated DNA (green) or an A to unmethylated DNA (red). The array is then stained and the intensity of these signals is captured and stored as .IDAT files. Files are then returned and the methylation value of each sample and probe are determined through bioinformatic approaches which are detailed further in section 2.5.6.

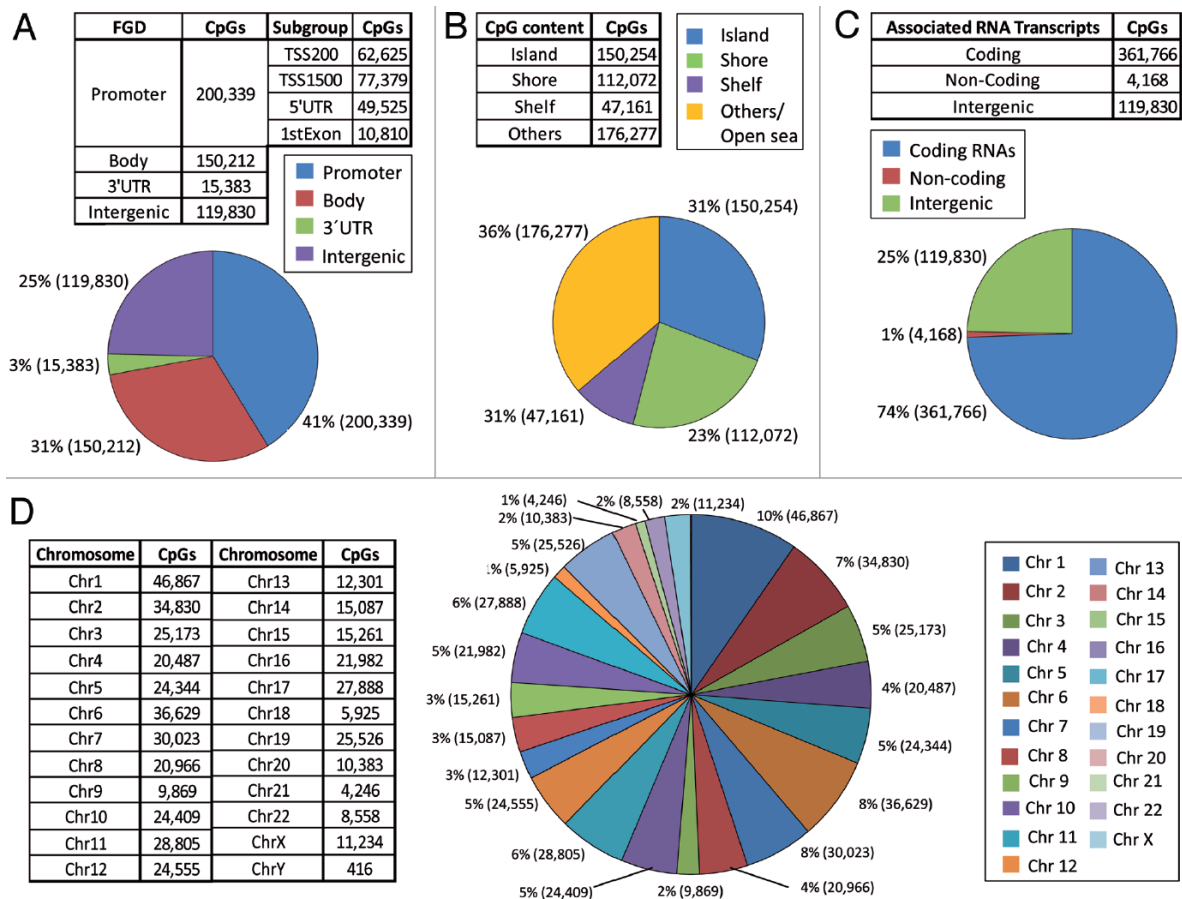


Figure 2.2 Description of the probes on the HM450K DNA methylation array

The probes used in HM450K DNAm array are distributed to ensure that the most useful probes are interrogated. (A) Functional genomic distribution (FGD) is classified in different groups depending on location relative to closest gene. (B) CpG content and neighbourhood is classified based on location relative to closest CpG island. (C) Association of RNA transcription in coding, non-coding and intergenic regions. (D) Chromosome location. Figure adapted from (Sandoval, Heyn et al. 2011)

2.5.3 Primer Design for Locus Specific Methylation Analysis

Prior to validation of findings using SEQUENOM, appropriate assay designs were required. Genomic regions of interest are detailed in their appropriate sections and full primer sequences and optimisations for each can be found in Appendix I. Primers were designed using multiple tools. Using the University of California Santa Cruz (UCSC) genome browser (GRCh37/hg19) in combination with the HM450K gene annotation details, sequence data of the area of interest was ascertained. Up to 500 base pairs either side of the desired CpG site were entered into the EpiDesigner tool (<https://www.epidesigner.com/index.html>). EpiDesigner provides automated, bisulphite-converted sequence primers that capture the desired area and gives basic

primer information such as melting temperature. NetPrimer (<http://www.premierbiosoft.com/netprimer/>) was then used to gather more detailed information about these primers, such as dimerization, hairpins and self-dimerization. Forward and reverse primers were to be designed to be within 1°C of one another with minimal CpG sites within them. Converted cytosine to thymine bases were highly sought after to ensure that only bisulphite-converted DNA would amplify during PCR. Finally, the ‘MassArray’ package was utilised in R for *in silico* validation of sites within the amplicon could be measured for their methylation level by SEQUENOM. Some sites may be unmeasurable due to the T-cleave chemistry (detailed in section 2.5.4). The resulting fragmentation of DNA may cause multiple CpG sites methylation to be obscured due to them having the same fragment weight or weights that are outside the detectable range.

Tags were added to each forward and reverse primer to enable *in-vitro* transcription. The T7 tag sequence (5’ CAGTAATACGACTCACTATAGGGAGAAGGCT 3’) was added to the reverse primer while a balance tag (5’ AGGAAGAGAG 3’) was added to the forward primer to match melting temperatures. Primers were ordered from Bioneer (South Korea) in dried form (25nmol, Bio-RP purified) and made up to a stock concentration of 100µM with RNase free water. Working solutions were created at a 1:10 dilution and stored at -20°C until further use.

Primers were tested for optimal annealing conditions using a ‘touch-down’ gradient PCR (detailed in section 2.5.4) at 2°C increments. PCR products were observed by electrophoresis on 2% agarose gel (TBE buffer, 1x gel red, 1kb ladder at 120V for 60 minutes). Negative controls (water) and positive controls (pre-converted bisulphite DNA from the MethylEasy Xceed Kit) were included to ensure both primer and DNA were amplified correctly without contamination.

2.5.4 MassArray EpiTYPER (SEQUENOM) Analysis

The SEQUENOM MassArray EpiTYPER platform was used to validate findings from the genome-wide DNAm results of the HM450K. SEQUENOM is used to measure small scale, region-specific DNAm. Similar to that of the HM450K, SEQUENOM requires DNA to be

bisulphite converted prior to measurement in order to distinguish between methylated and unmethylated nucleotides. By exploiting the 16 Dalton change in mass, induced by the aforementioned bisulphite induced SNP, SEQUENOM is able to distinguish the methylation status of any given CpG site within the targeted region through the use of base-specific cleavage and matrix assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS). The workflow for this protocol begins with the bisulphite conversion of DNA detailed in section 2.5.1. The target region is then amplified using primers designed to specifically bind bisulphite converted DNA. DNA products are then cleaned using a Shrimp Alkaline Phosphatase (SAP) treatment before being cleaved while also being treated with a T7 polymerase to facilitate *in vitro* transcription of RNA. Resulting products should be fragmented at all uracil nucleotides because of the T-cleavage cocktail before being fired in the MALDI-TOF MS. Methylation status is then calculated based on the unique fragment mass signal and is measured by the EpiTYPER software (Figure 2.3).

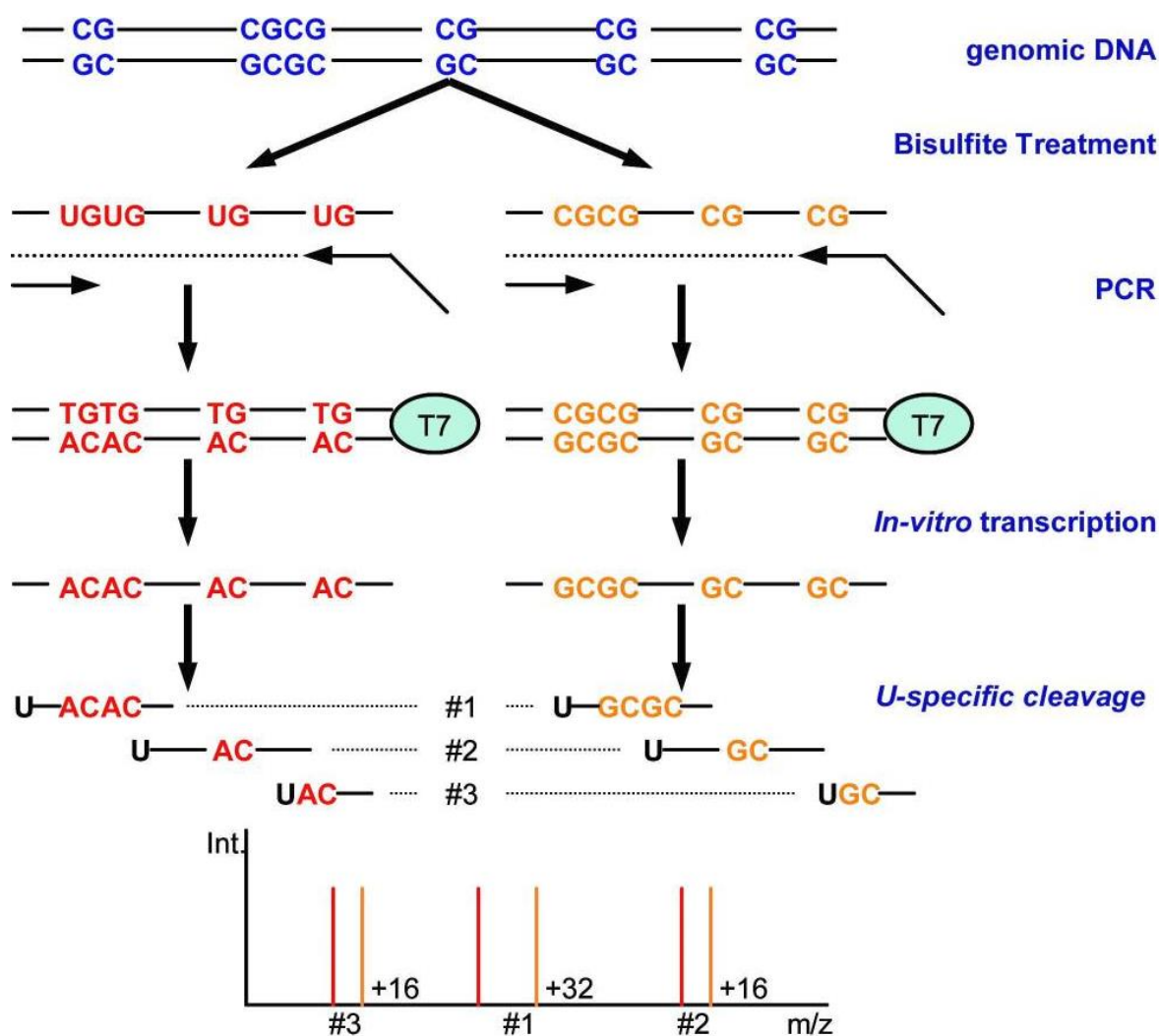


Figure 2.3 Analysis of methylation by base-specific cleavage and MALDI-TOF MS

Genomic DNA is extracted from the source and bisulphite treated. PCR amplification is undertaken with the T7 reverse primer tag to ensure *in vitro* transcription occurs. U-specific cleavage creates fragments of DNA for further analysis in the MALDI-TOF MS platform. DNAm status is defined by a 16 Dalton shift in molecular weight between unmethylated and methylated fragments. Multiple CpG sites can be within a single fragment, therefore this 16 Dalton change in weight is multiplied by each site (as found in fragment #1 of this example). Figure adapted from (Ehrich, Nelson et al. 2005).

Sample processing prior to being run on SEQUENOM firstly required a PCR step utilising the master mix (Table 2.5) and PCR conditions (Table 2.6). Samples were run in triplicate to minimise sample variability of the resulting measurements (Coolen, Statham et al. 2007). PCR products were visualised using the same gel conditions outline in section 2.5.1. Upon successful amplification of DNA, 5µL of the PCR reaction was transferred to a 384-well plate and incorporated with 2µL of the SAP master mix consisting of 159µL of SAP and 901µL of water.

The plate was then sealed and centrifuged for 1 minute at 3000 x g. The samples were then incubated in a thermal cycler under the following conditions: 37°C for 20 minutes, 85°C for 5 minutes and then held at 4°C until ready for the next steps (Ehrich, Nelson et al. 2005).

Table 2.5 Composition of full master mix for bisulphite converted DNA amplification PCR

Reagent	Volume for 1 Sample (μL)
2 x FastStart PCR Master Mix	7.5
Forward Primer (10μM)	0.6
Reverse Primer (10μM)	0.6
Nuclease-free Water	4.3
Bisulphite Converted DNA (1ng/μL)	2.0
Total Reaction Volume	15.0

Table 2.6 Cycling conditions for bisulphite converted DNA amplification PCR

Step	Cycle	Temperature	Time
1	1	95°C	10 minutes
2	5	95°C	10 seconds
		56 - 58°C*	30 seconds
		72°C	2 minutes
3	40	95°C	10 seconds
		56 - 62°C	30 seconds
		72°C	1.5 minutes
4	1	72°C	7 minutes
5	1	4°C	Indefinitely

Sample fragmentation was undertaken in a new 384-well plate utilising 5μL T-cleavage RNase cocktail (Table 2.7) with 2μL of sample carried over from the previous SAP reaction. This was mixed by gentle pipetting, sealed and centrifuged at 540 x g for 1 minute. This was incubated at 37°C for 3 hours and then centrifuged for the same time again. Finally, 20μL of RNase free water was added to each well while 6mg of clean resin was added to a dimple plate and left to dry for 20 minutes. The sample plate was turned upside down, placed over the dimple plate and both were flipped so the resin would fall into the sample wells and mix. The sample plate was rotated for a minimum of 10 minutes before being centrifuged for 5 minutes at 3200 x g. The contents of the sample plate were transferred to a SpectroCHIP using a MassArray nanodispenser. This chip was then placed in the MALDI-TOF MS and run with the assistance of Dr Benjamin Ong. The EpiTYPER program was employed to ensure correct samples, primers, target sequence and positions on the SpectroCHIP were used.

Table 2.7 Composition of T-cleavage RNase cocktail

Reagent	Volume for 1 Sample (μ L)	Volume for 384 Sample (μ L)
RNase-free dH ₂ O	3.21	1602
5 x T7 Polymerase Buffer	0.89	444
T-cleavage mix	0.22	110
DTT (100 mM)	0.22	110
T7 RNA & DNA Polymerase	0.40	200
RNase A	0.06	30
Total Volume	5.0	2496

2.6 Data Cleaning and Analysis

2.5.5 Overview

High-dimensionality data analysis, including array normalisation, machine-based learning and differential analyses, were all completed using the **R** Project for Statistical Computing. **R** is a freely available software environment designed around a statistical programming language. Initial machine-based learning algorithms were tested using Waikato Environment for Knowledge Analysis (Weka). Weka is a collection of machine learning algorithms for data mining, pre-processing, classification, regression, clustering, and visualisation. Finally, the ggplot2 package in **R** was used for basic data visualisation and statistical calculation.

2.5.6 Illumina Infinium HumanMethylation450 BeadChip Array (HM450k) Analysis

HM450K analysis was undertaken on 135 samples in the Australian AML cohort (including previously run, pooled samples), 232 samples from the TARGET AML cohort (Bolouri, Farrar et al. 2017) and a further 24 cell sorted, healthy bone marrow samples from the Jung *et al.* study (Jung, Dai et al. 2015). Pre-processing of data was undertaken by reading in the IDAT files, containing probe signal intensities for each sample, into **R**. Quality control was assessed using the *minfi* package, and any sample that did not meet the mean detection p-value of ≥ 0.01 were removed from further analysis. One sample from each the Australian and TARGET AML cohorts were removed as a result. Signal intensities for remaining samples were converted to methylated (*M*) and unmethylated (*U*) signals.

2.5.6.1 Data Normalisation

The final steps in pre-processing are targeted at potential sources of technical variation found within the data. As previously described, there are two types of Infinium probes, both with different manners in which they work. As a result, technical variation of methylation values can vary between probe types and subsequent regional variation as Type I probes are found more commonly in gene promoter regions (Maksimovic, Gordon et al. 2012). To minimise technical variation, several normalisation methods (including subset-quantile within array, functional,

quantile and variation of the Illumina normalisations) were tested to ensure only technical, and not biological, variation were eliminated from the dataset. After all normalisation methods were compared, the background corrected Illumina normalisation was selected as the optimal method for this dataset using the *minfi* package (Aryee, Jaffe et al. 2014). Probes that were removed from the dataset include those on the X and Y chromosomes to avoid a sex bias, probes that were associated with SNPs (minor allele frequency >1%), and probes that failed across >1 samples to ensure high quality data. Finally, the methylated (*M*) and unmethylated (*U*) signals were used to calculate *MVal* and β values through the following equations:

$$MVal = \log(M/U)$$

$$B = M/(M+U*100)$$

2.5.6.2 Differential Methylation Analysis

Linear regression modelling was implemented to define the average methylation differences between groups of interest. Models were generated using the Remove Unwanted Variance (RUV) package in R (Maksimovic, Gagnon-Bartsch et al. 2015). Prior to modelling, principal components analysis (PCA) was used to identify potential confounding sources of variation including HM450K chip used, position on chip, patient age, subgroup, cytogenetics, sample source, time of extraction, sample type, sex, and patient outcome. While RUV can take covariates into account, all analyses were undertaken without the use of covariates as suggested by Maksimovic *et al.* (Maksimovic, Gagnon-Bartsch et al. 2015). RUV utilises the negative probes on the HM450K array to correlate sources of variation with them, believed to be technical in nature, therefore allowing for the true variation of the group of interest to be left. This process of linear modelling using empirical negative probes is reiterated twice, but instead uses the bottom ranked 90% of all associated probes from the previous analysis as the new empirical negative probes. Probes that remain differentially methylated between groups are therefore believed to be due to variance of the factor of interest. The RUV output provides a $\Delta\beta$ (delta beta, average difference of methylation between two groups), standard error, t-statistic, associated p-value, and adjusted p-value (Benjamini-Hochberg method) for each individual probe. Probes with an adjusted p-value of ≤ 0.05 were considered to be statistically significantly differentially methylated. $\Delta\beta$ values for each probe were always calculated as:

$\Delta\beta$ = mean β -value of factor of interest (e.g. diagnostic) – mean β -value of alternative (e.g. remission)

2.5.7 SEQUENOM MassArray EpiTYPER (SEQUENOM) Data Cleaning and Analysis

Validation and replication data generated through the MALDI-TOF MS SEQUENOM platform was exported as a .csv file from the EpiTYPER program. As all samples were run in triplicate, quality control measures were implemented to ensure samples were as true to their methylation value as possible. To obtain the methylation percentage of any given CpG site, the median value of the triplicate was taken. If all three samples were within ± 0.10 of the median, then the mean was taken and used as the methylation value of that sample. If one of the triplicates did not fall within this range, it was removed and the remaining two values were averaged. If two or more triplicates failed, the entire sample for that CpG site was removed from further analysis. After quality control, each sample was appropriately grouped (e.g. diagnostic vs. remission) and the average methylation for each CpG site available was calculated. Statistical analyses (Student's T-test) between groups were performed for replication purposes while analyses for CpG sites that are shared with HM450K probes were performed for validation purposes.

2.5.8 Pathway Analysis

Pathway analysis was undertaken using the *missmethy1* R package gene ontology (gometh) function, which was carried out on each of the top probe lists used throughout this study. Top probes lists were built by the most significant probes in any of the differential methylation analyses in combination with the HM450K array annotation. However, considering the number of probes per gene vary, there is an inbuilt bias towards genes that have more probes within them. Missmethy1 gene ontology takes into account probe density per gene when calculating pathway/gene over-representation, thus making it the only true pathway analysis currently available compatible with array-based DNAm data.

Gene ontology terms (GO terms) have known genes associated with particular pathways. Over-representation of genes within pathways allow for insight into pathways that may be affected by differential methylation. Reduce + Visualise Gene Ontology (REViGO) was utilised to remove

redundancy in GO terms to ensure that appropriate ‘umbrella’ pathways associated with aberrant DNAm were being highlighted.

2.5.9 Additional Statistical Methods

All statistical tests were conducted in R using packages and codes described in Table 2.8. Comparisons of overall and event-free survival (OS/EFS) between groups used the Log-Rank Test. Comparisons of methylation between groups for each individual CpG site used the Student’s T-test in GraphPad Prism or R. Comparisons of categorical variables and epidemiological data (e.g. clinical features) between groups used Fisher’s Exact Test. Correlation matrices were created through the use of Pearson’s correlation coefficient. Receiver operating characteristic (ROC) analysis was utilised to calculate the area under the curve (AUC) and measure the predictive performance of models generated. Through the use of Weka and R, machine-based learning algorithms were tested for their predictive ability and returned Precision, Recall, False Positive Rate (FPR), F-Measure and AUC. Imputation of repeat sequences for the *Alu* gene were created through the REMP package for R.

Table 2.8 R packages used for statistical tests in this study

Statistical Test Used	R Package
Log Rank Test	survminer/survival
Fisher’s Exact Test	stats
Pearson’s Correlation Coefficient	corrplot
Receiver Operating Characteristic	pROC/ROCR
Precision/Recall/FPR/F-Measure & AUC	caret/randomForest/e1071

2.5.10 Graphical Methods

All graphs were produced using R or GraphPad Prism 7 and relevant packages are detailed in Table 2.9.

Table 2.9 R packages used for graphs presented in this study

Graph Type	R Package
Density/Scatter Violin Plot	ggplot2/ggsci/geom_flat_violin
Density Plot	ggplot2
Heatmap	heatmap2
Correlation matrix/Correlogram	corrplot
Boxplot	ggplot2
Line graph	ggplot2
Kaplan-Meier Plot	survminer/survival
Receiver Operating Characteristic Curve	pROC/ROCR
Unsupervised Principal Component Analysis Plot	WCGNA
Venn Diagram	VennDiagram
Confusion Matrix/Four-Fold Plot	caret
Hyperparameter Model Tuning	caret
t-Distributed Stochastic Neighbor Embedding Plot	Rtsne

This page is intentionally left blank.

Chapter 3 Interrogation of Global Methylation in Paediatric Acute Myeloid Leukaemia

3.1 Rationale

Global methylation describes the average methylation across the entire genome. Aberrant global DNAm has historically been associated with cancers of all types. For example, a globally hypomethylated profile is often associated with cancer, with widespread loss of methylation in oncogenes, retrotransposons, endogenous retroviral repeat elements and other satellite repeats (Feinberg and Vogelstein 1983). Conversely, targeted hypermethylation is also a feature of most cancers, particularly at transcription start sites and enhancer regions of tumour suppressor genes. While hypermethylation events are not as common as hypomethylation, they both contribute to an aberrant methylation profile; allowing cancer cells to proliferate and avoid apoptosis or immune detection (Costello, Frühwald et al. 2000). Global methylation can be measured using many assays, such as high-performance liquid chromatography (HPLC), array-based platforms such as the HM450K, or through proxy gene measures including methylation of *LINE1* and *Alu* repeat elements.

While global DNA hypomethylation is a common feature of most cancers, there are exceptions for certain cancers (Estecio, Gharibyan et al. 2007). For instance, globally hypermethylated profiles have been identified in pancreatic and breast cancer (Neale, Clark et al. 2014; Reyngold, Turcan et al. 2014). However, the significance of these total 5mC changes in cancer are yet to be fully elucidated (Kaneda, Tsukamoto et al. 2004; Stumpel, Schneider et al. 2013; Karsy, Guan et al. 2018). Changes in total 5mC have been shown to be associated with microsatellite instability (MSI) and loss of heterozygosity (LOH) in some cancers (Rodriguez, Frigola et al. 2006; Richards, Zhang et al. 2009). Matsuzaki *et al.* suggests that microsatellite stable (MSS) colorectal cancers have a hypomethylated total 5mC profile, and therefore increases in global methylation could contribute to microsatellite instability (Matsuzaki, Deng et al. 2005). Furthermore, global hypomethylation has also been associated with a LOH, and cancer patients with neither of these abnormalities have significantly higher total 5mC levels (Matsuzaki, Deng et al. 2005). One study investigating LOH in paediatric AML identified very few instances where there was complete LOH (deletion and SNP), although cases with a LOH had several recurring regions identified (Radtke, Mullighan et al. 2009). Studies of LOH induced by DNAm have yet to be explored in paediatric AML. Therefore, it is reasonable to

suggest that interrogating the global methylation profile of primary AML samples may allow us to better understand LOH in paediatric acute myeloid leukaemia patients.

Additionally, changes in global methylation have been associated with many diseases such as *Anaplasma phagocytophilum* pathogenic infection (Sinclair, Yegnasubramanian et al. 2015), high-grade cervical intraepithelial neoplasia caused by human papilloma virus infection (Barchitta, Quattrocchi et al. 2017) and gestational diabetes (Wu, Farrell et al. 2018). The true dynamic nature of global DNAm can also be shown in the methylation response to environmental influences, such as discordance in IVF-conceived twins compared to naturally conceived newborns (Loke, Galati et al. 2015), the effects of maternal alcohol intake on placental methylation (Loke, Muggli et al. 2018) and in the response to certain therapeutics, such as IFN β (Pinto-Medel, Oliver-Martos et al. 2017). While global DNAm analysis rarely reveals details about specific genes that are dysregulated, its use as a biomarker through proxy measures can identify disease state and implicates aberrant DNAm as a contributor to disease (Neale, Clark et al. 2014; Barchitta, Quattrocchi et al. 2017)

In the current study we attempt to define the global methylation status of diagnostic and matched remission samples in paediatric AML using two widely accepted proxy measures. We also investigate the association of global methylation levels with different AML subtypes and LOH status. This was done by deriving a Genome-Wide Average Methylation (GWAM) value across all Infinium Bead Chip probes or the proxy measure of global methylation based on repetitive *LINE1* and *Alu* elements. By defining differences in global methylation, the association was tested for utility to define patient subtype and risk of poor outcome at diagnosis, as well as exploring whether global methylation could be used as a marker of disease.

We hypothesised that:

- (i) Childhood acute myeloid leukaemia will have disrupted total 5mC levels;
- (ii) Unique global methylation signatures can be attributed to different subtypes of AML, and;
- (iii) The total 5mC profile of patients with LOH will be lower than that of patients without LOH.

3.2 Study Cohort

This study used 198 diagnostic and 143 matched remission bone marrow samples from two separate cohorts of childhood AML, along with publicly available DNAm data from 24 healthy, cell sorted bone marrow samples derived from 5 participants for a total cohort size of 365 samples from 231 participants (Table 3.1). Of all remission samples used, 26 did not have a matched sample at another time point. Diagnostic bone marrow samples were assayed for standard cytogenetic abnormalities after collection. The most common fusion genes associated with AML are well represented in this cohort, but the cohort also includes rare and novel translocations (Table 3.1). The majority of patients were between 3-15 years of age. Males were slightly over-represented, as expected by epidemiological studies into paediatric AML (WHO 2014).

Clinical patient characteristics were obtained from the CLARA database for RCH (VIC) cases while other cohort information was received upon receipt of samples from their respective biobanking repositories. All data is correct as of April 2018.

Table 3.1 Clinical characteristics of diagnostic, matched remission and cell sorted, healthy bone marrow samples from the Australian and TARGET AML cohorts.

Clinical Characteristics	Diagnostic Samples (n = 198)	Remission Samples (n = 143)	Relapse Samples (n = 60)	Healthy Bone Marrow Samples (%) (n = 24)
Sex				
Male	105 (53.0%)	69 (48.3%)	29 (48.3%)	Unknown
Female	93 (47.0%)	74 (51.7%)	31 (51.7%)	Unknown
Age at diagnosis				
Infants (<3 y/o)	56 (28.3%)	37 (25.9%)	19 (31.7%)	Unknown
Children (3-15 y/o)	109 (55.1%)	80 (55.9%)	30 (50.0%)	Unknown
AYA (15+ y/o)	33 (16.6%)	26 (18.2%)	11 (18.3%)	Unknown
Outcome				
Good	46 (23.2%)	41 (28.7%)	0	Unknown
Poor	143 (72.2%)	102 (71.3%)	60 (100%)	Unknown
Unknown	9 (4.6%)	0	0	Unknown
Cytogenetics				
inv(16)/t(16;16)	16 (8.1%)	7 (4.9%)	6 (10.0%)	0
t(8;21)	11 (5.6%)	7 (4.9%)	3 (5.0%)	0
<i>KMT2A</i> rearranged	41 (20.7%)	33 (23.1%)	7 (11.7%)	0
Trisomy 8	9 (4.6%)	4 (2.7%)	1 (1.7%)	0
Trisomy 21	2 (<1%)	0	1 (1.7%)	0
CN	47 (23.7%)	43 (30.1%)	8 (13.2%)	24 (100%)
Complex karyotype *	14 (7.1%)	5 (3.5%)	1 (1.7%)	0
Other [‡]	44 (22.2%)	39 (27.3%)	33 (55.0%)	0
Unknown	14 (7.1%)	5 (3.5%)	0	0
Sample source				
RCH(VIC)	41 (20.7%)	12 (8.4%)	10 (%)	0
WM	38 (19.1%)	15 (10.5%)	9 (%)	0
RCH(QLD)	21 (10.6%)	13 (9.1%)	2 (%)	0
PMCH	9 (4.6%)	0	0	0
TARGET AML	89 (45.0%)	103 (72.0%)	39 (%)	0
Other	0	0	0	24 (100%)

Abbreviations: EFS - Event free survival; y/o - Years old; AYA - Adolescent/Young adult; inv - Inversion; t - Translocation; CN - Cytogenetically normal; RCH - Royal Children's Hospital; VIC - Victoria; WM - Westmead; QLD - Queensland; PMCH - Princess Margaret Children's Hospital; TARGET AML - Therapeutically Applicable Research To Generate Effective Treatments in Acute Myeloid Leukaemia.

* Complex karyotypes are defined by patients having two or more cytogenetic abnormalities or abnormalities that are otherwise unrelated to AML pathogenesis. [‡]Other cytogenetic abnormalities are defined as del5/7/9/13/20q, inv(5q), -7, t(5;6), t(6;9), t(7;14), t(15;17), t(16;21). Genetic abnormalities are mostly undefined.

3.3 Genome-wide Average Methylation in Paediatric Acute Myeloid Leukaemia

3.3.1 Comparison between Diagnostic, Remission and Healthy Bone Marrow Samples - Discovery & Pooled-Analyses

Genome-wide average methylation was initially used to determine if there were significant differences in total 5mC between diagnostic, remission and healthy bone marrow samples. Analysis of the average methylation β -value of 373,253 450K methylation probes in the Australian cohort (n=91) showed that the overall distribution of methylation was similar between sample types. However, diagnostic samples trended towards a slightly more hypermethylated profile compared to remission samples, with a 3.72% increase in GWAM-derived 5mC (Figure 3.1A, $p < 1.0 \times 10^{-8}$, Student's t-test). This difference is encapsulated by the 5330 probes that went from hypomethylated (<50% methylation) to hypermethylated (>50% methylation).

Although the overall bimodal distribution of methylation was similar for all sample types (diagnostic, remission and control), the density of methylation at the ends of the distribution varied. Many more probes were hypomethylated in remission and healthy samples compared to diagnostic bone marrow. This is primarily seen as an increase in intermediate methylation density (β -values= 0.25 ± 0.05), rather than an increase in the number of highly methylated probes (β -value=0.80) in diagnostic samples compared to control/remission samples (Figure 3.1B).

The hypermethylated probes also varied across the sample types. While density of probes is similar between diagnostic and remission samples, the shift of methylation peaks suggest that many probes have become less methylated during disease progression and treatment (Figure 3.1). Furthermore, considering the similarity of bimodal distribution of methylation between all sample types, this suggests that specific genes and pathways have been targeted rather than a global dysfunction of DNAm machinery, which would change the distribution entirely.

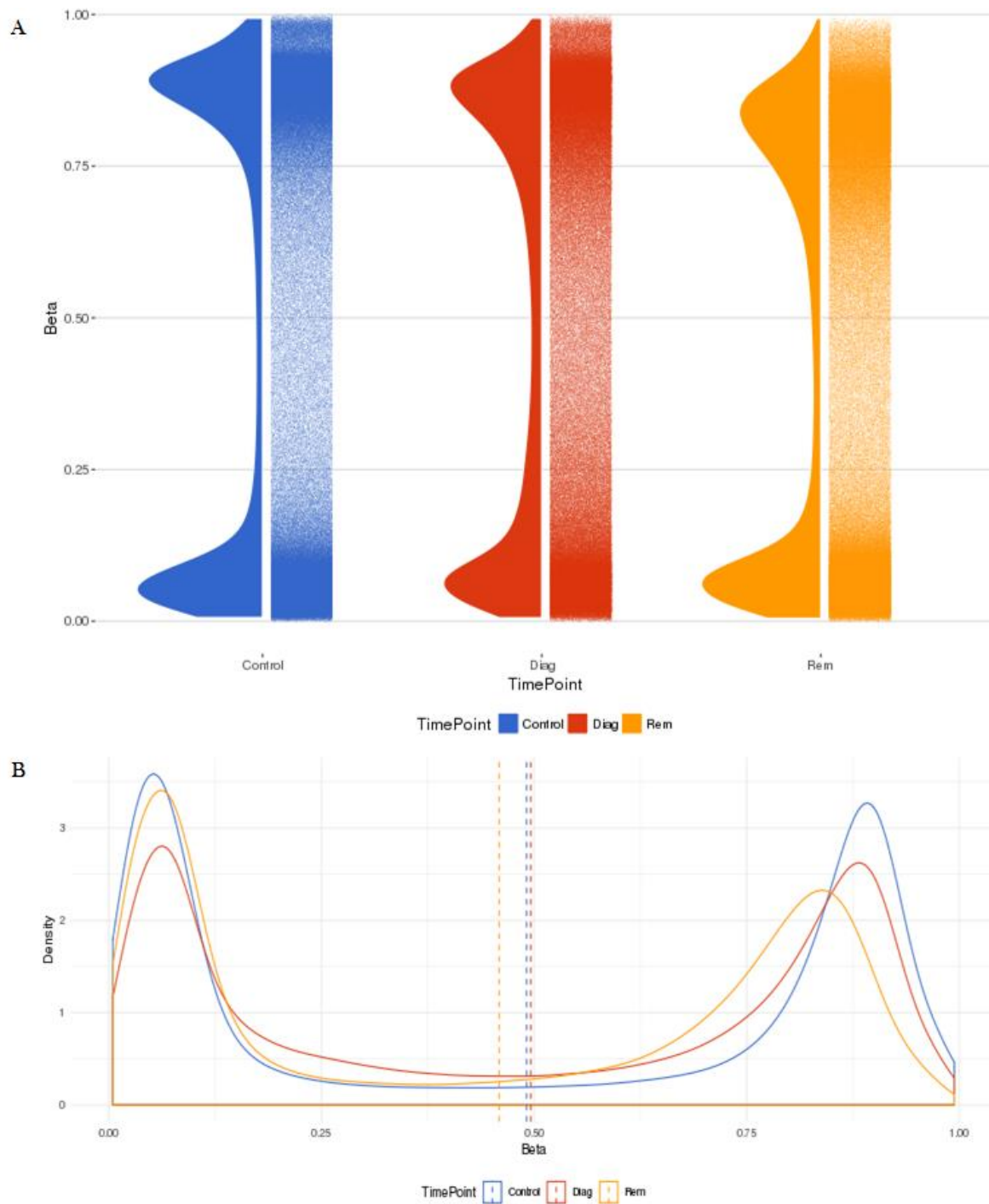


Figure 3.1 Inspection of global methylation differences between diagnostic, remission and healthy bone marrow samples through density and scatter analysis
Caption on next page.

Beta values from all 373,253 probes were averaged for each sample type (n=131, 82 diagnostic, 25 remission and 24 control samples) and used to produce A) Vertical density /scatter plot show each sample types typical bimodal distribution of methylation, with the majority of probes being hypo or hyper methylated (<0.2, corresponding to 20% or >0.8, 80% respectively). B) When overlayed, there is a clear shift in the relative density of probes between diagnostic and remission samples, with a loss of hypomethylated probes and a clear gain of hypermethylated probes. Density plots show little difference in the bimodal distribution of beta values between sample types but do suggest that the extent to which they are methylated is defined by the time point at which they are taken ($p < 1.0 \times 10^{-8}$). Scatterplot does not reveal any underlying patterning concealed by density analysis.

Pooled-analysis of global methylation in a combined cohort (n=298) yielded similar results to the Australian discovery cohort, with a clear bimodal distribution of methylation consisting of peaks around 10% and 90% for all samples time points. However, the density of methylation at these peaks differed between groups (Figure 3.2A). The average total 5mC difference between diagnostic and remission samples remained consistent between analyses at 2.42% (Figure 3.2B, $p < 1.0 \times 10^{-8}$). The ends of the bimodal distribution showed similar patterns to that of the original discovery cohort findings (Figure 3.2B). Overall, we find a shift of 4587 probes going from hypomethylated (<50% methylation) to hypermethylated (>50% methylation) between diagnostic and remission samples.

The addition of 109 samples from the TARGET cohort had very little impact upon the average diagnostic beta value (0.43% increase), although there was a 1.73% average increase in total 5mC for remission samples and a higher density of probes at higher degree of methylation compared to remission samples solely used in the discovery analysis. Considering that remission samples were taken at different time points in each cohort (TARGET - Day 14, AUS - Day 30), this data suggests that total 5mC is progressively decreased throughout treatment before returning to a 'normal' global profile, as defined by healthy samples. As probes at the hypermethylated end of the spectrum are slightly less methylated in diagnostic samples, and even less in remission samples compared to healthy controls, further justifies the need to identify what individual probes are being hypomethylated before potentially returning to a normal level of methylation.

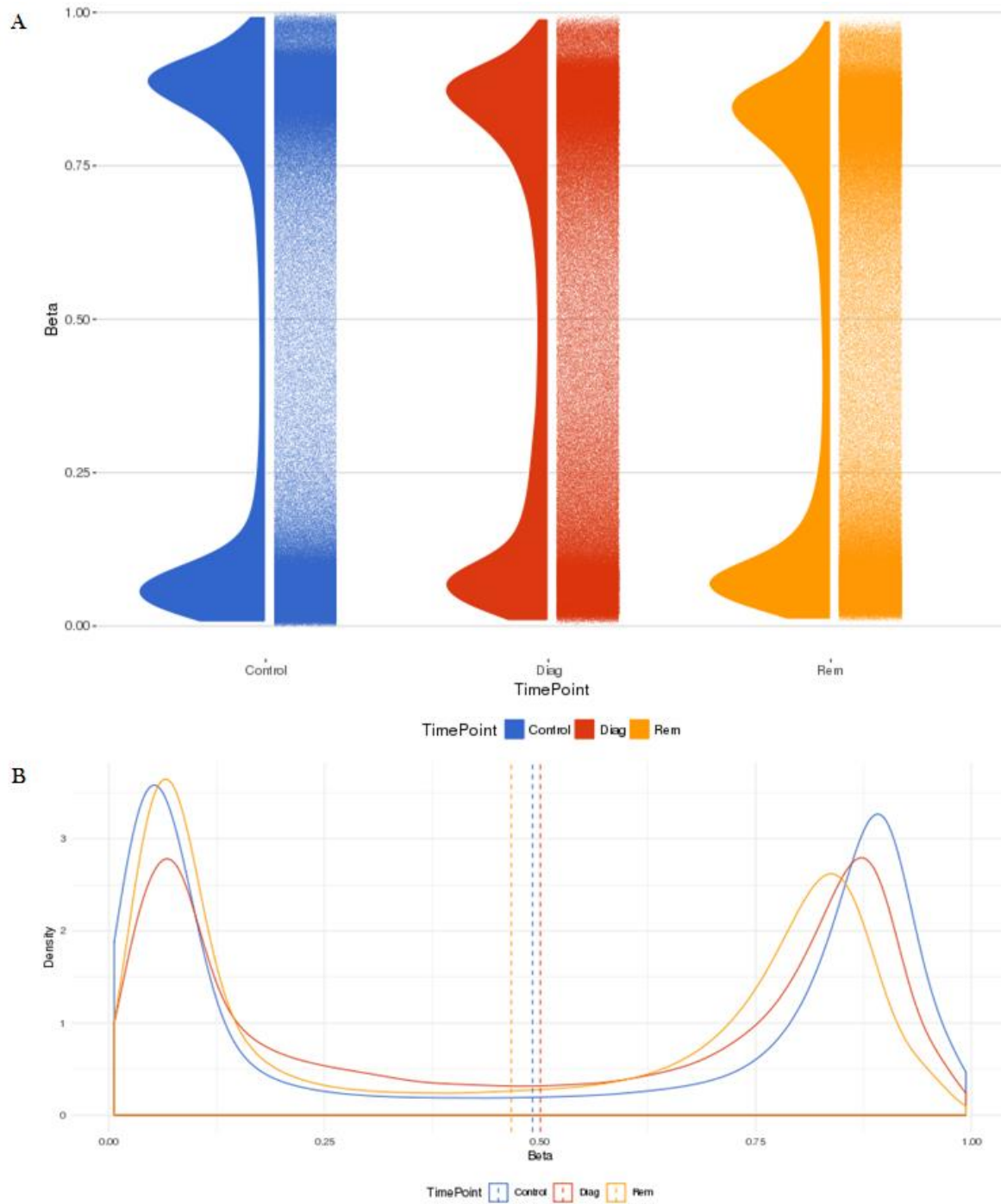


Figure 3.2 Inspection of global methylation differences between diagnostic, remission and healthy bone marrow samples in a pooled cohort through density and scatter analysis

Caption on next page.

Beta values from all 373,253 probes were averaged for each sample type (n = 298, 171 diagnostic, 103, remission and 24 control samples) and used to produce A) Vertical density/scatter plot show each sample types typical bimodal distribution of methylation, with the majority of probes being hypo or hyper methylated (<0.2, corresponding to 20% or >0.8, 80% respectively). B) When overlayed, there is a clear shift in the relative density of probes between diagnostic and remission samples with a loss of hypomethylated probes and a clear gain of hypermethylated probes. Density plots show little difference in the bimodal distribution of beta values between sample types but do suggest that the extent to which they are methylated is defined by the time point at which they are taken ($p < 1.0 \times 10^{-8}$). Scatterplot does not reveal any underlying methylation patterning concealed by density analysis.

As expected, when comparing diagnostic to matched relapse samples in the pooled cohort, we find that there is only a small difference in GWAM ($\Delta\beta = 0.046\%$, $p < 1.0 \times 10^{-8}$). Although the difference was statistically significant, it is hard to image that such a small overall change in GWAM would prove any large-scale functional consequence. More interestingly, this finding shows that, while diagnostic GWAM is higher than remission, this level of GWAM returns upon relapse showing the plasticity of DNAm over a small period of time (average time to relapse being 3 months), and that GWAM is consistent between multiple AML sample types. This suggests that, while aberrant on a genome-wide scale, precise methylation changes are occurring to provide an optimal leukaemic cell environment. Individual gene methylation contribution is further explored in Chapter 4.

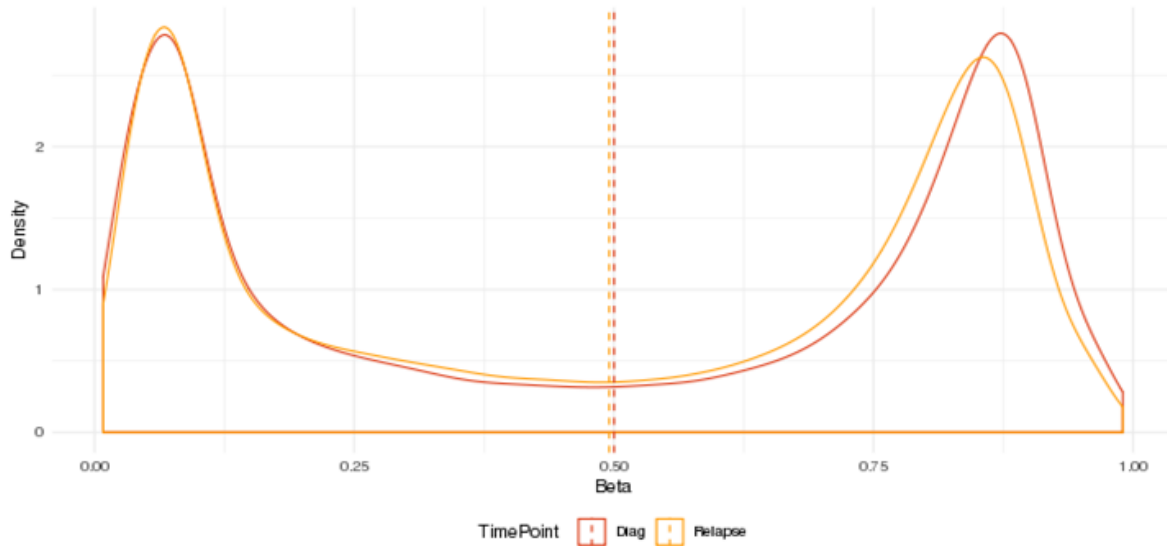


Figure 3.3 Inspection of global methylation through density analysis between diagnostic and relapsed paediatric AML samples

Beta values from all 373,253 probes were averaged for each sample type ($n = 231$, 171 diagnostic and 60 relapse samples) and used to produce an overlaid density plot where each sample time points shows a typical bimodal distribution of methylation, with the majority of probes being hypo or hyper methylated (<0.2 , corresponding to 20% or >0.8 , 80% respectively). There is little shift in overall GWAM between diagnostic and relapsed samples ($\Delta\beta=0.046\%$, $p<1.0\times10^{-6}$) suggesting that the GWAM profile does not differ upon recurrence.

Global methylation was broken down into genomic regions in accordance with the IlluminaHumanMethylation450k manifest (build hg19) to determine whether the location of probes relative to specific genomic features impacted their likelihood of altering methylation in transition to AML (Figure 3.4). Generally, probes within 1500bp or 200bp of the transcription start, a 5' untranslated region, or the first exon of a gene were hypomethylated in all samples, irrespective of disease state. In contrast, probes in the gene body, 3'UTR or intergenic regions were hypermethylated according to the median. Following transition from diagnosis to remission, all probes, irrespective of their genomic location relative to genes, showed a decrease in median methylation, consistent with a global hypomethylation in remission of AML. In other words, the total 5mC hypermethylation event previously described between diagnostic and remission samples was evident within all regions of the genome. All regions were identified as significantly differentially hypermethylated in diagnostic samples and were most pronounced within intergenic regions (Table 3.2, average $\Delta\beta=2.9\%$, $p<1.95\times10^{-8}$) and the 1st exon of genes (Table 3.2, average $\Delta\beta=2.57\%$, $p<8.69\times10^{-11}$). Global density analysis of genomic features mirrors findings of total 5mC density analysis previously discussed.

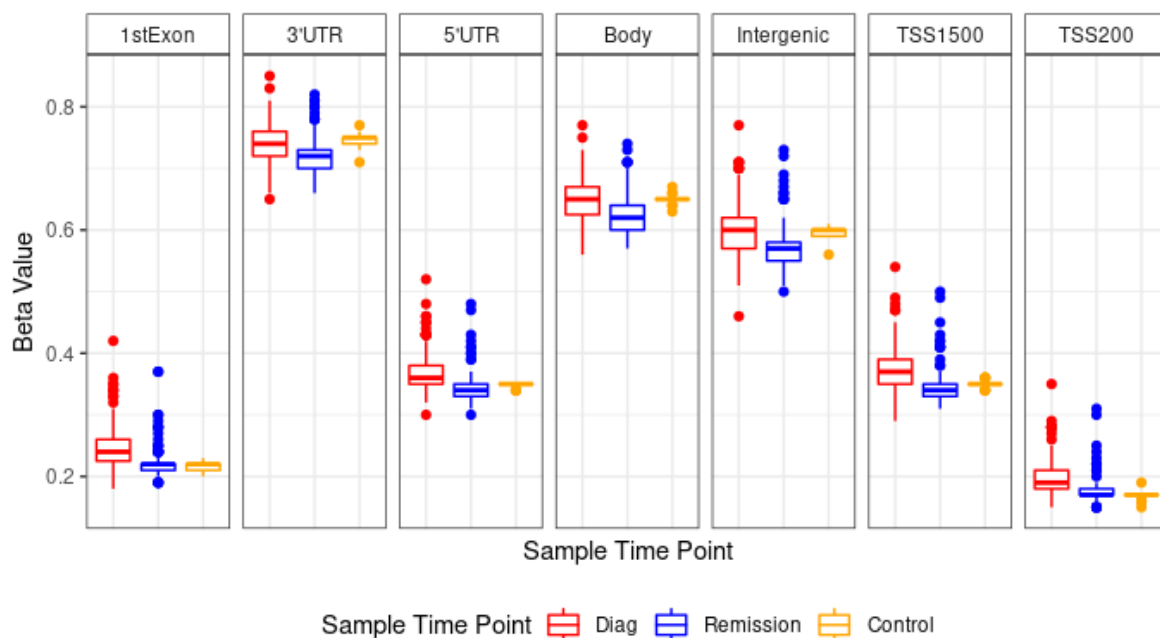


Figure 3.4 Boxplot of global methylation differences between diagnostic and remission sample restricted to genomic region in a pooled AML cohort

All 373,253 probes were sorted according to genomic location in accordance with the IlluminaHumanMethylation450kmanifest (build hg19). Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between sample time points (n=324 total samples). All genomic regions show a slightly more hypermethylated profile in diagnostic AML samples when compared to remission and control samples ($p < 1.0 \times 10^{-6}$).

Table 3.2 Comparison of GWAM between diagnostic and remission AML bone marrow samples restricted by genomic feature

Region	Diagnostic Average β	Remission Average β	$\Delta\beta^*$	p-value [†]
TSS1500	37.26	34.96	2.3	6.49×10^{-8}
TSS200	19.66	17.72	1.94	4.71×10^{-9}
5'UTR	36.79	34.45	2.34	5.49×10^{-10}
1 st Exon	24.76	22.19	2.57	8.69×10^{-11}
Body	64.78	62.41	2.37	5.37×10^{-9}
3'UTR	73.83	71.75	2.08	3.377×10^{-7}
Intergenic	59.97	57.07	2.9	1.95×10^{-8}

Abbreviations: TSS - Transcription start site; UTR - Untranslated region; β - Beta; $\Delta\beta$ - Delta beta (Difference of methylation)

*Delta beta of diagnostic average minus remission average.

[†]Student's t-test between diagnostic and remission samples.

3.3.2 Global Methylation Levels Assessed Using *LINE1* and *Alu* between Diagnostic and Remission Bone Marrow Samples in Paediatric AML

In addition to assessing global methylation by deriving an average methylation score, global methylation levels were also assessed through locus-specific methylation analysis using both imputed HM450K data and repeat-specific amplification assays, MassARRAY chemistry and SEQUENOM. Specifically, we used previously validated assays of *LINE1* and *Alu*, which are repetitive elements that represent over 30% of the human genome. These elements are known for their generally stable methylation levels and, because of their highly repetitive nature, can serve as a proxy for total 5mC. Utilising the **REMP** package in **R**, we were able to impute methylation levels of both *Alu* and *LINE1* in thousands of CpG sites across all diagnostic, remission and control samples.

REMP was able to impute 2462 *Alu* CpG sites across 320 bins (*Alu* repeats within a 1000bp flanking region) after quality control measures were considered. In *Alu* we observed a 1.5% increase in diagnostic methylation compared to remission samples (Figure 3.5), in which findings further corroborate that at diagnosis, paediatric AML is globally hypermethylated. Control samples were consistently methylated with a similar average, yet vastly different distribution when compared to diagnostic samples in both genes. This is largely due to the control samples being derived from five patients; thus the global methylation profile would be more uniform.

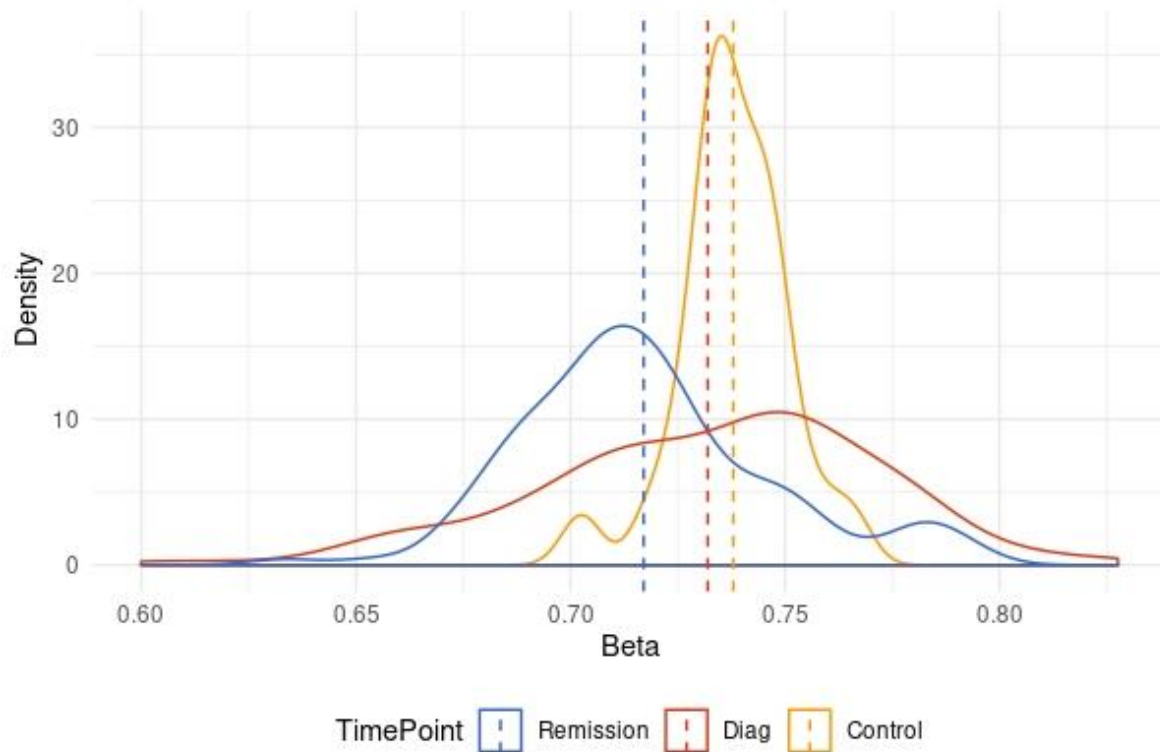


Figure 3.5 Density plot of average imputed methylation of 2462 Alu CpG sites across the whole genome in diagnostic, remission and healthy bone marrow samples

Beta values of 2462 imputed Alu CpG sites across the genome were interrogated for their methylation and averaged between diagnostic (red), remission (blue) and control (yellow) samples. Both diagnostic and remission samples varied in *Alu* methylation, however there was an average hypermethylation of 1.5% in diagnostic samples. Control samples were slightly more hypermethylated than diagnostic samples, but the difference is negligible and similar to findings of total HM450k global methylation.

The *LINE1* assay generated data for 3 individual CpG sites, which was then combined to generate an average methylation level. We compared the methylation of each individual CpG site between sample types (n=100; 73 diagnostic, 27 remission). Median methylation levels for all sites and the average were between 0.6 and 0.7 (60%–70% respectively), in both diagnostic and remission samples.

LINE1 was significantly differentially methylated between diagnostic and remission samples at all CpG sites, with an average beta difference of 1.8% (Figure 3.6, $p=0.005$, Student's t-test).

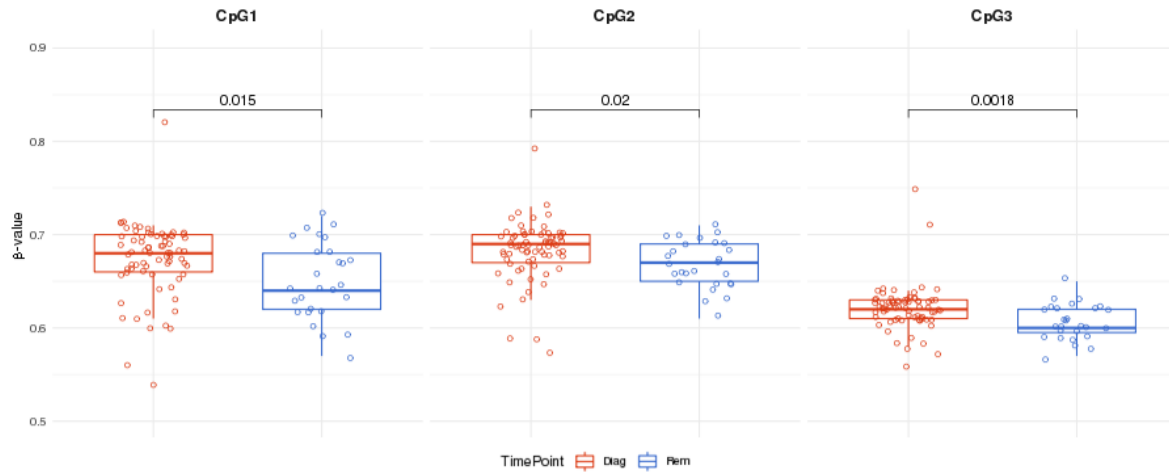


Figure 3.6 Boxplot of differential methylation between patient time points at the global methylation proxy *LINE1*

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation within patient outcome groups (n=100, 73 diagnostic, 27 remission). Outliers lie beyond the whiskers. All 3 sites assayed in *LINE1* and the average were significantly differentially methylated between diagnosis and remission ($p \leq 0.05$, Student's t-test).

The *A/u* assay generated data for a total of 19 CpG sites, of which 10 passed quality control. Of these, CpG2, CpG6, CpG12, CpG13, CpG14 and CpG17.18 were significantly differentially methylated between diagnostic and remission samples ($p \leq 0.05$, Student's t-test). The average of all 10 CpG sites was significantly different between time points, showing a 1.1% difference in DNAm (Figure 3.7, $p = 0.009$, Student's t-test). With both this independent validation and the imputed *A/u* data previously discussed, we find a 0.4% difference between the two analyses, both suggesting that diagnostic samples are hypermethylated in paediatric AML. Discrepancies between actual beta values can be explained by data acquisition of different CpG sites for both analyses.

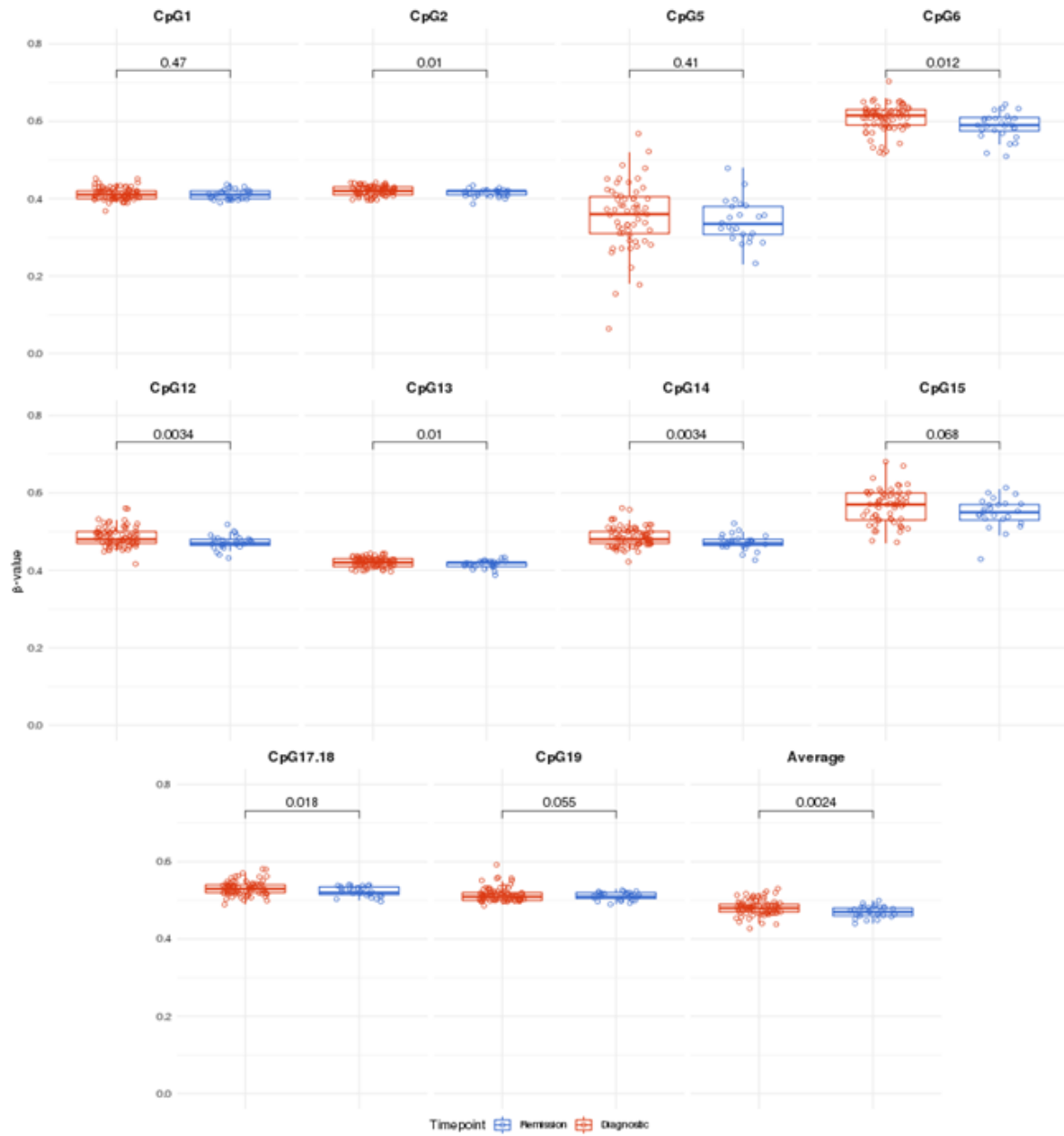


Figure 3.7 Boxplot of differential methylation between patient time points at the global methylation proxy *Alu*

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation within patient outcome groups (n=100). Outliers lie beyond the whiskers. Of the 10 sites that passed quality control, 6 sites in *Alu* and the average were significantly differentially methylated between diagnosis and remission ($p \leq 0.05$, Student's t-test).

Correlation analysis was employed to determine whether these genes would validate genome-wide findings. Average *LINE1* methylation was significantly positively correlated with total 5mC derived from the HM450k array (Figure 3.8A, $R=0.40$, $p=0.004$). While not significant, *Alu* methylation showed an overall weak positive correlation with total 5mC (Figure 3.8B, $R=0.26$, $p=0.07$). Division of samples showed that there was little difference, and no driving force in the correlative effects of either *LINE1/Alu* with total 5mC between time points.

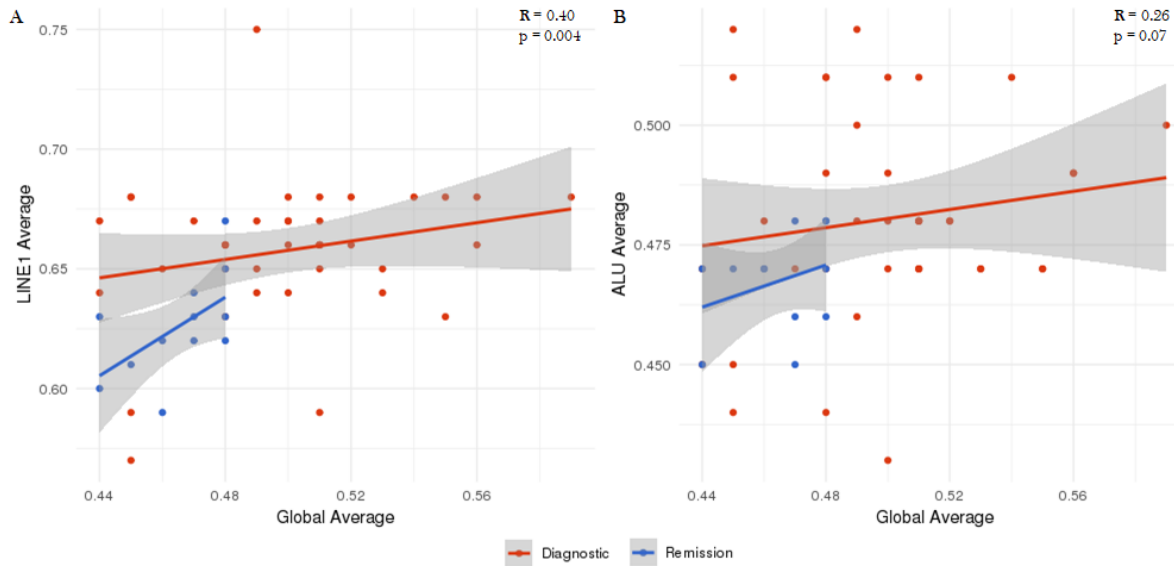


Figure 3.8 Correlation of average *LINE1/Alu* locus-specific methylation with global methylation from the HM450k array

Average locus specific methylation of *LINE1/Alu* of 3 and 10 CpG sites respectively, positively correlate with global methylation derived from the HM450k array. Line of best fit and error bars (standard error) were used to demonstrate correlation. Pearson's correlation coefficient was determined for both A) *LINE1* ($R=0.40$, $p=0.004$) and B) *Alu* ($r=0.26$, $p=0.07$) average methylation, suggesting that a respective positive and weak positive association exists between both genes with total 5mC determined by HM450.

We also compared average methylation of *LINE1* and *Alu* in diagnostic and matched relapse samples (Figure 3.9). Interestingly, we found that there was no significant difference in average DNAm of both *LINE1* ($p=0.32$, Student's t-test) and *Alu* ($p=0.35$, Student's t-test). Furthermore, each individual CpG site of both *LINE1/Alu* were also not significantly different between diagnostic and relapse samples. This finding further solidifies the hypothesis that DNAm changes between diagnostic and remission time points, and that the DNAm profile returns to a similar level found in diagnostic samples after relapse.

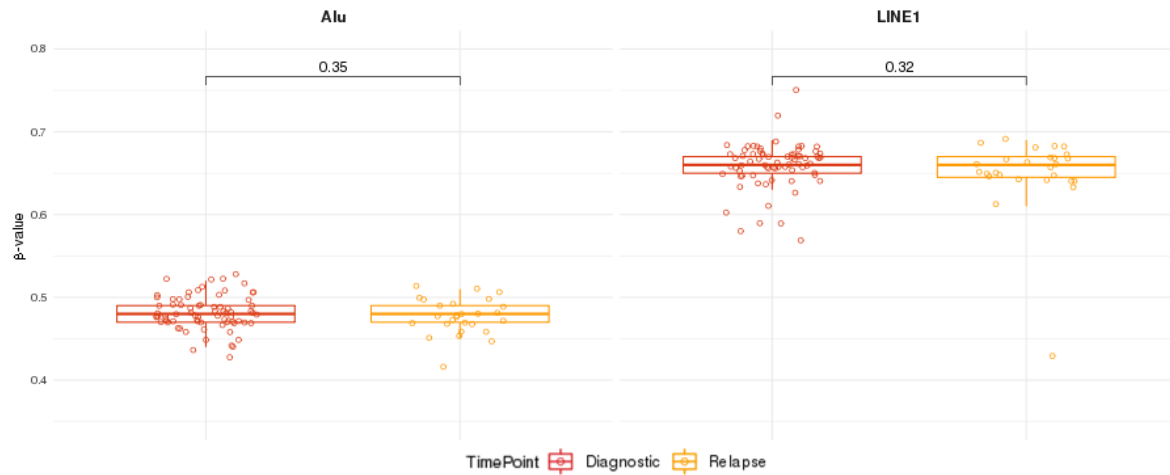


Figure 3.9 Boxplot of average differential methylation between diagnostic and matched relapsed AML samples at the global methylation proxies *LINE1* and *Alu*

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation within patient outcome groups (n=78). Outliers lie beyond the whiskers. Average methylation of all sites that passed quality control in A) *LINE1* and B) *Alu* were compared between diagnostic (n=73) and relapsed samples (n=27, 22 matched). Both *LINE1* (p=0.32, Student's t-test) and *Alu* (p=0.35, Student's t-test) showed no significant methylation difference between patient time points.

3.4 Global Methylation Levels in Association with Specific Genetic Lesions

3.4.1 Comparison of Global Methylation between AML Subtypes

Traditionally, acute myeloid leukaemia is classified into several subtypes according to specific chromosomal aberrations and genetic mutations. The resulting individual phenotype of these subtypes often involves disruption of distinct cellular pathways (Pui, Carroll et al. 2011). To date, the relative impact of specific chromosomal aberrations on global methylation has not been investigated in paediatric AML.

We interrogated 6 different subtypes of AML for their global methylation profiles (Figure 3.10) and found that there were significant differences in total 5mC as assessed by GWAM ($n=121$, $p<1.0\times10^{-8}$, ANOVA). All subtypes average total 5mC fell within a $\pm 2\%$ range of one another. Both the *inv(16)* ($n=15$, avg. $\beta=50.4\%$) and the cytogenetically normal ($n=43$, avg. $\beta=50.1\%$) subtypes showed a significant differential methylation when taking the average methylation of each individual patient subgroup and comparing them to control bone marrow (Table 3.3, $p<0.05$, Student's t-test). Trisomy 8 ($n=5$, avg. $\beta=50.6\%$) showed a similarly large $\Delta\beta$ as the aforementioned subtypes, however significance was not achieved primarily due to low patient numbers. The *KMT2A* ($n=37$, avg. $\beta=48.9\%$), complex karyotype ($n=12$, avg. $\beta=50.0\%$), and *t(8;21)* ($n=9$, avg. $\beta=49.4\%$) subtypes of AML all showed similar profiles to one another, in which these probes were irregularly distributed with more probes deviating from the two major hyper/hypomethylated peaks towards the mean, potentially suggesting a more disturbed or erratic biological pathway to (de)methylate DNA. The average methylation of these subtypes were more similar to control bone marrow averages than the other subtypes (Table 3.3).

This data suggests that each subtype of AML has their own methylation profile with individual differences that could potentially define disease pathways and help identify patient subtype in unknown or more complex cases (explored further in Chapter 5).

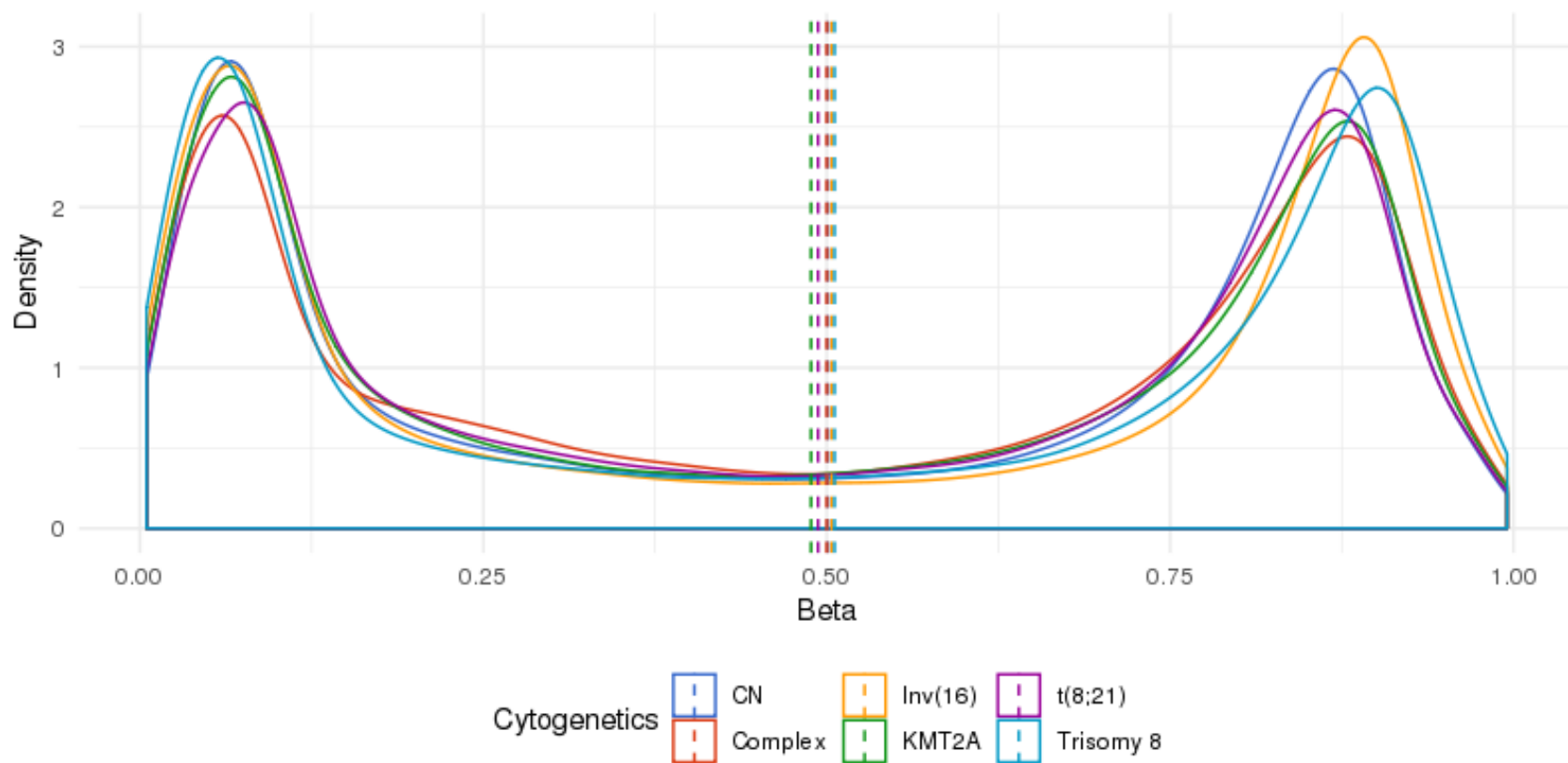


Figure 3.10 Interrogation of global methylation between AML subtypes through density analysis

Beta values from all 373,253 probes were averaged between AML subtypes (n=121). Overlaid density plots were used to define the global methylation landscape for each analysis. Significant differences in average global methylation were discovered between subtypes ($p < 1.0 \times 10^{-8}$, ANOVA). The inv(16) subtype (n=15) had the most normally distributed probes amongst all subtypes, however it also had the highest hypermethylated peak too. The KMT2A (n=37), Complex (n=12) and t(8;21) (n=9) subtypes had the most irregular methylation distribution.

Table 3.3 Comparison of GWAM of patient subtypes against control bone marrow samples

Subtype (<i>n</i>)	Average β	$\Delta\beta^*$	p-value [†]
Cytogenetically normal (<i>43</i>)	50.08	0.96	0.048
inv(16) (<i>15</i>)	50.35	1.23	0.001
t(8;21) (<i>9</i>)	49.37	0.25	0.823
<i>KMT2A</i> (<i>37</i>)	48.86	-0.26	0.67
Trisomy 8 (<i>5</i>)	50.57	1.45	0.428
Complex karyotype (<i>12</i>)	50.00	0.88	0.533

Abbreviations: inv – Inversion; t – Translocation; β – Beta; $\Delta\beta$ – Delta beta (Difference of methylation)

*Delta beta compared to control bone marrows (n=24, Average β =49.12).

†Student's t-test between subtype and control samples.

3.4.2 Effects of Loss of Heterozygosity upon Global Methylation in Paediatric Acute Myeloid Leukaemia

Global methylation, as assessed by an array-based platform (GWAM), allows for detection of copy number variation. Given that AML is a cancer commonly shown to have chromosomal aberrations, variation in chromosomal copy number, including deletions and amplifications, are likely contributors to disease (Tarlock, He et al. 2017). Furthermore, as one of the known features contributing to tumorigenesis in a number of cancers is hypermethylation of tumour suppressor genes (Herman 1999), it is feasible that a combination of DNA hypermethylation of one allele, coupled with deletion of the remaining allele could potentially lead to a complete suppression of expression. This has previously been demonstrated in several cancers, and represents an alternative version of the classical Knudson two-hit hypothesis, ultimately leading to a LOH. Despite this, it remains unclear if such a combined genetic/epigenetic silencing of genes is associated with global 5mC levels in AML.

Copy number variation data, including binning and segmentation data, was generated for each diagnostic sample from 450K methylation data and compiled for Genomic Identification of Significant Targets in Cancer (GISTIC) analysis. GISTIC analysis (and individual CNV validation) revealed 11 heterozygous genomic regions that were frequently deleted in the combined AML cohort (Figure 3.11, $q\text{-value} \leq 10^{-10}$). These regions contained a total of 135 genes, of which 40 genes were hypermethylated (Figure 3.12, average $\beta \geq 75\%$ in promoter or enhancer regions). These 40 genes were run through an enrichment analysis where ‘regulation of multicellular organismal processes’ was the sole upregulated biological process identified (Gene Ontology Consortium, AmiGO 2). The results of the enrichment analysis suggest that 71 patients contain both a deletion of a gene and hypermethylation in the promoter/enhancer of the same gene, resulting in a potential loss of heterozygosity.

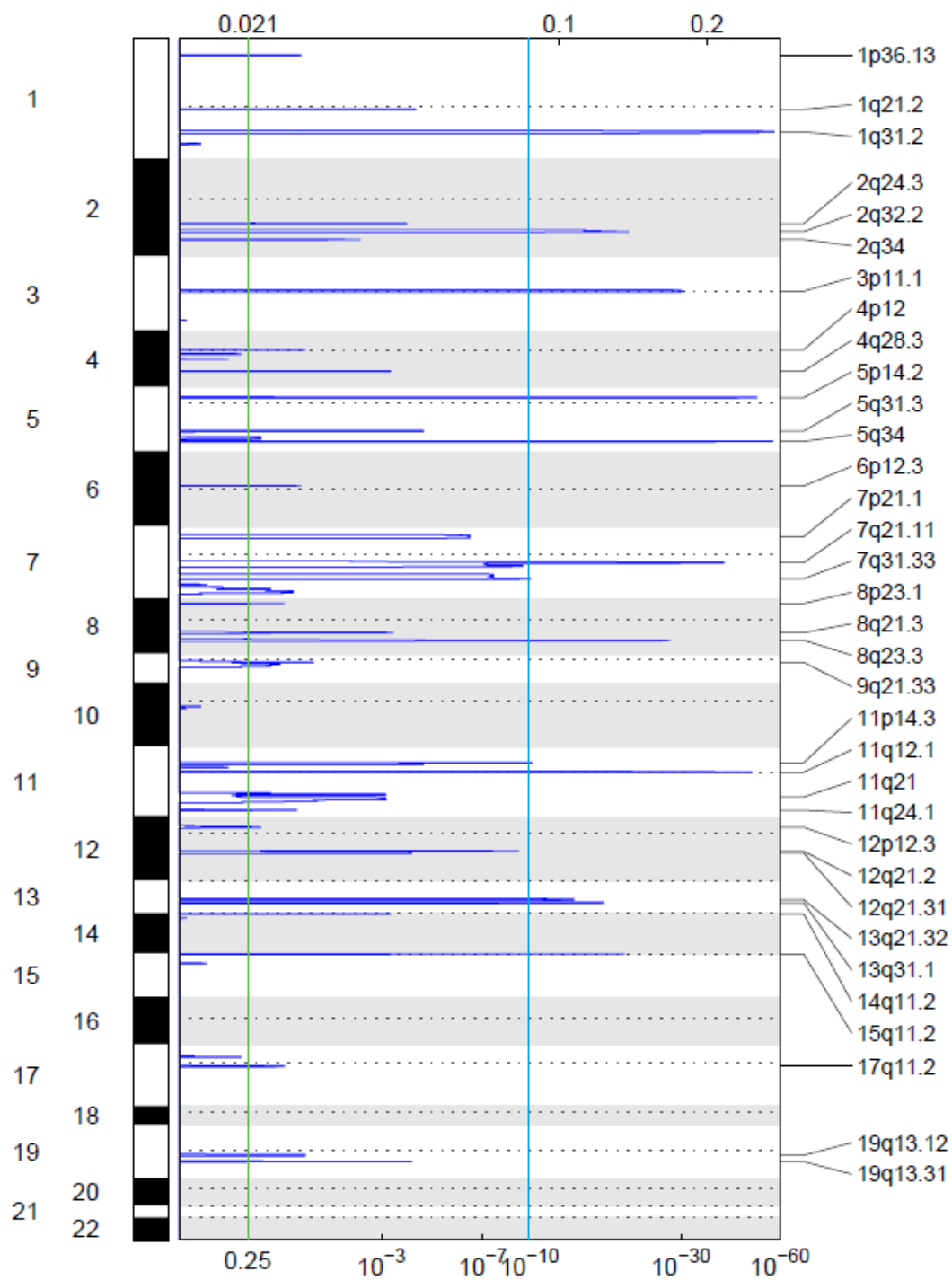


Figure 3.11 GISTIC deletion plot of all diagnostic pooled AML samples
Caption on next page.

Deletion plot showing which sections of chromosomes (right annotation) are most commonly deleted on each chromosome (left annotation) in the pooled AML cohort (n=171). G-scores (a unique GISTIC derivative considering the amplitude of the aberration as well as the frequency) shown across the top, measuring the degree of amplitude and frequency within the entire cohort. False discovery rate q-values are calculated for aberrant regions and are presented as $-\log_{10}(q)$. Default FDR q-value significance is set to 0.25 (green line), however for this analysis, manual significance threshold was set to 10^{-10} (blue line). All genes found within the regions below the set manual significance threshold were taken for downstream LOH analysis. Genomic peak coverage is defined by automated binning algorithms found within the conumee R package.

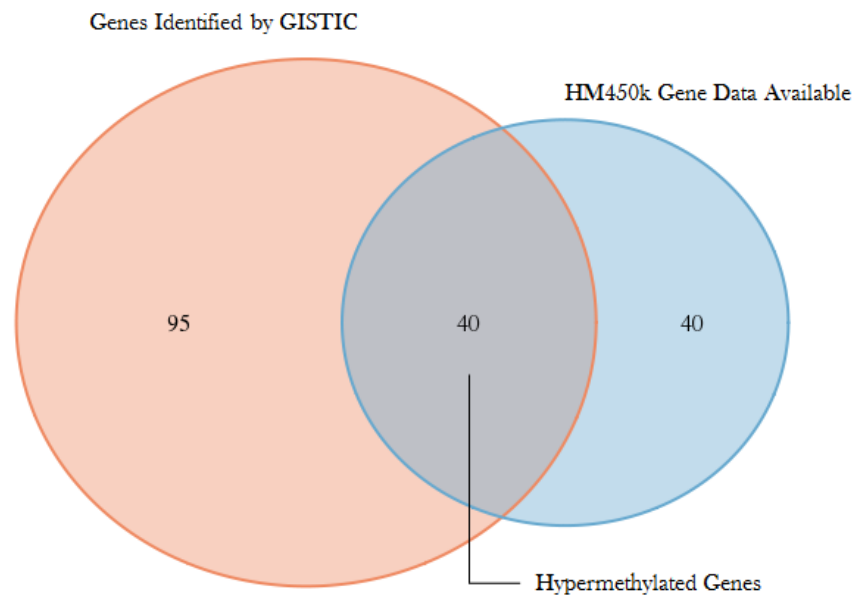


Figure 3.12 Venn diagram of genes that were both identified in the GISTIC analysis and had hypermethylated HM450k data available

Of the 11 regions identified in GISTIC analysis as serially deleted in the pooled AML cohort, 135 genes were identified to be in these regions with a deletion in at least one sample. Of these 135 genes, 80 had HM450k data available, 40 of which were hypermethylated in promoter or enhancer regions of the gene. Patients with both a deletion and hypermethylation in these 40 genes are noted to be at risk of DNAm associated LOH.

Patients in the combined AML cohort previously identified by GISTIC and CNV variation analyses as having one of the former 11 regional genomic deletions (as defined by conumee binning) had decreased GWAM levels compared to cases that did not show such genetic variation (Figure 3.13 A, average $\Delta\beta=0.5\%$, $p<1.0\times10^{-8}$). However, samples from the TARGET AML cohort, which were often identified to have allelic deletions and constitutes patients diagnosed with rare or difficult to treat AML, have an increased global methylation profile. We found a significant difference between cohorts (Figure 3.13B, average $\Delta\beta=0.8\%$, $p<1.0\times10^{-8}$) suggesting that there is an association of increased average global methylation with patients of poor outcome (further explored in Chapter 6). These two findings are seemingly at odds with

the suggestion that the minimal changes in total 5mC (<1%) found are fairly inconsequential to overall disease pathogenesis.

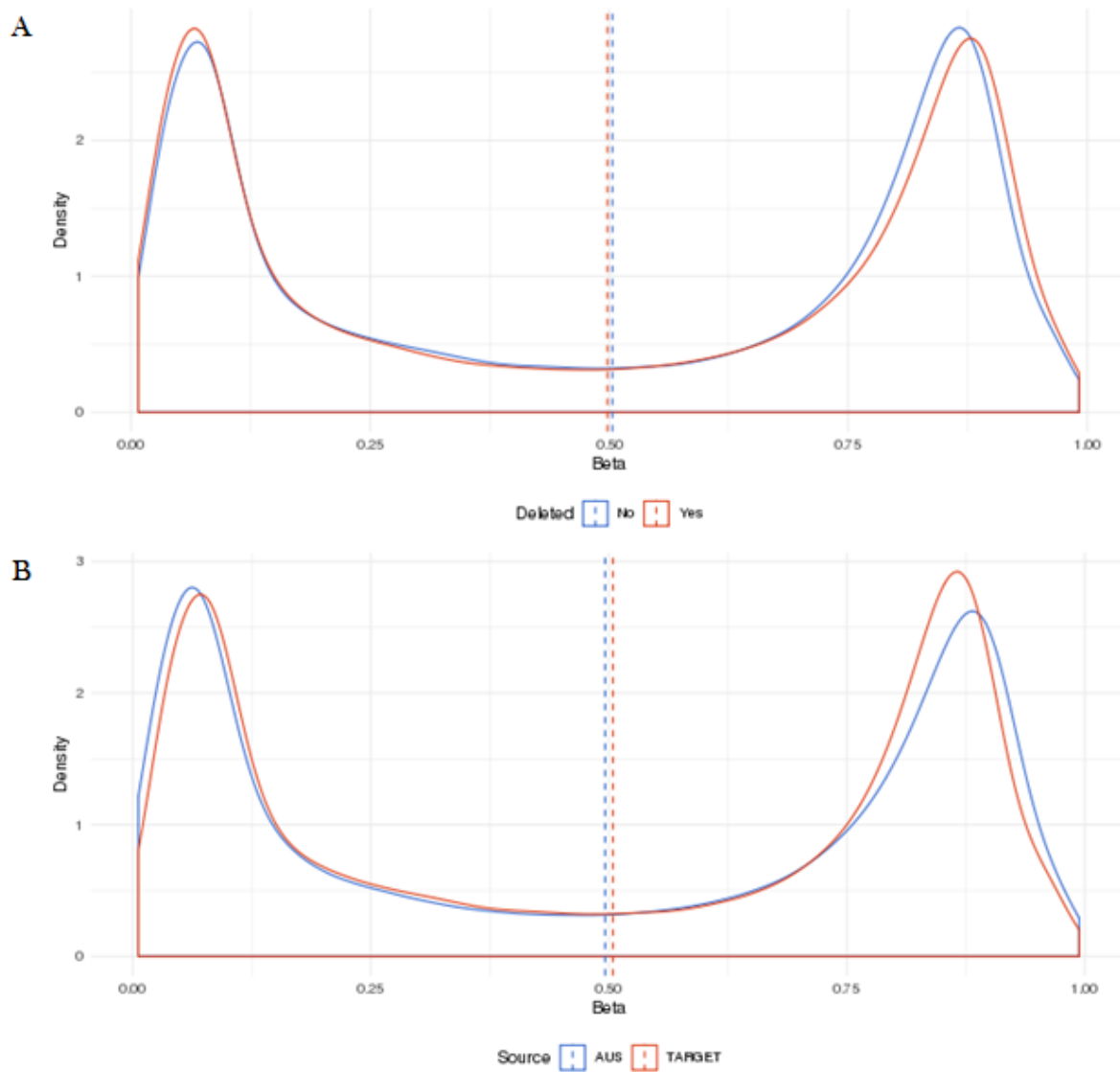


Figure 3.13 Inspection of global methylation differences between patients with and without allelic deletion and differences in patient cohort

Beta values from all 373,253 probes were averaged between A) cases with (n=71) and without allelic deletion (n=100), and B) between patient cohorts (n = 171). Overlaid density plots were used to define the global methylation landscape for each analysis. Distribution and density of beta values remain similar between both analyses, however a hypermethylated profile was found in samples A) without allelic deletion and B) TARGET cohorts when compared to their respective counterparts ($p < 1.0 \times 10^{-6}$). This was primarily driven by changes at the hypermethylated end of the spectrum.

3.5 Discussion

In this study we interrogated the global methylation status of over 200 individual paediatric AML patient bone marrow samples using two widely accepted ‘proxy’ measures. Two collections of patient samples were used; one from Australia and the other from the USA. The primary hypothesis tested was that there would be significantly aberrant total 5mC DNAm levels in diagnostic compared to remission and healthy samples. Our secondary hypotheses were that different subtypes of AML would display unique total 5mC levels and finally, that patients with a potential DNAm associated loss of heterozygosity would also have aberrantly methylated total 5mC levels compared to those that did not. To test this, we employed two widely used proxies of total 5mC. An initial GWAM value was derived using HM450K and density analyses. These total 5mC levels were then validated using MALDI-TOF MS of the repeat regions *LINE1* and *Alu*. Using the HM450k platform we were also able to ascertain copy number variation data that, in combination with regression analyses, were able to identify patients with both a deleted allele and hypermethylated promoter/enhancer region. Being able to detect these differences allowed us to identify the major trends in the total 5mC levels, and is the first of its kind in paediatric AML.

3.5.1 Interrogation of Global DNA Methylation Reveals Differences between Diagnostic and Remission AML Bone Marrow Samples

Multiple studies that have identified aberrant global DNAm levels in cancerous tissue compared to normal tissues (Gama-Sosa, Slagel et al. 1983; Jones and Baylin 2007; Portela and Esteller 2010; Sharma, Kelly et al. 2010). These studies found that global DNAm is generally decreased in cancer, however the opposite effect has been reported in adult AML and MDS (Toyota, Kopecky et al. 2001; Boulton and Wainscoat 2007; Zhang, Yuan et al. 2016). In this study we found that this trend of global hypermethylation in AML is consistent within an Australian paediatric cohort, when comparing diagnostic to matched remission bone marrow samples. A 3.7% increase in average global methylation was observed between the two groups, primarily driven by hypermethylation of diagnostic samples and widespread hypomethylation in remission samples. Pooled-analysis by addition of 109 samples from the

TARGET cohort resulted in similar findings, albeit with a smaller total $\Delta\beta$ (2.42% difference). This change is primarily attributed to the slightly more hypermethylated remission samples in the TARGET cohort compared to the Australian cohort. The difference may be due to Australian remission samples being derived from day 30 bone marrow post-induction therapy while the TARGET cohort was taken at day 14 during induction therapy. Therefore, it is possible that the reduction of average total methylation occurs from diagnosis and throughout induction therapy. This suggests that not only do we find a hypermethylated GWAM upon diagnosis of paediatric AML, but treatment is altering GWAM levels over time. All remission samples were verified to have <5% blast count, so residual disease from refractory samples is unlikely. This is further supported by ‘normal’, control bone marrow having a more similar profile to diagnostic samples than to remission samples, suggesting induction therapy may be generally hypomethylating the genome, requiring time before returning to a more normal state.

As evidenced by the regional breakdown of probes on the HM450k according to genomic location relative to genes, average global methylation is hypermethylated in diagnostic samples across all regions. This suggests that the effects witnessed are not solely found in promoter regions. Interestingly, we found that the largest regional gain of global methylation was in intergenic regions, which may be due to the higher density of enhancer or regulatory elements within these regions (Pennacchio, Bickmore et al. 2013). While these total diagnostic and regional hypermethylation findings are in line with previous studies in adult AML (Deneberg, Grövdal et al. 2010; Hopp, Lübbert et al. 2017), this is the first time it has been shown in not only one, but two paediatric AML cohorts. Further studies, preferably with additional time points post-diagnosis, will need to be completed to elucidate the longitudinal trend of aberrant global methylation as the patient returns to a truly healthy state.

While global methylation can be identified through genome-wide analyses, using proxies of DNAm is far more convenient for reasons regarding scale and ease of access (Vryer and Saffery 2017). *LINE1* and *Alu* are repetitive elements that together cover approximately 28% of the entire human genome (Cordaux and Batzer 2009) and have previously been shown to be accurate identifiers of global methylation in patient samples. In this study we found that *LINE1* and *Alu* both had positive linear relationships with HM450k derived global methylation as well as with imputed *Alu* methylation derived from HM450k data in combination with the REMP

package. We were also able to further validate our findings that diagnostic samples are hypermethylated compared to remission samples; the first time this has been shown in the paediatric AML setting. While not every individual site was statistically significantly different, the average methylation of both genes across all sites was significant. This validation of global methylation difference in both replicated and new samples suggests that the hypermethylated global signature found in diagnostic samples is valid finding. Previous studies have identified the global re-methylation of the genome from myeloma to leukaemia effects pathways involved in cell-cell signalling and cell adhesion (Walker, Wardell et al. 2011).

Patients that had matched relapse samples available showed that there was little difference in *LINE1* and *A/u* methylation levels, suggesting that the global methylation signature returns to its prior state after recurrence of AML. Considering the dynamic nature of DNAm in *LINE1* and *A/u* in response to disease progression, this opens the interesting possibility of using these repeat elements longitudinally to reflect the global DNAm landscape, and correlate their methylation with subsequent disease progression. Using DNAm proxies to observe disease progression and recurrence would allow effective monitoring and better decision making regarding patient treatment in a personalised manner. However, this is complicated by the potential of the ‘normal’ global epigenome being more similar to the diagnostic signature than the remission and would require further studies to truly elucidate the methylation signature in normal paediatric bone marrow.

The cumulative findings of GWAM from the HM450K and *LINE1* and *A/u* further solidify the gain of methylation found across all diagnostic samples. It shows that regional hypermethylation is not restricted to regions that are directly influencing gene expression but also repeat elements that comprise a much larger proportion of the genome. Considering the low mutation burden of paediatric AML and that progression of adult AML is intertwined with aberrant epigenetic mechanisms, one could speculate that there may be influences other than mutation acting upon epigenetic machinery that is causing these results in paediatric AML. The likely scenario of a total global hypermethylation event by disruption (via mutation that is not commonly found in AML or by targeted epigenetic repression) of the *TET2* protein or recruiters/cooperative factors of *TET* such as *Lin28A* (Zeng, Yao et al. 2016) or *IDH* (Yang, Ye et al. 2012), would disturb demethylation maintenance while simultaneously effecting self-

renewal capacity of haematopoietic stem cells (Hu, Chen et al. 2015). Furthermore, studies in adult AML have shown that *TET2* and *IDH* mutations occur frequently, are major contributors to the phenotype and are integral to adult AML progression (Im, Sehgal et al. 2014).

To summarise, this study has identified that there is global hypermethylation in diagnostic paediatric AML bone marrow samples when compared to matched, remission samples, and is further validated using the DNAm proxies *LINE1* and *Alu*. This confirms our hypothesis that paediatric AML is in line with previous adult leukaemic literature regarding global methylation status. This aberrant signature is indistinguishable from matched relapse samples but varies depending on how far into treatment the remission sample was taken. This indicates that the change in total 5mC begins with the original leukaemogenesis and decreases in response to induction therapy. This dynamic nature of total 5mC suggests potential for *LINE1* and *Alu* to be used as a biomarker of disease progression and recurrence.

3.5.2 Varying Chromosomal Aberrations Reveal Unique Global Methylation Profiles

Acute myeloid leukaemia has several subtypes, primarily driven by differing mutations and chromosomal aberrations (Figuerola, Lugthart et al. 2010; Taskesen, Babaei et al. 2015). This heterogeneity has been linked to subtype specific epigenetic alterations that have been shown to influence disease progression and severity in adult leukaemia (Larmonie, van den Heuvel-Eibrink et al. 2014). Whilst few studies have been conducted in the paediatric setting, one has shown that subtype-specific DNAm alterations are also apparent in some paediatric AML subtypes (Morenos 2014). With this in mind, we interrogated the global methylation level of various subtypes to test whether any are associated with altered 5mC levels. We found that global methylation did vary between the most common subtypes, with average β values ranging from 48.9-50.4% (maximum $\Delta\beta=1.5\%$). While global methylation averages do not tell the whole story, seeing that each subtype has their own distribution of methylation across all probes suggests that the aberrant methylation found is certainly subtype specific and each subtype may potentially have distinct mechanisms driving DNAm changes.

Interestingly, the *KMT2A* rearranged group of patients, which have a poorer prognosis (Balgobind, Raimondi et al. 2009), had the lowest total 5mC of all subtypes despite previous studies suggesting that patients with lower levels of global methylation tend to have a higher complete remission rate (Deneberg, Grövdal et al. 2010). Considering the *KMT2A* subtype has one of the worst rates of overall survival, while inv(16) and trisomy 8 have some of the best (Dohner, Estey et al. 2010), our data suggests that global methylation alone may not be a useful predictor of clinical outcome in paediatric AML. However, with an increase in sample size, there is potential for global methylation to be used to identify subtypes of AML that could otherwise remain undiagnosed by conventional methods. This is similar to studies conducted by Nordlund *et al.* or Bolouri *et al.*, where subtype-specific methylation signatures identified common subtypes and classified, as well as informed treatment for, previously cytogenetically unknown cases (Nordlund, Bäcklin et al. 2013; Bolouri, Farrar et al. 2018).

In this study we aimed to investigate if there was a difference in total 5mC of patients that could potentially have DNAm associated LOH compared to those that did not. Our study identified 71 patients with a chromosomal deletion and at least one gene in the corresponding region that was hypermethylated ($\beta > 75\%$). Many of the deleted gene regions identified by GISTIC analysis, including *CD36*, *CDC73*, *DTX4*, *MED19* and *STAT1*, are involved in cancer or differentiation pathways (Perea, Domingo et al. 2005; Shao, Liu et al. 2016). All genes identified pose interesting questions about how they may function in disease progression and need to be further investigated in paediatric AML specific functional models. Subsequently, this also leads to future hypotheses investigating demethylating agents and their potential to rescue patients from their potential LOH.

In summary, this study has identified unique global methylation signatures associated with patient subtype. Furthermore, this study identified a hypomethylated global signature associated with patients with potential loss of heterozygosity, and a hypermethylated signature associated with poor patient outcome. However, considering the minimal changes identified and that previously associated findings have opposite associations with global methylation this does bring into question if there is truly any prognostic and functional significance associated with total 5mC and patient LOH. Despite this, global methylation does serve as a useful tool for identifying patients that have a region specific deletion that may otherwise go undetected by

traditional diagnostic methods. Further studies need to be conducted to define what the exact changes in methylation are that may be able to delineate these groups. Finally, our data, while limited, supports our hypothesis of global hypomethylation being associated with LOH in paediatric AML, however the functional effects of this finding are yet to be elucidated.

3.5.3 Limitations & Future Directions

The primary limitation of this study centres on the relatively limited sample size, particularly in analyses stratifying samples into subgroups. Global methylation analysis using the HM450K is useful in that it can discern trends in the data and can often predict direction of effect in validation analyses, however each individual sample and subgroups are different as evidenced by the various signatures associated with patient subtype. As a result, global methylation analyses can lead to different results depending on how one stratifies groups, platform used, potential technical confounding factors, and what samples make up the initial cohort. Global methylation still serves as valid ‘first-look’ approach to genome-wide methylation data analysis. However, it should only be used, at best, as an associative marker of disease outcome. Further investigation into causes of genome-wide change in methylation should be examined upon discovery of differential total 5mC between groups. In future it would be productive to utilise a larger cohort and the gold standard of total 5mC interrogation in whole genome bisulphite sequencing (WGBS), which allows both total 5mC and distribution to be obtained. When combined with proxy measures of DNA, this approach provides a more holistic view of total and regional 5mC that are subject to gain or loss of methylation over multiple time points.

Lacking a true ‘normal’ control to compare global methylation data is also a limitation in interpretation of total 5mC change. While using the cell-sorted publicly available bone marrow methylation data from GSE63409 serves as a comparison, it is not a perfect control as it was derived from adults, cell-sorted and minimal in numbers. Therefore, we also used whole bone marrow at remission as a non-cancerous control. With methylation being dynamic in response to environment, age becomes a major focal point as it is unknown whether the difference in methylation profile of the adult controls to paediatric diagnostic and remission is due to age rather than disease status. Furthermore, without a more refined control cohort, delineation of subtype specific total 5mC stops at intra-cohort differences and disallows comparison of

subtype to normal methylation signatures to find what subtypes have a generally more aberrant total 5mC profile. While it is highly unlikely that the perfect control sample of bone marrow prior to diagnosis of AML will ever be available for research, it may be possible to use healthy donor bone marrow from children in future studies.

Proxies of global DNAm were utilised as an alternative to using genome-wide data to infer disease state and potentially by-pass the limitation of scale that genome-wide approaches impose. However, using *LINE1* and *Alu* as proxies of global DNAm have their own set of limitations. As shown in Figure 3.8, there was a positive correlation between global and average *LINE1* and *Alu* methylation, however *Alu* was identified as non-significant. Utilising a larger cohort, preferably with cell-sorted samples, will help elucidate whether sample time point or cell constitution has any effect upon global DNAm proxy results. With respect to the cohorts used in this study, we did not have any samples that had data available across all diagnostic, remission and relapse sample types. A sample set of this nature would be required to determine the practicality of using *LINE1* and *Alu* as a marker of disease over time, requiring demonstration of global methylation being dynamic in response to a patient going into remission and having their disease recur. As it stands, we can only suggest that *LINE1* and *Alu* methylation lowers upon remission and, considering the minimal differences between diagnostic and relapse samples, returns to the original diagnostic state upon recurrence. This, unfortunately, does not rule out the chances that the identified effect is driven by the different samples used in both analyses. Using a greater number of remission time points (day 14, day 30, 3 months, etc.) and a continuum of time points for multiple patients across all diagnostic, remission and relapse samples would determine the true potential of using *LINE1* and *Alu* as a dynamic biomarker of both disease progression and potential recurrence.

The final limitation to this study refers to GISTIC analysis. While this analysis is an acceptable approach for identifying frequently deleted genomic regions in a cohort, the binning mechanism used to create the initial input groups had large sections of the genome together, encompassing multiple genes. While parts of these areas may be deleted, the genes identified by the analysis can only be truly proven to be deleted using more conventional comparative genome hybridisation approaches. Subsequently, genes identified to be hypermethylated and deleted may not be deleted at all, and a true patient LOH can't be determined without further

testing. Therefore, the resulting future direction for this type of analysis would be to accurately determine the break points, and by extension, the specific genes deleted, of each patient identified to have a deletion and subsequently identifying those genes to have an aberrant methylation profile. This will allow us to determine patients with a truly DNAm induced LOH and further refine analyses pertaining to LOH and global methylation associations.

This page is intentionally left blank.

Chapter 4 DNA Methylation Profiling of Paediatric Acute Myeloid Leukaemia

4.1 Rationale

Acute myeloid leukaemia, like all cancers, is universally associated with aberrant DNAm. For some time, studies in adult AML have shown that targeted changes of DNAm occur in key tumour-suppressor genes such as *P15/16/73*, *MGMT*, *RB*, *H1C1*, *CDH1* and *SOCS1* (Melki, Vincent et al. 1999; Galm, Wilop et al. 2005). Furthermore, mutations in epigenetic modifiers such as *TET2*, *IDH1/2*, *ASXL1* and *DNMT3A* are common in adult AML (Abdel-Wahab and Levine 2013). This implicates DNAm as a functional factor behind disease progression rather than a consequence of genomic aberrations. However, while this is true for adult AML, the DNAm landscape of paediatric AML has not been widely explored. A few studies have looked at this, but data on the frequency of epigenetic modifier gene mutations are almost non-existent (Ho, Kutny et al. 2011; Liang, Liu et al. 2013). In contrast, the frequency of other common mutations in childhood relative to adult AML is becoming clear (Creutzig, Zimmermann et al. 2016). Despite the lack of knowledge on epigenetic modifier mutations, there is emerging evidence of differential methylation in the paediatric setting (Morenos 2014; Larmonie, Arentsen-Peters et al. 2017; Stieglitz, Mazor et al. 2017). However, these studies are largely limited to subtype specific analyses and do not explore changes that may be ubiquitous across all diagnoses. Therefore, knowing that aberrant DNAm is a contributing factor in paediatric AML it is necessary to investigate and understand the specifically targeted DNAm changes that occur during initiation and progression of all forms of AML.

A major study dissecting DNAm in the paediatric setting was conducted by Bolouri *et al.* (2017). This focused broadly on paediatric AML, further verifying that mutations in epigenetic modifying genes are rare and that distinct methylation signatures exist between diagnostic and remission bone marrow samples (Bolouri, Farrar et al. 2018). Signatures of DNAm were observed through non-negative matrix factorisation that associated variable methylation levels with patient subtype and age. Bolouri *et al.* identified 119 pan-leukaemic genes recurrently silenced by hypermethylation including *ULBP1/2/3*, *HES1/4/6/7*, *MGMT*, *SMAD1/5*, *SOCS2/6*, and *TET1/2*. These genes are commonly involved in Wnt- β -catenin signalling and are active ligands in NK cells, suggesting that they might be worth targeting therapeutically. Furthermore, subtype specific genes were also identified, particularly a group of relatively

unknown zinc finger-encoding genes, associated with the *KM2TA* rearranged subtype. Larmonie *et al.* have shown evidence that *MNI* and *TRIM71* are both hypermethylated in the inv(16) subtype of paediatric AML compared to other subtypes and are subsequently repressed (Larmonie, Arentsen-Peters *et al.* 2017). They also suggest that *MNI* is co-regulated by expression levels of *DNMT3B*, further solidifying the role of aberrant DNAm in disease progression. Morenos *et al.* found that in the M5 French-American-British (FAB) defined subtype that *DLEU2* and the encompassed tumour suppressor *miR-15a/16-1* were hypermethylated (Morenos 2014). Methylation and expression similar to that of remission samples at these loci were able to predict low-risk prognosis in particular trisomy 8 and *KM2TA* rearranged leukaemia.

The functional contribution of aberrant DNAm to disease progression is relatively unknown in paediatric AML, however aberrant methylation has been associated with the progression of adult myelodysplastic syndromes (MDS) to secondary AML. Jiang *et al.* have found that out of 1505 candidate tumour-suppressor genes (TSG) 91 were silenced in MDS and this increased up to 179 upon leukaemic blast transformation. *FZD9* was identified as a potential candidate TSG as this was one of the most frequently silenced genes and was also, when deleted and hypermethylated, associated with poor clinical outcome (Jiang, Dunbar *et al.* 2009). Furthermore, MDS and secondary AML were identified to have a more extensively aberrant promoter DNAm profile compared to CD34+ cells (Figueroa, Skrabanek *et al.* 2009). This was particularly apparent in *WNT* and *MAPK* signalling pathways. Given that DNAm changes in response to disease progression from MDS to secondary AML, it is plausible that DNAm may also change from AML diagnosis to remission and again upon recurrence.

In this study we have used both the Australian and TARGET AML cohorts in combination with linear regression modelling to identify signatures associated with AML, AML subtype and disease progression to relapse. We identified specific differentially methylated probes (DMPs) and regions, (DMRs), including gene-based markers of differential methylation, between disease states and time points. The overarching aim of this study was to define targeted DNAm differences that could be contributing to disease in a pan-leukaemic manner across the entire paediatric AML population.

We have hypothesized that;

- (i) There is a distinct DNAm profile that can accurately define diagnostic, remission and healthy bone marrow, that is independent of patient subtype;
- (ii) Differential DNAm can be identified in specific genes that may be contributing to initiation and progression of disease, and;
- (iii) Subtle aberrant DNAm can be identified between diagnostic and matched relapse AML samples.

4.2 Study Cohort

For this study we utilised the entire Australian and TARGET AML cohorts. A breakdown of these samples can be found in Table 2.4. Overall, this included 198 diagnostic, 143 matched remission and 76 matched relapsed samples across two major cohorts and publicly available DNAm data from 24 healthy, cell sorted populations derived from 5 adult donor bone marrows. In total the number of unique samples available was 231. Diagnostic bone marrow samples were taken on the day of diagnosis and tested for cytogenetic abnormalities. Of remission samples, 26 did not have a matched diagnostic sample. The most common AML-associated chromosomal rearrangements are well represented in this cohort. Patients were between 3-15 years of age with males slightly over-represented.

Patient clinical information was obtained from a local hospital database (CLARA) for Victorian cases at the RCH. For other cases cohort information was obtained from respective biobanking repositories. All data is correct as of April 2018.

4.3 Genome-wide DNA Methylation of Diagnostic, Remission & Healthy Samples - Australian Cohort

4.3.1 Categorisation of Samples According to Disease State

Unsupervised analyses by principal components analysis (PCA) was undertaken on all diagnostic, remission and healthy bone marrow samples (n =131). Principal component 1 (PC1), accounting for 13.1% of the variability found in the dataset, is primarily attributed to the time of sample taken between diagnostic and remission samples (Figure 4.1, $p = 2.0 \times 10^{-7}$, Student's asymptotic p-value for correlation). However, it was also influenced by common confounding factors such as HM450K chip, age, study site and patient outcome (Appendix II). These factors clearly contribute to the overall variation in the dataset. Variation within PC2 was also associated with patient time point, this time separating diagnostic/remission samples from healthy bone marrow, which may be due to true biological variation, age of sample or cohort (Figure 4.1). Overall dimensionality reduction (the process of reducing the amount of random dimensions within the data to a principal few to best encompass total variation) utilising DNAm data shows that majority of diagnostic AML bone marrow samples can be distinguished from matched remission samples (PC1) and from healthy adult bone marrow (PC2).

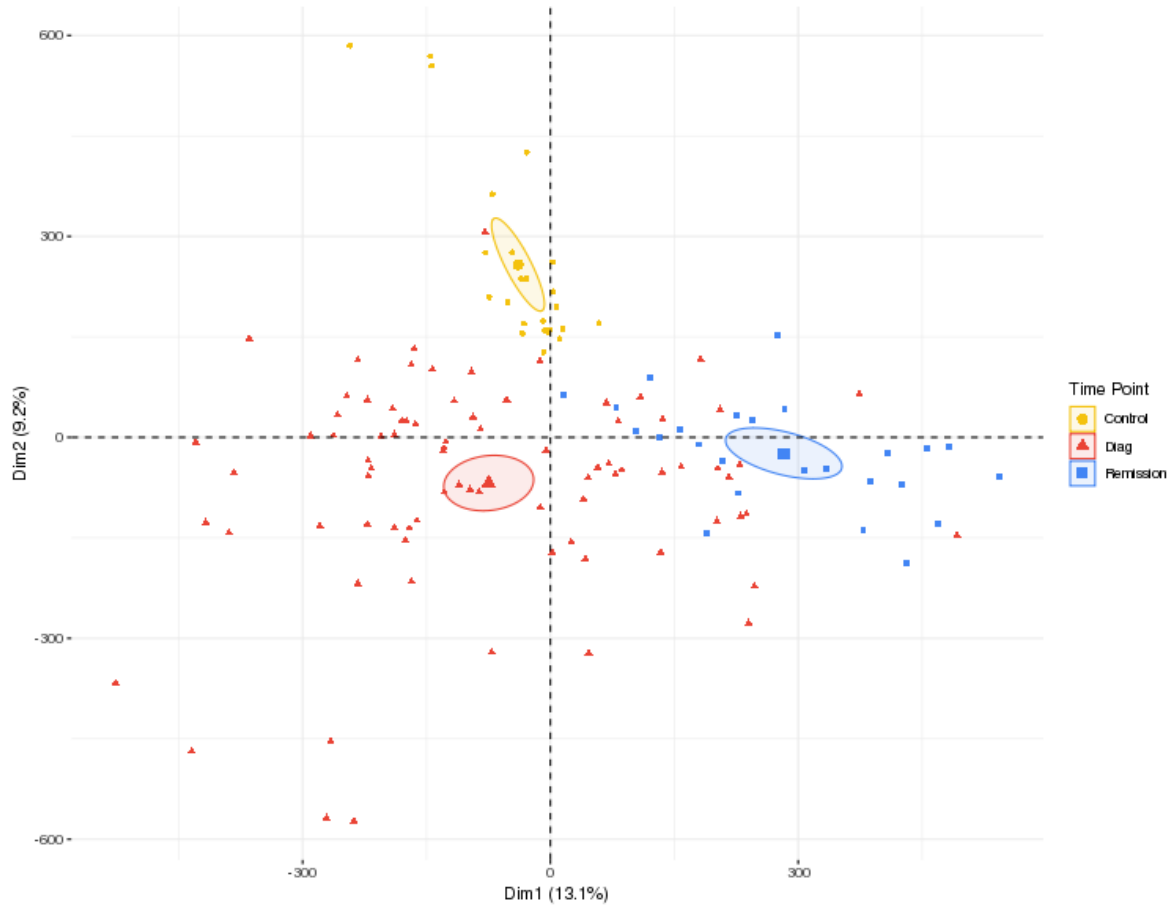


Figure 4.1 Unsupervised principal components analysis plot defining diagnostic from remission and healthy adult bone marrow samples

Beta values from all 373,253 remaining probes from the HM450k across 131 samples were used in a PCA to determine sources of variation in the dataset. The variation in PC1 (accounting for 13.1% total variation) can be partially attributed to differences in sample time point but this does not explain the full source of variation as patients are still undefined on the x-axis. Variation in PC2 (accounting for 9.2% total variation) is defined by age of sample (adult vs paediatric) or by the varying cohorts used (Australian cohort vs. public data). The larger circle, square and triangle represent their respective group means and ellipses drawn around these averages are defined by 0.95 confidence interval. Diagnostic samples are in red, remission samples in blue and control samples in yellow.

In order to further identify specific methylation variation associated with disease state as suggested by the unsupervised analysis, least squares regression analysis was carried out using RUVm (Section 2.5.6). This revealed 4219 DMPs following correction for FDR (Benjamini-Hochberg adjusted $p < 0.05$) between diagnostic and healthy bone marrow samples (remission and control bone marrow samples combined). Figure 4.2 shows that methylation levels at these DMPs delineated diagnostic from non-AML bone marrow samples. Furthermore, this signature also distinguished remission samples from healthy, control bone marrow suggesting

that while remission bone marrow is more closely related to healthy compared to diagnostic samples, minor differences exist between these time points. Considering the close relatedness of remission and healthy samples compared to their diagnostic counterparts, this suggests that the cancerous cell composition (blast population in diagnostic samples), as opposed to differences in cohort or age is primarily driving the differential methylation profile discovered.

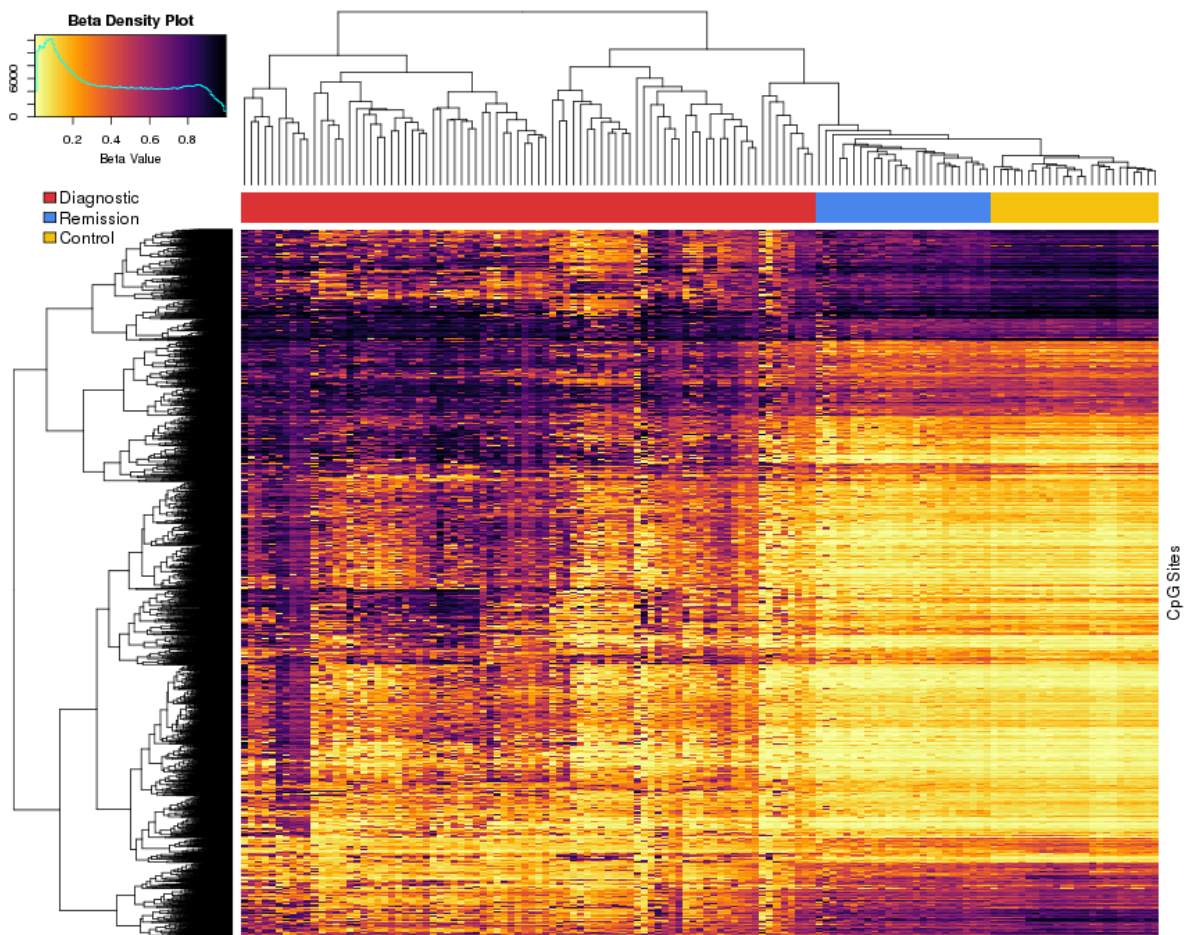


Figure 4.2 Heatmap of methylation levels across the 4219 highest ranked DMPs associated with AML status

Heatmap with the y-axis representing 4219 differentially methylated CpG sites while the x-axis shows 131 patient samples where both are ordered by hierarchical clustering as depicted by the dendrograms. This heatmap shows that DNA methylation can accurately distinguish diagnostic AML bone marrow from remission and healthy adult bone marrow samples.

4.3.2 Genes Differentially Methylated in Diagnostic AML Samples

After identification of a DNAm signature associated with paediatric AML, we sought to identify specific genes that may provide clues as to the pathogenesis of AML through differences between diagnostic and non-AML bone marrow. Figure 4.3 shows a Manhattan plot of all 373,253 probes and their $-\log_{10}$ transformed p-value after regression analysis. We identified several genes showing multiple instances of differential methylation. Several of these were prioritised according to effect size, number of probes and location within gene regulatory regions and mapped to the genes Wilms Tumour 1 (*WT1*), Diacylglycerol Kinase Gamma (*DGKG*), Zinc Finger and SCAN Domain Containing 1 (*ZSCAN1*) and Zinc Finger Protein 876, Pseudogene (*ZNF876P*). A list of the top 100 differentially methylated probes are detailed in Table 4.1.

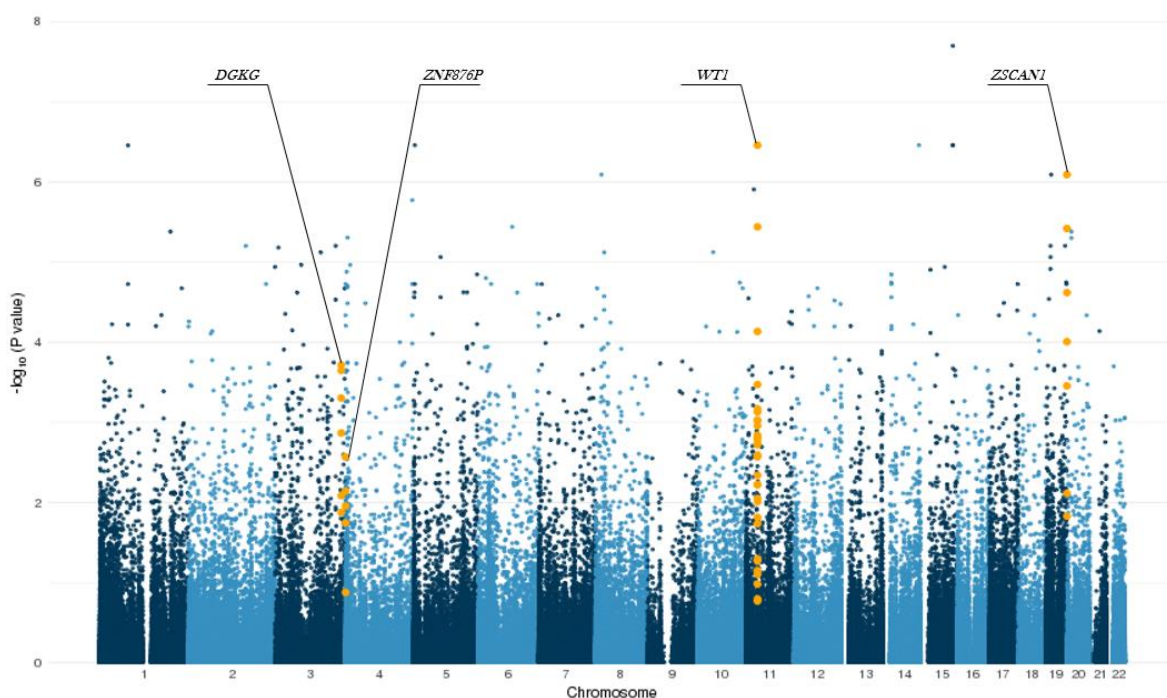


Figure 4.3 Manhattan plot showing significance of all probes after least squares regression analysis between diagnostic and control bone marrow samples

Manhattan plot details the significance of each probe after differential methylation analysis (n=131 samples). Significance is shown on the y-axis and has been $-\log_{10}$ transformed. Probes are separated on the x-axis by chromosomal location. Highlighted probes pertain to the significantly differentially methylated probes in the genes of *WT1* (chr11), *ZNF876P* (chr4), *ZSCAN1* (chr19) and *DGKG* (chr3). These genes were selected for further validation.

A probe in the *WT1* gene, cg09234616, was identified as the second most significantly differentially methylated probe (average $\Delta\beta$ 0.57, Benjamini-Hochberg adjusted $p=3.47 \times 10^{-7}$). *WT1* is a gene of interest due to the already established pathogenic contribution to all forms of AML, via both mutational status and its known function in the maintenance of DNAm (Gaidzik, Schlenk et al. 2009; Ho, Zeng et al. 2010). An additional 23 DMPs were also identified in *WT1* (Figure 4.4A, Benjamini-Hochberg adjusted $p<0.05$) with a total average $\Delta\beta$ of 39.1%. Many of these probes have previously been identified to be cancer associated DMPs in accordance with research conducted by Doi *et al* (Doi, Park et al. 2009); while three probes were located within a known enhancer region in the gene.

We also identified six adjacent DMPs in *DGKG* that were significantly differentially methylated between diagnostic and non-AML bone marrow samples (Benjamini-Hochberg adjusted $p<0.05$). An average $\Delta\beta$ of 54.8% across all significant probes was found and 5/6 of these probes were within a known enhancer and DNase hypersensitivity (DHS) region as defined by the ENCODE study in over 125 cell types including hESC and myeloid cell lines (Figure 4.4B, (Consortium 2012)). *DGKG* has been identified as a gene of interest due to its involvement in cell cycle regulation and macrophage differentiation, as well as previous studies having found hypermethylation in acute lymphoblastic leukaemia (Figueroa, Chen et al. 2013).

Another probe that was identified in the top 10 differentially methylated probes was located within *ZSCAN1* ($\Delta\beta$ 49.5%, Benjamini-Hochberg adjusted $p=8.07 \times 10^{-7}$). We also identified a total of seven adjacent probes within *ZSCAN1* forming a DMR with an average $\Delta\beta$ of 43.2% (Figure 4.4C, Benjamini-Hochberg adjusted $p<0.05$). Furthermore, this DMR was also found to be within promoter region and 5'UTR of the gene that has also been identified to be in a region of DHS which is associated with open chromatin according to ENCODE (Consortium 2012). Probes in *ZSCAN1* has previously been identified as a cancer-related DMR (Doi, Park et al. 2009) and believed to be involved in transcription regulation.

Finally, *ZNF876P* had four probes that were identified as DMPs (Figure 4.4D, Benjamini-Hochberg adjusted $p<0.05$). Probes identified had an average $\Delta\beta$ of 45.6%. These probes were found in all parts of the gene including the promoter region, 5'UTR and introns, all of which have been previously shown to be in regions of DNase hypersensitivity (Consortium 2012).

Besides the large disparity in DNAm between diagnostic and healthy bone marrow samples, *ZNF876P* was selected as it may play a role in transcriptional activity.

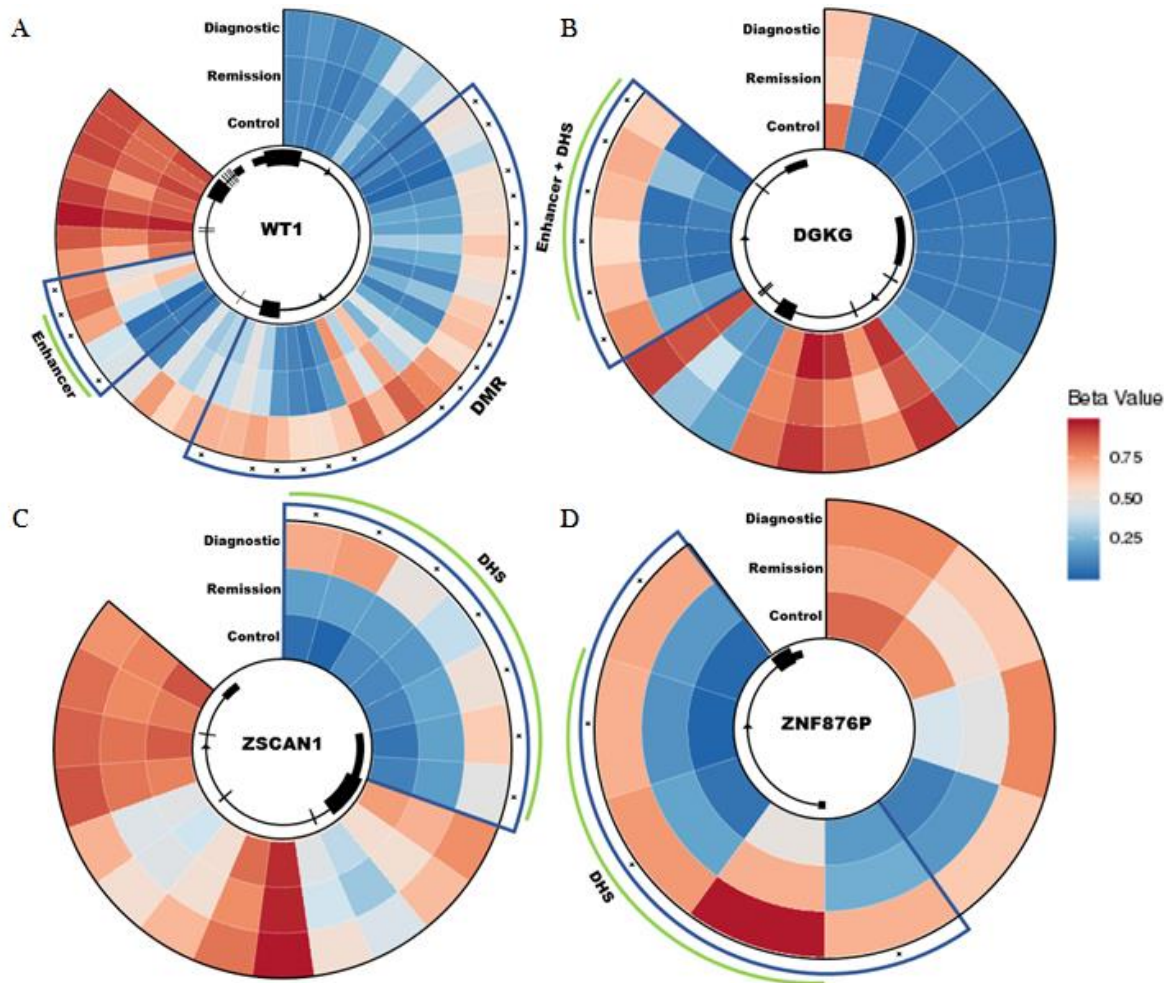


Figure 4.4 Circular heatmaps showing DNA methylation levels within specific probes for diagnostic, remission and control samples

Circular heatmaps shows DNA methylation at each probe that is found in the gene or within 1500 base pairs of the associated untranslated regions (n=131 samples). Inner circles identify the geographical position of any given probe in the gene (not to scale). Methylation levels are depicted by a spectrum of colour, with completely unmethylated probes dark blue while complete methylation is represented by dark red. Significantly differentially methylated probes are indicated by asterisk ($p < 0.05$, Student's t-test). Differentially methylated regions (DMRs), are indicated by blue wedges. Enhancer regions or regions of DNase hypersensitivity found within DMRs are indicated by the outer green line as defined by ENCODE and Doi *et al.* (Doi, Park et al. 2009; Consortium 2012). A) Methylation levels at 57 *WT1* probes. 23 probes were identified to be significantly differentially methylated across two DMRs with an average $\Delta\beta$ of 39.1%. B) Methylation levels at 26 *DGKG* probes. Six probes were identified to be differentially methylated across one DMR with an average $\Delta\beta$ of 54.8%. C) Methylation levels at 20 *ZSCAN1* probes. Seven probes were identified to be significantly differentially methylated across one DMR with an average $\Delta\beta$ of 43.2%. D) Methylation levels at 9 *ZNF876P* probes. Four probes were identified to be differentially methylated across one DMR with an average $\Delta\beta$ of 45.6%.

Table 4.1 Top 100 probes as defined by least squares regression analysis comparing AML to non-AML bone marrow samples

Probe ID	Gene	Adjusted p-value*	$\Delta\beta$	Location	Probe ID	Gene	Adjusted p-value*	$\Delta\beta$	Location
cg20738500	LOC254559	1.99E-08	0.573	TSS200	cg23948811		1.88E-05	-0.506	
cg09234616	WT1	3.47E-07	0.565	Body	cg25325512	PIM1	1.88E-05	0.396	3'UTR
cg15084543	ELTD1	3.47E-07	0.680	5'UTR	cg06335867	NXPH1	1.88E-05	0.589	Body
cg02287710	DIO3	3.47E-07	0.579	TSS200	cg05542338	EVC2	1.97E-05	0.194	5'UTR
cg00481644	LOC254559	3.47E-07	0.534	TSS200	cg06426416	EVC2	2.07E-05	0.271	5'UTR
cg12598235	LOC254559	3.47E-07	0.593	TSS200	cg11527913	GALNT2	2.12E-05	0.477	Body
cg13729210		3.47E-07	0.326		cg09924848	FAM53B	2.12E-05	-0.377	3'UTR
cg12296772	MTMR7	8.07E-07	0.490	TSS200	cg08120226	MUC4	2.12E-05	0.379	Body
cg08173263	LPHN1	8.07E-07	-	Body	cg23378033	MSRB3	2.12E-05	0.434	TSS1500
			0.150						
cg10132208	ZSCAN1	8.07E-07	0.495	5'UTR	cg23495279	CSMD1	2.12E-05	0.262	Body
cg23181580		1.24E-06	0.544		cg07408835	PCDHAC1	2.38E-05	0.313	1stExon
cg09473510	MTNR1A	1.68E-06	0.490	TSS200	cg19014419	ZNF300	2.39E-05	0.421	TSS200
cg22533573	WT1	3.62E-06	0.391	Body	cg05903444	IRX2	2.39E-05	0.394	Body
cg01758512	FUT9	3.62E-06	0.378	5'UTR	cg24368848	ZSCAN1	2.40E-05	0.600	TSS1500
cg25537993	ZSCAN1	3.80E-06	0.630	TSS1500	cg17152981	GPR6	2.40E-05	0.353	TSS1500
cg07504763		4.15E-06	0.430		cg02221302	C3orf14	2.40E-05	0.302	TSS1500
cg04682802	BTBD3	4.15E-06	0.559	5'UTR	cg20585869	NEFM	2.66E-05	0.359	TSS200
cg03544320	CRMP1	4.93E-06	0.532	1stExon	cg14625113	CNTN1	2.66E-05	0.296	5'UTR
cg05822633	BTBD3	5.00E-06	0.582	5'UTR	cg07516252	REC8	2.74E-05	0.426	TSS200
cg05135521	RBMS1	6.26E-06	0.253	Body	cg14619949	F2RL1	2.74E-05	0.571	Body
cg03192598		6.26E-06	0.593		cg07565505		2.74E-05	0.488	
cg05937737	SLC7A14	6.26E-06	0.577	5'UTR	cg08185661	SYT9	2.83E-05	0.260	1stExon
cg11254700		6.26E-06	0.718		cg02961707	EVI5L	2.88E-05	0.407	Body
cg11084334	LHFPL4	6.57E-06	0.407	Body	cg07766263	SLC7A14	2.93E-05	0.446	5'UTR
cg03171770		7.52E-06	0.363		cg08658787		3.01E-05	0.521	
cg17078116	NEFM	7.52E-06	0.453	TSS200	cg15138339	COASY	3.23E-05	-0.196	1stExon
cg11083235		7.52E-06	0.224		cg07963234	CRMP1	3.26E-05	0.406	TSS200
cg00499700	F2RL1	8.64E-06	0.552	Body	cg25444386	KDR	3.26E-05	0.368	Body
cg17971015		8.64E-06	0.368		cg15928093	FZD10	3.33E-05	0.241	5'UTR
cg05502701	PDZRN3	1.08E-05	0.375	TSS200	cg14243778	CNTN1	3.94E-05	0.476	5'UTR
cg07333191		1.08E-05	0.497		cg26980244	NEFM	3.94E-05	0.419	5'UTR
cg17232357	SMAD6	1.15E-05	0.157	Body	cg00973876		4.01E-05	0.270	
cg10603275	CHL1	1.15E-05	0.528	5'UTR	cg19896824		4.13E-05	0.465	
cg22195627		1.22E-05	0.610		cg10312186		4.36E-05	-0.311	
cg05678749	GABRG3	1.24E-05	0.389	Body	cg08276289		4.41E-05	0.234	
cg15230030	RGS12	1.32E-05	-	Body	cg03817667	VSTM2A	4.59E-05	0.542	TSS200
			0.446						
cg17840719	ZNF454	1.42E-05	0.641	1stExon	cg20779964	CBLN4	4.59E-05	0.493	TSS200
cg18556834	REC8	1.42E-05	0.454	TSS200	cg26894523	ZNF718	4.59E-05	0.322	Body
cg07914200		1.58E-05	0.500		cg08404702	TSC2	4.59E-05	-0.335	Body
cg25510090	ZNF787	1.79E-05	0.171	5'UTR	cg20192747		4.59E-05	0.383	
cg16473141	REC8	1.80E-05	0.514	1stExon	cg08915603	MTNR1A	4.60E-05	0.461	TSS200
cg17453460		1.81E-05	0.422		cg12765382		4.61E-05	-0.409	
cg01294808	IRX1	1.88E-05	0.523	Body	cg01196858	SCRN1	5.09E-05	0.364	5'UTR
cg08146323	ZNF835	1.88E-05	0.287	Body	cg06942649		5.09E-05	-0.116	
cg06594404		1.88E-05	0.575		cg19517718	PXDN	5.51E-05	0.270	Body
cg23302682	T	1.88E-05	0.274	Body	cg11474701		5.61E-05	0.306	
cg03431846	REC8	1.88E-05	0.448	TSS200	cg10806146	SLC20A2	5.67E-05	0.550	5'UTR
cg21443659		1.88E-05	0.364		cg26731131		5.93E-05	0.291	
cg04360793	ELTD1	1.88E-05	0.621	5'UTR	cg02165355	ZNF454	5.94E-05	0.304	TSS200
cg26238041	MTNR1A	1.88E-05	0.369	TSS200	cg22859727	FLRT2	5.96E-05	0.442	3'UTR

*Adjusted p-value calculated by Benjamini-Hochberg method.

$\Delta\beta$ = Difference in methylation calculated as AML average minus non-AML average

The most significant individual probes identified in this regression analysis were attributed to *MIR9-3HG* (Benjamini-Hochberg adjusted $p=1.99 \times 10^{-8}$, Figure 4.5A), *WT1*, as previously mentioned (Figure 4.5B), *ELTD1* (Benjamini-Hochberg adjusted $p=3.47 \times 10^{-7}$, Figure 4.5C), *DIO3* (Benjamini-Hochberg adjusted $p=3.47 \times 10^{-7}$, Figure 4.5D), *MIR9-3HG* again (Benjamini-Hochberg adjusted $p=3.47 \times 10^{-7}$, Figure 4.5E/F), *MTMR7* (Benjamini-Hochberg adjusted $p=8.07 \times 10^{-7}$, Figure 4.5H) and *LPHN1* (Benjamini-Hochberg adjusted $p=8.07 \times 10^{-7}$, Figure 4.5I). All probes identified as the most significant DMPs have shown either previous pathogenic involvement in AML/other forms of cancer or are involved in pathways that, when disrupted, could plausibly contribute to a cancerous phenotype. However, thorough inflation of findings has been identified in this study ($\lambda=1.53$) suggesting that top probes and previous discussed DMRs need to be validated in an independent cohort, as discussed in Section 4.4.1.

Figure 4.5 and Table 4.1 shows that, of the top nine probes, eight were identified as hypermethylated in diagnostic samples which is as expected considering elevated global methylation findings (as discussed in Chapter 3). Remission and control bone marrow samples are separated to show that there is minimal difference between non-AML samples, and that the change in methylation is specifically associated with leukaemogenesis. However, while this does hold true for most DMPs, Figure 4.5I demonstrate that there are remission associated changes of DNAm. Furthermore, while the average differential methylation between diagnostic and healthy bone marrow samples does exceed 50%, Figure 4.5 does indicate that there is a high amount of variability between patient samples which may be due to patient subtype or cellular heterogeneity in diagnostic samples.

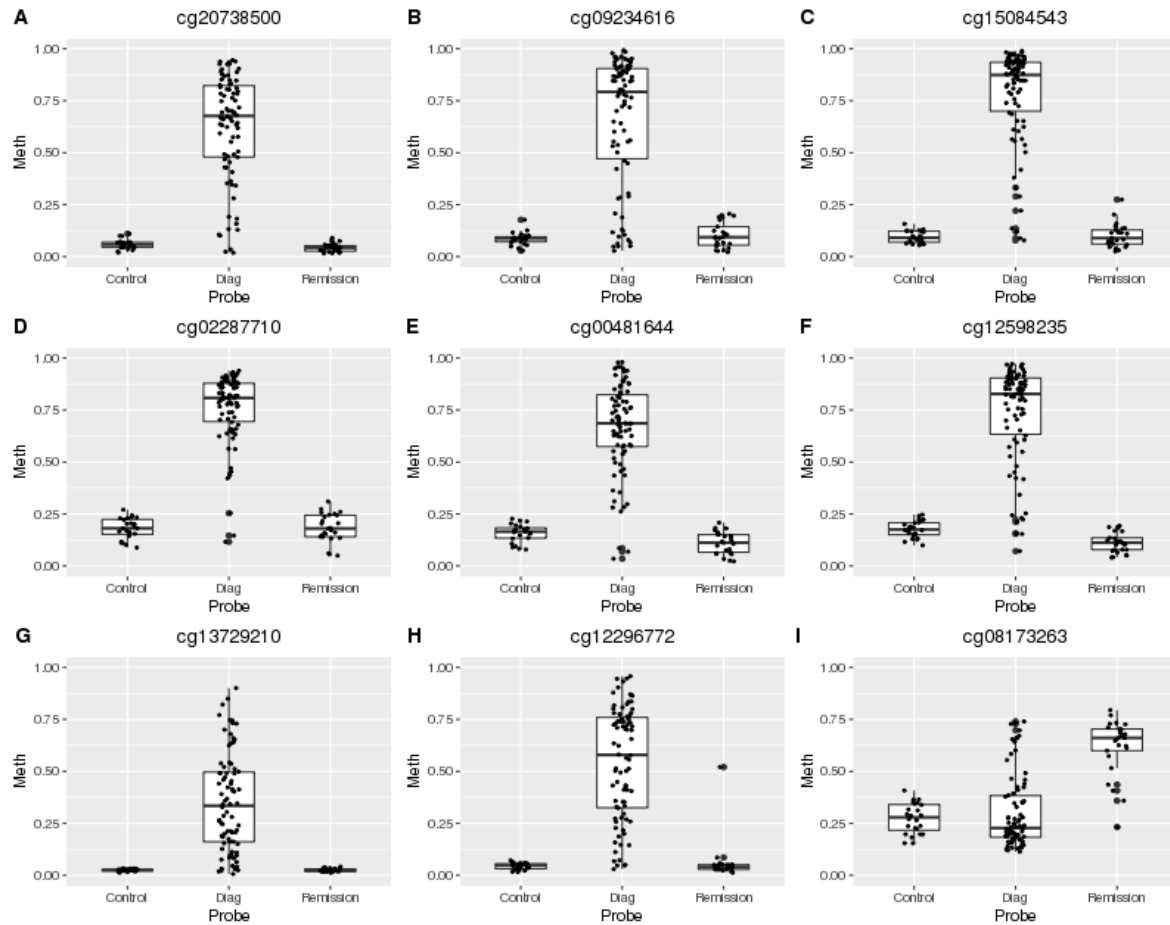


Figure 4.5 Boxplots of the most significantly differentially methylated probes between diagnostic and non-AML bone marrow samples

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between patient bone marrow biopsy time points (n=131). The median is represented by the line in the middle of the boxplot while outliers lay beyond the whiskers. Probes were identified as the most statistically significant DMPs (Benjamini-Hochberg adjusted $p < 8.07 \times 10^{-7}$). Probes identified can be attributed to the genes A) *MIR9-3HG*, B) *WT1*, C) *ELTD1*, D) *DIO3*, E/F) *MIR9-3HG*, G) Intergenic region, H) *MTMR7*, I) *LPHN1*.

4.3.3 Pathway Analysis

Pathway analysis (gene ontology) was undertaken on the genes associated with the top probes identified through regression analysis. Utilising the GOMeth package in R we found 617 gene ontology terms that were significantly associated with the 4219 significant probes. This was further refined using the Reduce + Visualise Gene Ontology (REVIGO) software which summarises and removes redundant GO terms. This analysis identified a total of 43 GO umbrella terms that was visualised using Cytoscape (Figure 4.6). Anatomical structural development, multicellular organismal process, developmental process, cell-cell signalling and signalling were the top 5 GO terms associated with these differentially methylated genes.

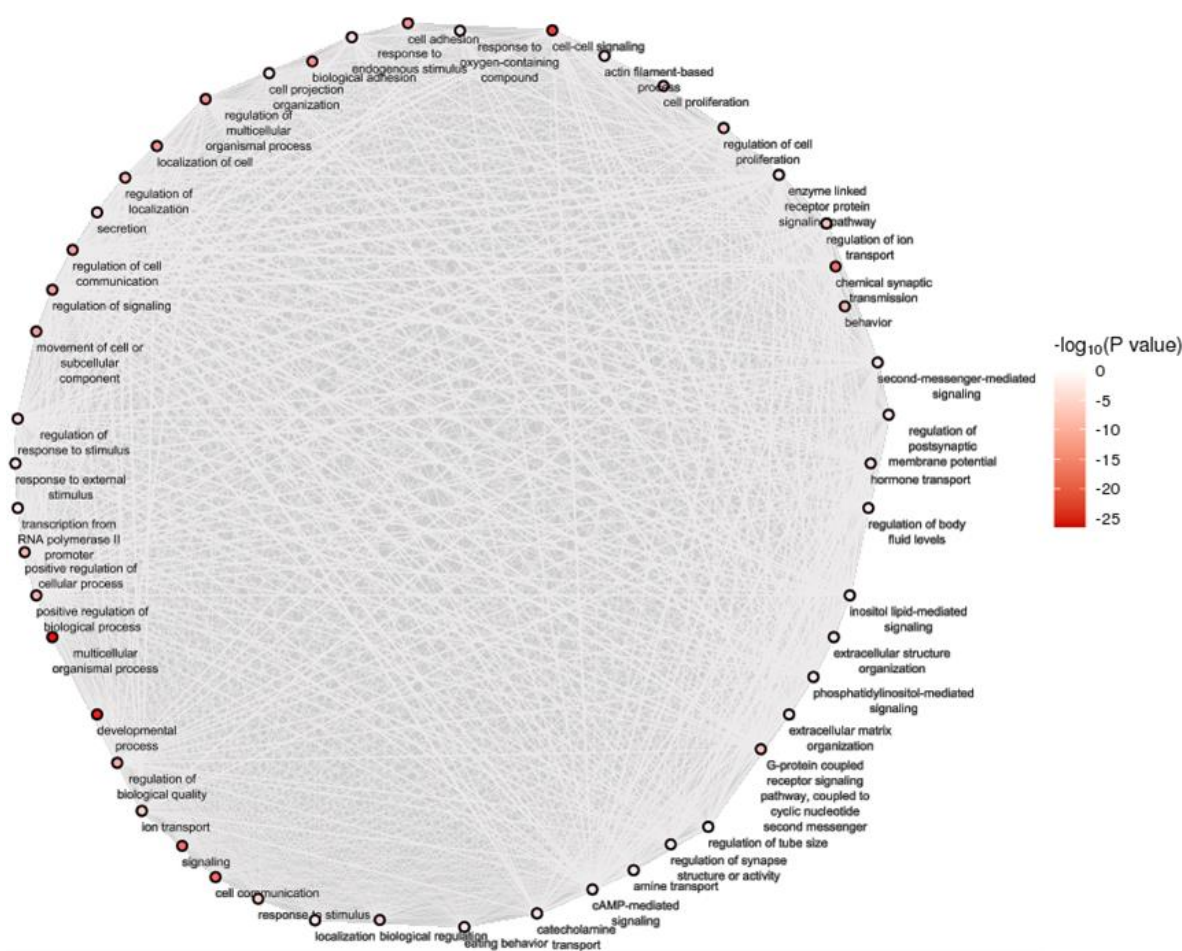


Figure 4.6 REVIGO gene ontology network associated with 4219 differentially methylated probes

Caption on next page.

REVIGO gene ontology reduction identified 43 different GO terms associated with the 4219 significantly differentially methylated probes derived from regression analysis. $\log_{10}$ p values are shown for each GO term to indicate the significance of the number of probes that are associated with that term appearing by chance.

4.4 Locus Specific Replication of DNA Methylation

4.4.1 Validation and Replication of Methylation between HM450K Array and SEQUENOM Assayed CpG Sites

Locus specific DNA methylation analysis was undertaken on a mixture of samples previously used in HM450K genome-wide analyses, in addition to new samples, to create both a validation and replication cohort. SEQUENOM amplification assays covering the genes *WT1*, *DGKG* and *ZNF876P* were designed to cover HM450K probes of interest, plus several other CpG sites that are not found on the HM450K platform, expanding the regional analysis of differential methylation.

WT1

Two assays were created covering the first intron of *WT1* but also covered the 5'UTR, transcription start site and gene body of an alternatively spliced version of the gene. Between these assays, four previously identified differentially methylated probes were encompassed including; cg01693350, cg19211915, cg09234616 and cg13641903 (Figure 4.7 A/D/G/J, $p < 0.05$, Student's t-test corrected for multiple testing). Examining methylation levels through least-squares regression showed that a large proportion of probes were aberrantly methylated when comparing diagnostic to remission samples. After quality control measures were considered, the replication cohort consisted of 63 new samples and 33 previously run samples, for a total of 96.

When comparing the HM450K methylation of probes to the SEQUENOM output of the validation cohort we found that each sample is highly correlated (Figure 4.7 B/E/H/K, $R > 0.8$, Pearson's correlation coefficient). Results from the entire replication cohort ($n=96$) largely validated findings from the original regression analysis. Three out of four CpG sites were significantly differentially methylated (Figure 4.7 C/F/L, $p < 0.05$, Student's t-test) while cg09234616 did not successfully validate, however it did approach significance (Figure 4.7 I, $p = 0.055$, Student's t-test). It should be noted that the replication of cg01693350 and cg09234616 in the SEQUENOM assay were both measured where another CpG site was

incorporated into the fragment, thus methylation output was an average of the two sites. However, considering the correlation of these sites with their HM450K counterpart, it can be assumed that these averaged CpG sites both individually correlated with one another and are not significantly skewing the overall data.

Each of these genes also had several other CpG sites surveyed that were not found on the HM450K. For *WT1* a total of 52 CpG sites were interrogated, however after quality control measures were considered, only 41 CpG sites remained. Several of these CpG sites are found within a single fragment and had their methylation values averaged, therefore a total of 24 data points (CpG units) were obtained from these two assays. All CpG sites showed that diagnostic samples were hypermethylated compared to remission samples with an overall average $\Delta\beta$ of 17.0%. With the exclusion of CpG 39 and 43, all sites were significantly hypermethylated in diagnostic samples (Figure 4.8, $p < 0.05$, Student's t-test).

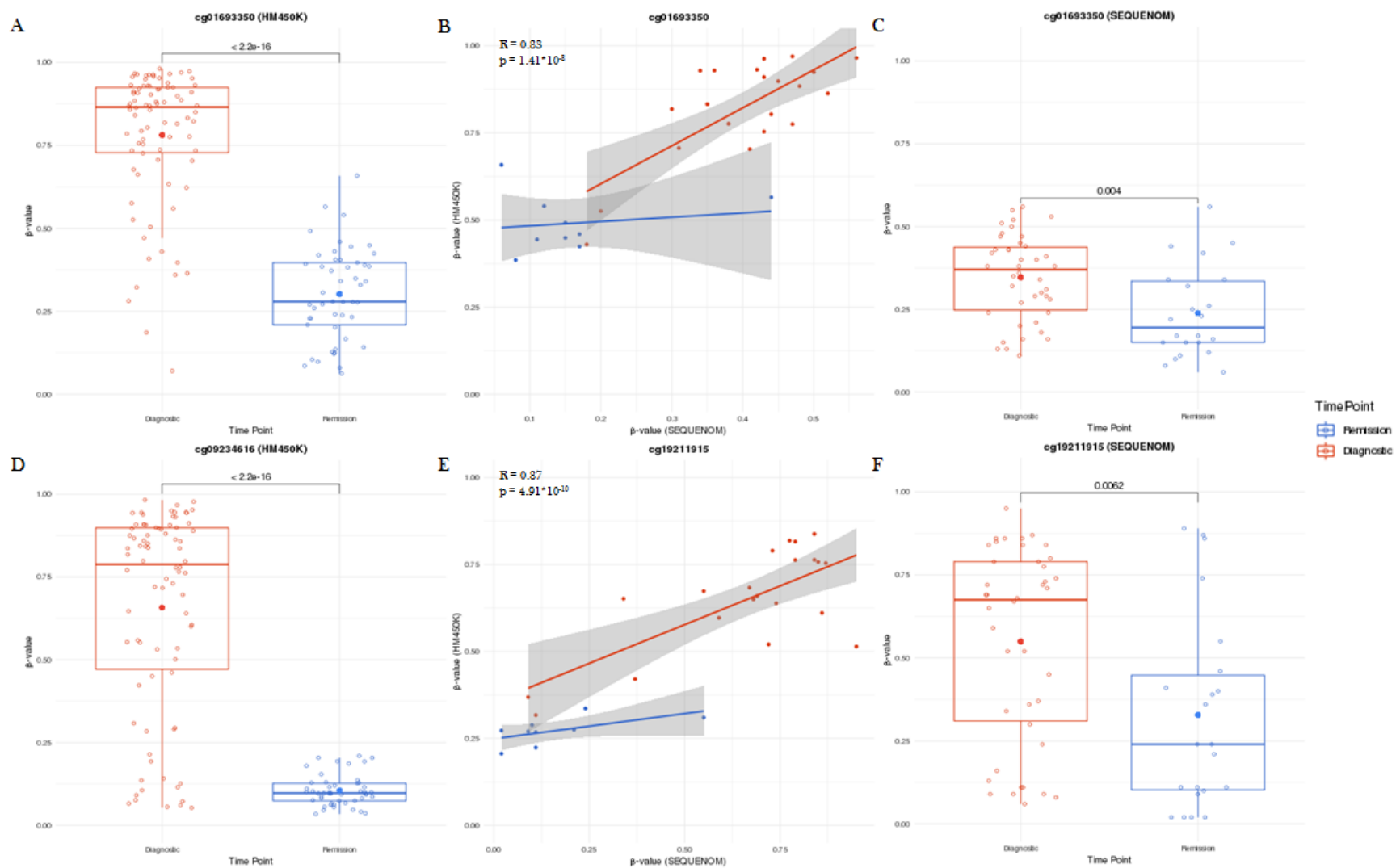


Figure continued on next page.

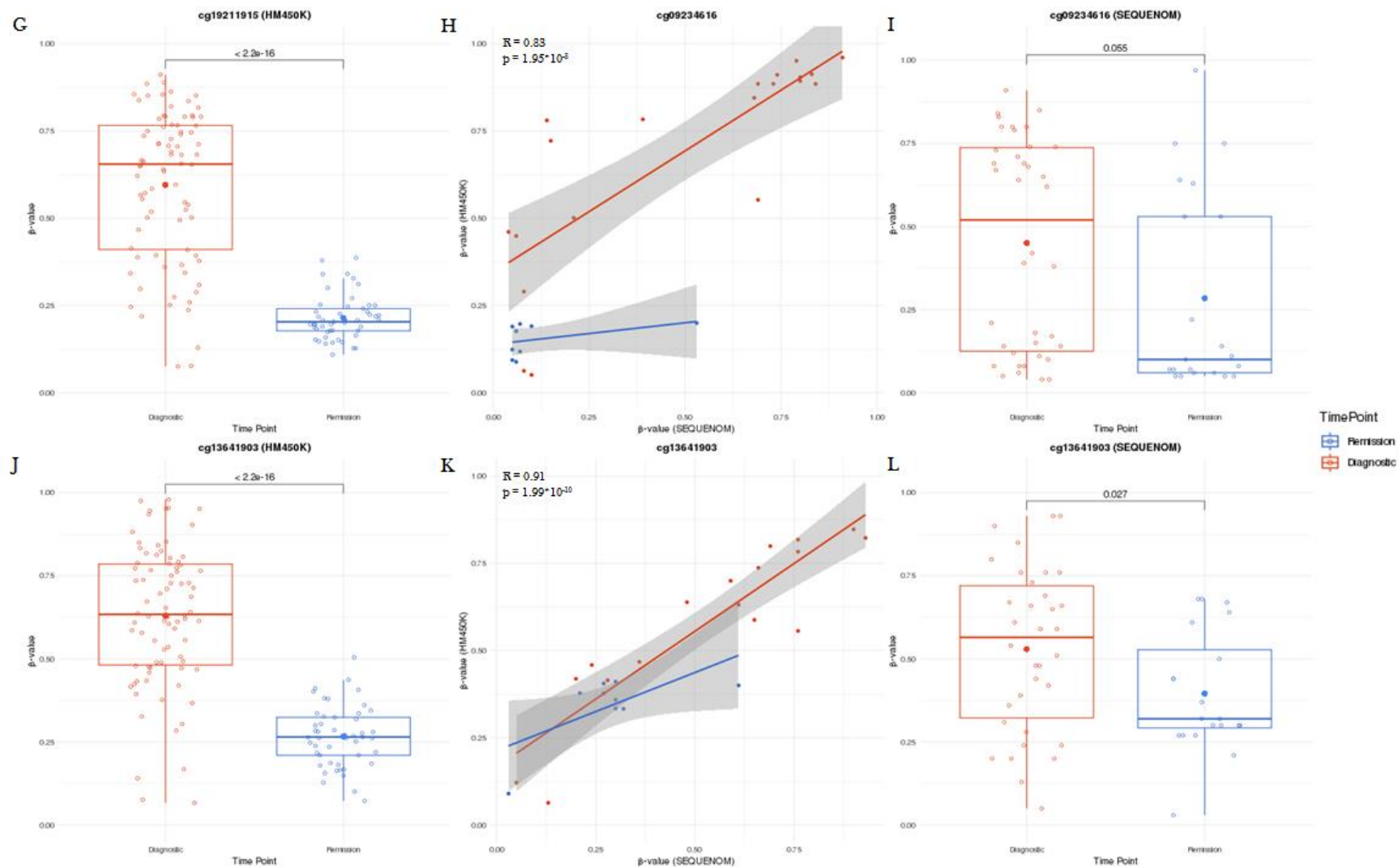


Figure 4.7 Validation of HM450K methylation using MALDI-TOF MS – WT1
Caption on next page.

Boxplots show the median, 0.25/0.75 **IQR**, minimum and maximum methylation between diagnostic and remission time points (n=96). The mean is represented by the coloured circle in the middle of the boxplot and outliers lay beyond the whiskers. Scatterplot show the original **HM450K** methylation on the y-axis and the validation (**SEQUENOM**) methylation on the x-axis. Line of best fit ($\pm 95\%$ **CI**) is added for both diagnostic and remission samples, however statistical analysis (**Pearson's R**) is undertaken on all samples combined. Figures A/B/C) shows a **HM450K** methylation boxplot between patient time points, scatterplot of correlation between **HM450K/SEQUENOM** validation, and **SEQUENOM** methylation boxplot between patient time points, respectively, at cg01693350. Figures D/E/F) are as above for cg19211915. Figures G/H/I) are as above for cg09234616. Figures J/K/L) are as above for cg13641903. All CpG sites were validated ($p < 0.05$, Student's t-test), with the exclusion of cg09234616 which is approaching significance. All sites were highly correlated with their **HM450K** counterpart ($R > 0.8$, Pearson's correlation coefficient).

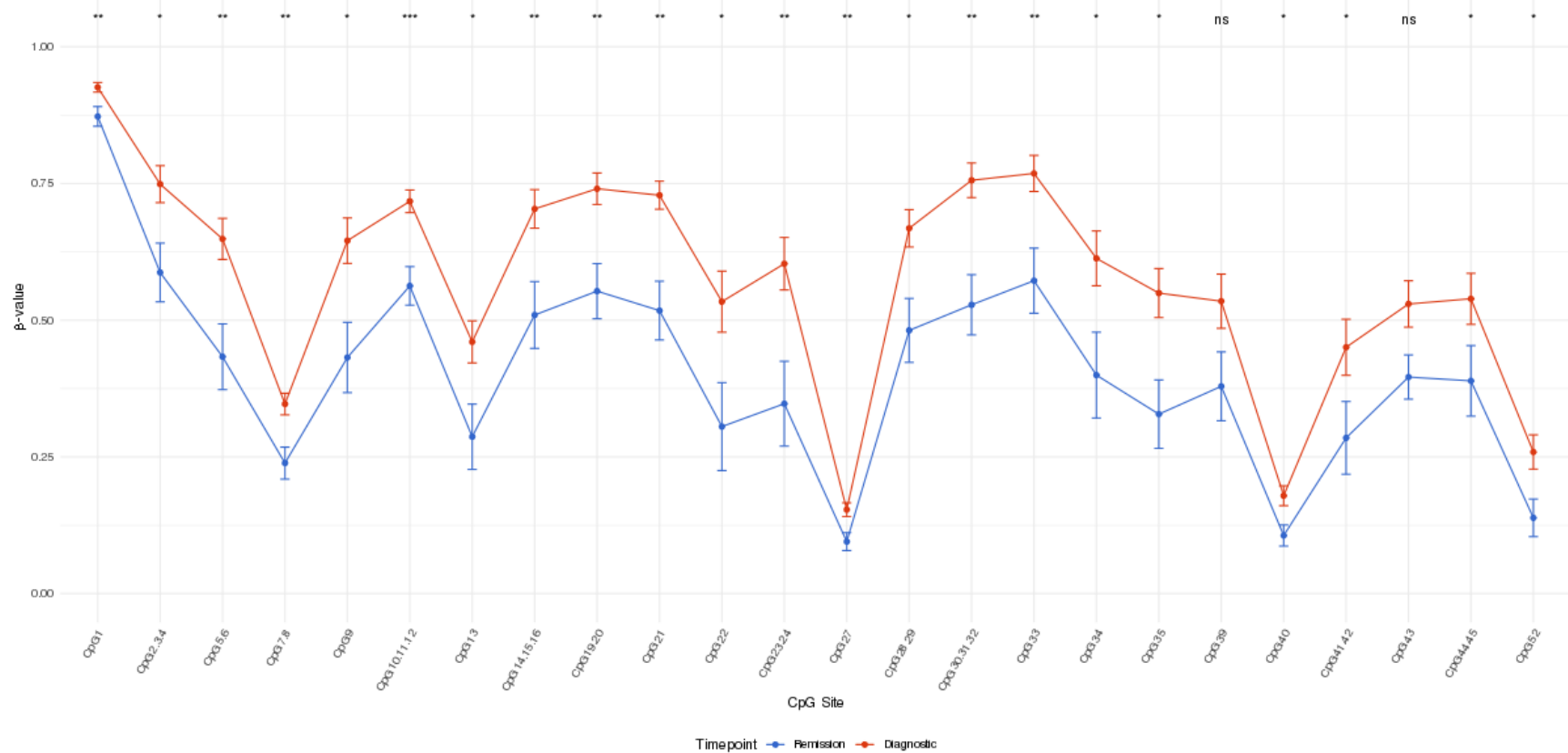


Figure 4.8 Regional differential methylation in *WT1* between diagnostic and remission AML samples

Line graph denoting the level of methylation of each CpG site assayed. Groups are split between diagnostic and remission samples and Student's t-test was used to test for significant differences between groups. Significance is shown as; ns= not significant, *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$. All points are stated as the mean of the group with SEM for error bars. Figure shows the methylation in *WT1* ($n = 96$) across 24 CpG sites. All sites were significantly differentially methylated ($p < 0.05$, Student's t-test) with the exception of CpG 39 and 43.

DGKG

DGKG emerged as a highly likely target to be involved in paediatric AML due to previous studies suggesting it has epigenetically induced tumour suppressor properties (Kai, Yamamoto et al. 2017). An assay was created to encompass as many previously identified significant probes. This assay encompassed cg04391718 and cg24155429 within intron 21 of the gene. Unfortunately cg04391718 was represented by a fragment containing two CpG sites, thus an average of the two values were taken. After quality control measures were taken, methylation of both HM450K samples and samples from the validation cohort (n=29) showed that there was a high correlation of data between the two platforms (Figure 4.9 B/E, $R>0.9$, Pearson's correlation coefficient). Regression analysis of HM450K data revealed that *DGKG* was hypermethylated in diagnostic samples compared to remission samples (Figure 4.9 A/D, $p<0.05$, Student's t-test). Both probes were found to be significantly differentially methylated between patient time points in the replication cohort (n=83), thus validating the previous findings (Figure 4.9 C/F, $p<0.05$, Student's t-test).

For *DGKG*, a total of 20 CpG sites were captured in one assay, however after quality control only 10 CpG sites were accurately captured over 8 CpG units. All CpG sites captured in this analysis were found to be significantly differentially hypermethylated in diagnostic samples with an average $\Delta\beta$ of 27.0% (Figure 4.10, $p<0.05$, Student's t-test). This suggests that *DGKG* is most likely a completely differentially methylated region and may be disrupted in paediatric AML pathogenesis.

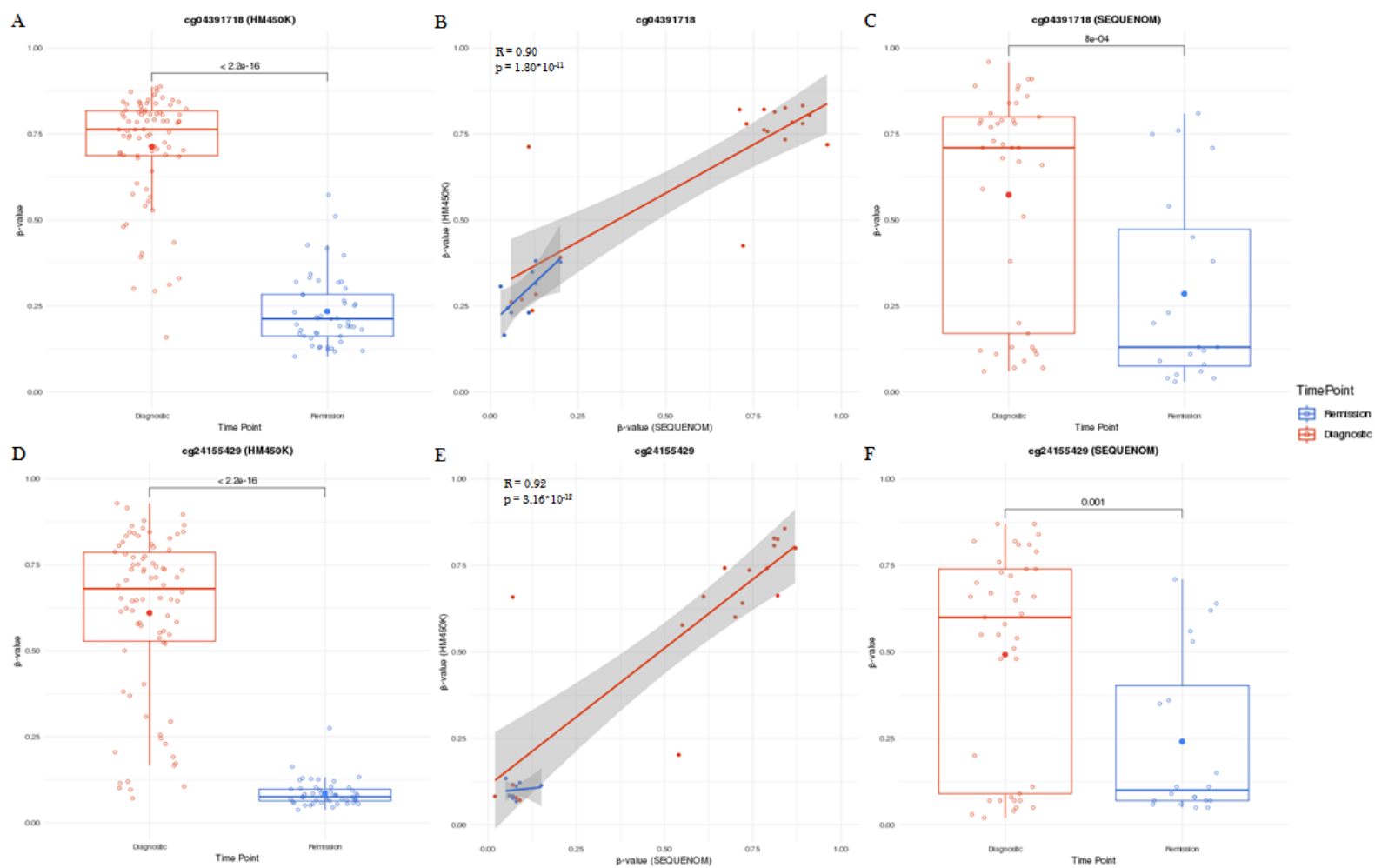


Figure 4.9 Validation of HM450K methylation using MALDI-TOF MS – DGKG
 Caption on next page.

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between diagnostic and remission time points (n=83). The mean is represented by the coloured circle in the middle of the boxplot and outliers lay beyond the whiskers. Scatterplot show the original HM450K methylation on the y-axis and the validation (SEQUENOM) methylation on the x-axis. Line of best fit ($\pm 95\%$ CI) is added for both diagnostic and remission samples, however statistical analysis (Pearson's correlation coefficient) is undertaken on all samples combined. Figures A/B/C) shows a HM450K methylation boxplot between patient time points, scatterplot of correlation between HM450K/SEQUENOM validation, and SEQUENOM methylation boxplot between patient time points, respectively, at cg04391718. Figures D/E/F) are as above for 24155429. All CpG sites were validated ($p < 0.05$, Student's t-test) and were highly correlated with their HM450K counterpart ($R > 0.9$, Pearson's correlation coefficient).

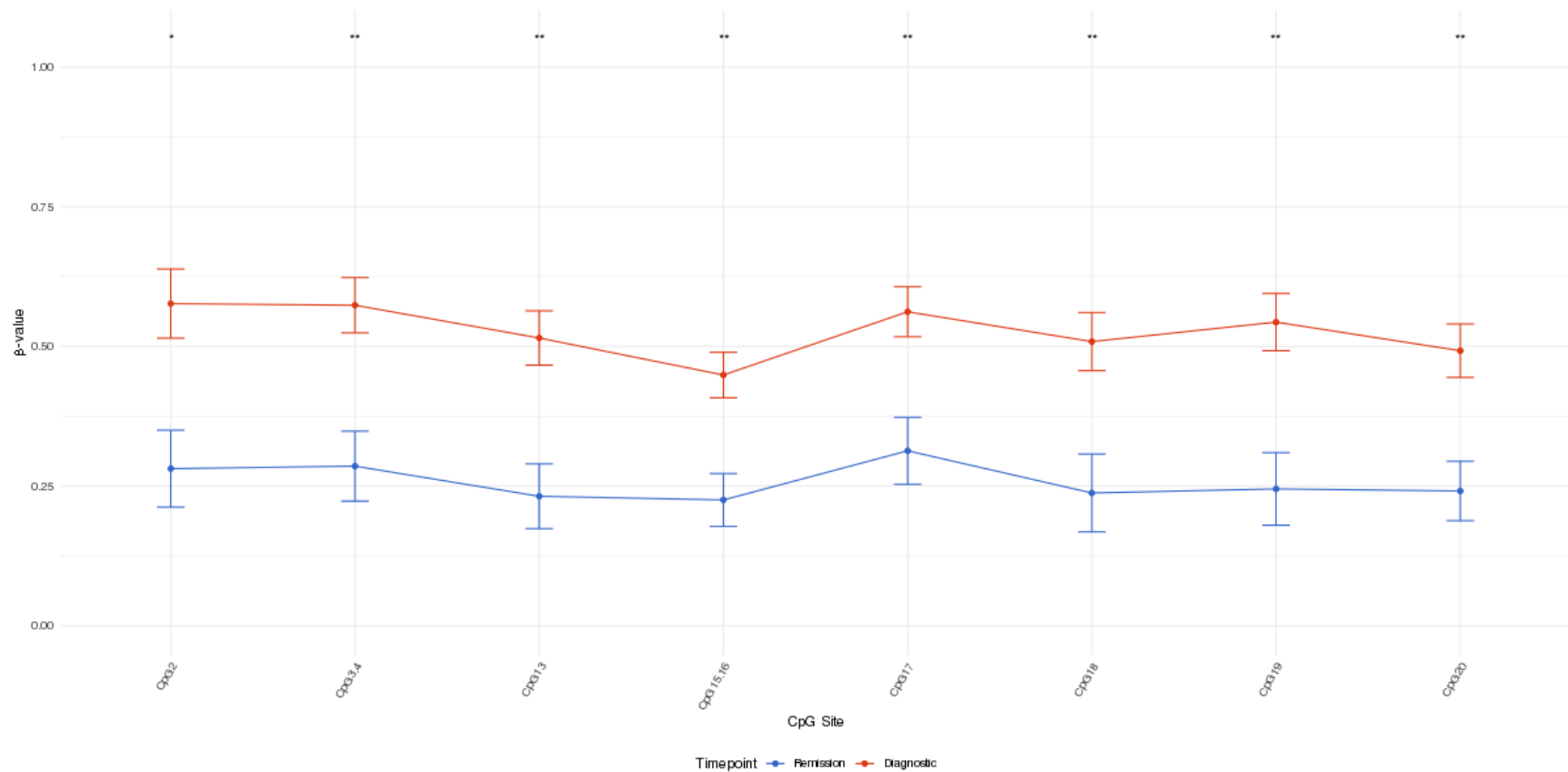


Figure 4.10 Regional differential methylation in *DGKG* between diagnostic and remission AML samples

Line graph denoting the level of methylation of each CpG site assayed. Groups are split between diagnostic and remission samples and Student's t-test was used to test for significant differences between groups. Significance is shown as; ns= not significant, *= $p<0.05$, **= $p<0.01$, ***= $p<0.001$. All points are stated as the mean of the group with SEM for error bars. Methylation is measured in *DGKG* (n=83) across 8 CpG units. All units were significantly differentially methylated ($p<0.05$, Student's t-test).

ZNF876P

ZNF876P had one assay created for validation purposes. This assay encompassed a single previously identified significant CpG site, namely cg17201760 found within the 5'UTR of the gene (Figure 4.11 A, $p < 0.05$, Student's t-test). Correlation analysis between HM450K and SEQUENOM methylation data showed that there is little correlation between samples in the validation cohort ($n=29$, Figure 4.11 B, $R=0.28$, Pearson's correlation coefficient). Despite this lack of correlation, MALDITOF-MS did find that there was a significant difference in methylation in the replication cohort ($n=84$, Figure 4.11 C, $p < 0.05$, Student's t-test). However, the $\Delta\beta$ from the original analysis (48.3%) was much lower in the replication analysis (9.3%). Considering the lack of correlation of samples run both on the HM450K and SEQUENOM, there is potential of measurement error.

Overall, 35 CpG sites were interrogated for methylation. Following quality control 25 CpG sites remained over 18 CpG units. We discovered that all of these CpG units were hypermethylated in diagnostic samples with an average $\Delta\beta$ of 10.0%. This average is lower than the other genes mainly due to the large amount of variation in methylation between individual CpG sites. All sites, with the exclusion of CpG 1, 7, 16 and 17 were found to be significantly differentially methylated (Figure 4.12, $p < 0.05$, Student's t-test).

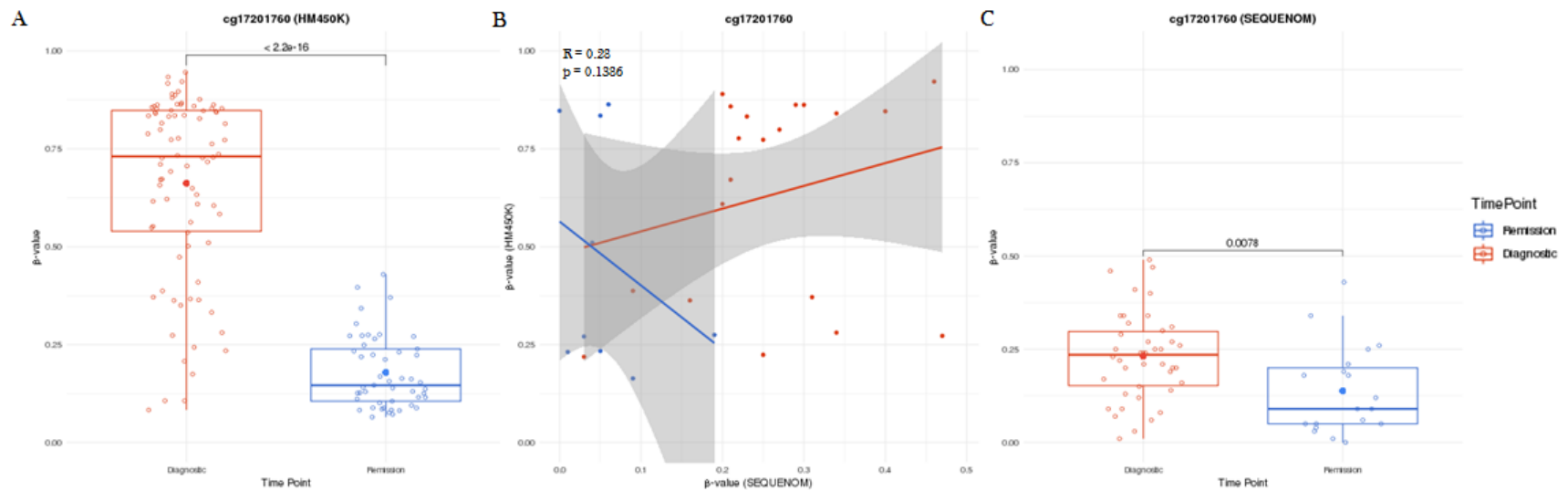


Figure 4.11 Validation of HM450K methylation using MALDI-TOF MS – *ZNF876P*

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between diagnostic and remission time points (n=84). The mean is represented by the coloured circle in the middle of the boxplot and outliers lay beyond the whiskers. Scatterplot show the original HM450K methylation on the y-axis and the validation (SEQUENOM) methylation on the x-axis. Line of best fit ($\pm 95\%$ CI) is added for both diagnostic and remission samples, however statistical analysis (Pearson's correlation coefficient) is undertaken on all samples combined. Figures A/B/C shows a HM450K methylation boxplot between patient time points, scatterplot of correlation between HM450K/SEQUENOM validation, and SEQUENOM methylation boxplot between patient time points, respectively, at cg17201760. This CpG site was validated ($p < 0.05$, Student's t-test) however this was not highly correlated with their HM450K counterpart suggesting that this may yet be a false positive despite validation ($R = 0.28$, Pearson's correlation coefficient).

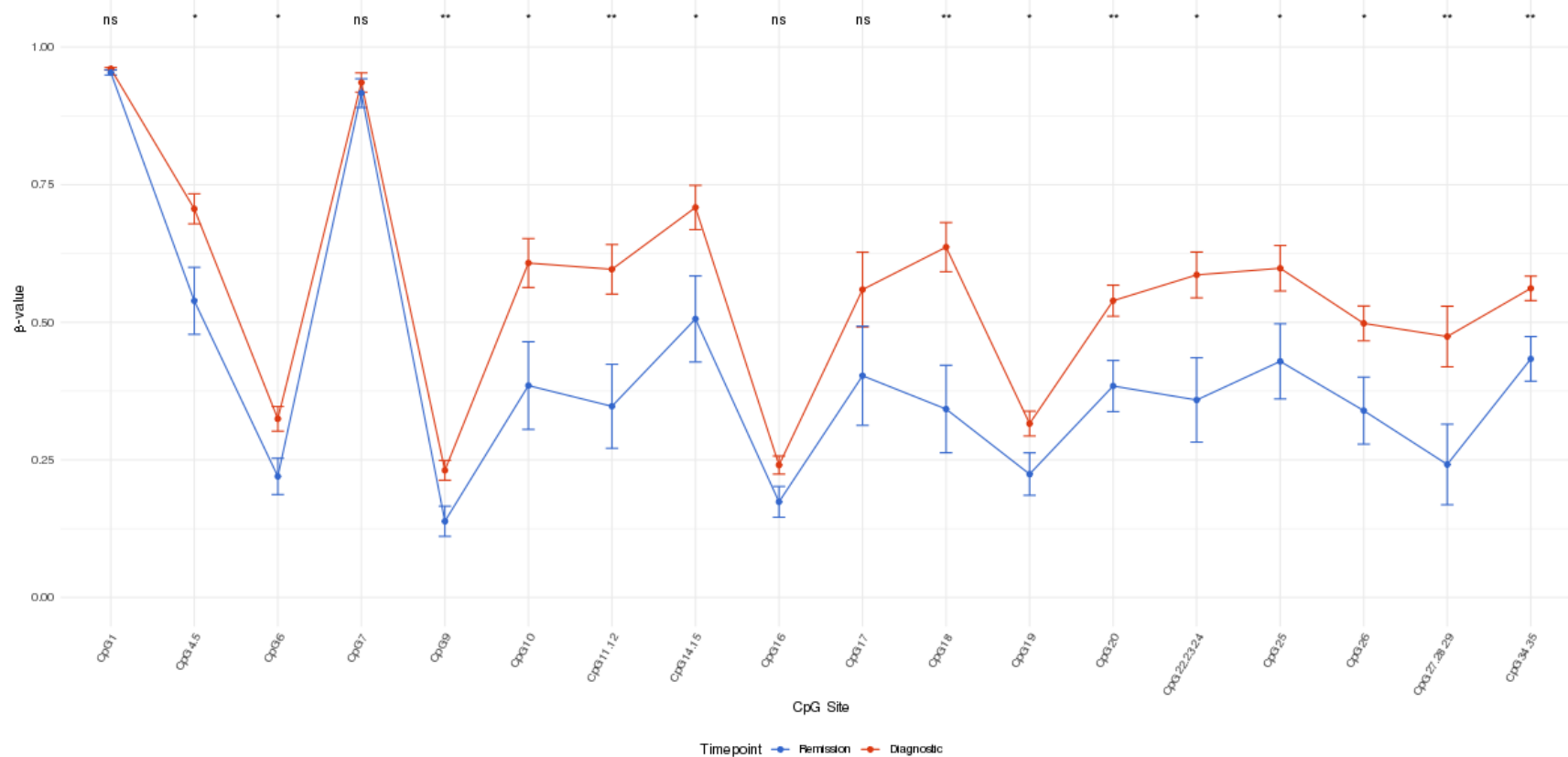


Figure 4.12 Regional differential methylation in *ZNF876P* between diagnostic and remission AML samples

Line graph denoting the level of methylation of each CpG site assayed. Groups are split between diagnostic and remission samples and Student's t-test was used to test for significant differences between groups. Significance is shown as; ns= not significant, *= $p<0.05$, **= $p<0.01$, ***= $p<0.001$. All points are stated as the mean of the group with SEM for error bars. Methylation measured in *ZNF876P* ($n=84$) across 18 CpG units. All units were significantly differentially methylated ($p<0.05$, Student's t-test) with the exception of CpG1, 7, 16 and 17.

4.5 Replication of Differential Methylation in a Combined AML Cohort: A Pooled-Analysis

4.5.1 DNA Methylation Profiling to Identify AML Signatures

Unsupervised analyses were undertaken on the combined Australian and TARGET cohorts (n=298 total samples, diagnostic = 171, remission = 103, healthy = 24). Principal components analysis validated the findings in the Australian cohort as PC1 (containing the majority of variation within the combined dataset) was primarily due to differences between diagnostic and matched remission samples, accounting for 19.4% of total variation (Figure 4.13). PC2, accounting for 9.7% of the variation found in the dataset, was attributed to the differences between the diagnostic/remission and healthy control samples which is likely indicative of age of sample. While close to 30% of variation can be attributed to differences in sample time point that still means that 70% of variation may be due to other unidentified factors.

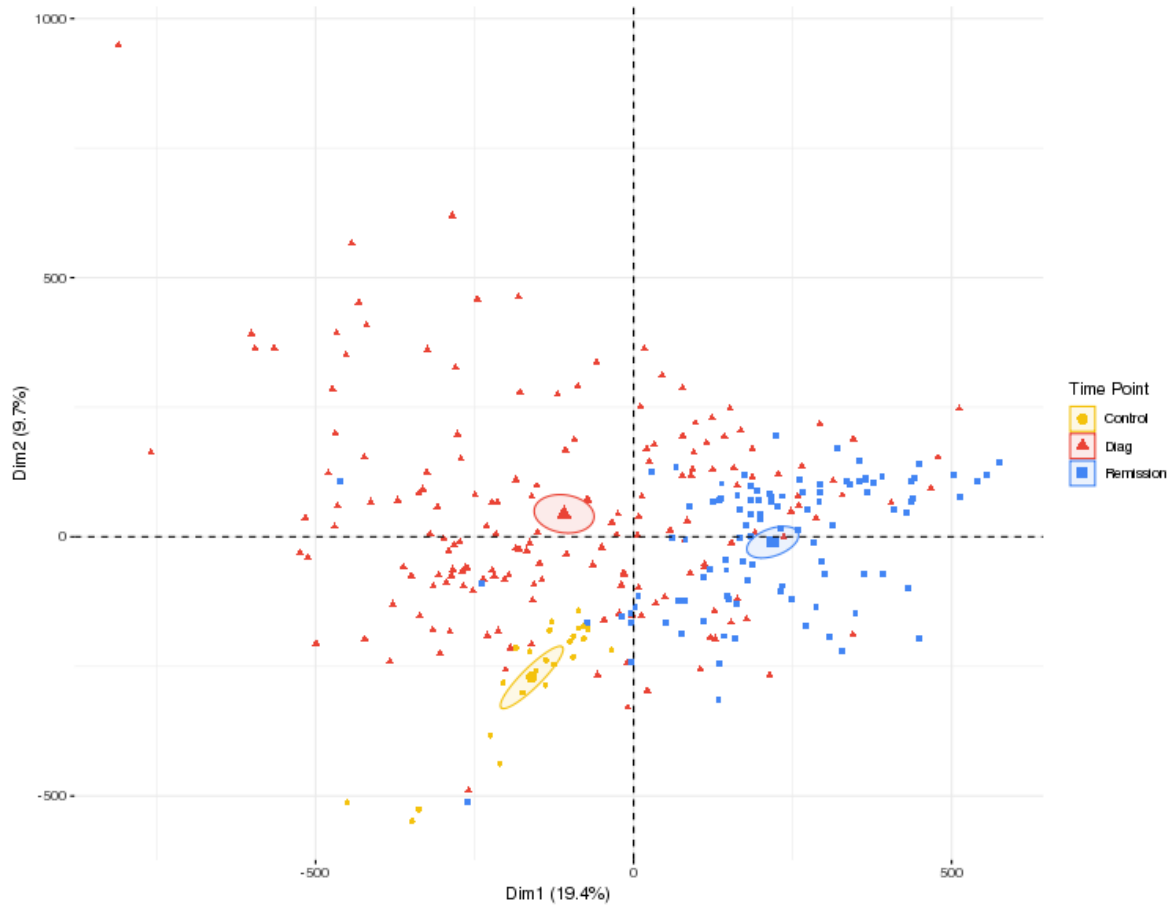


Figure 4.13 Unsupervised principal components analysis of diagnostic and remission AML and healthy adult bone marrow samples in a combined cohort

Beta values from all 373,253 remaining probes from the HM450k across 298 samples were used in a PCA to determine sources of variation in the dataset. The variation in PC1 (accounting for 19.4% total variation) can be partially attributed to differences between matched diagnostic and remission samples. Variation in PC2 (accounting for 9.7% total variation) is partially defined by age of sample (adult vs paediatric). The larger circle, square and triangle represent their respective group means and ellipses drawn around these averages are defined by 0.95 confidence interval.

Least squares regression analysis was utilised using the statistical package RUVm and identified 1622 differentially methylated probes (DMPs, Benjamini-Hochberg adjusted $p < 0.05$) between diagnostic and non-AML samples. Of these DMPs, 878 of them were found in both the Australian only and combined analyses, therefore 744 were newly revealed by the pooled-analysis. Figure 4.14 shows the top differentially methylated probes derived from unsupervised hierarchical clustering, revealing clear separation between diagnostic AML samples from remission and healthy, control bone marrow. Comparing these findings to the Australian cohort only analysis, we do find that there is a lowered ability to distinguish samples, however

this is due to the lowered number of significant probes found in the pooled-analysis and the differing timings of remission samples taken between the two cohorts. Most importantly this further validates the notion that the most important source of variation in the DNA methylome is derived from blast cell composition (cancer vs. non-cancerous sample) as opposed to other confounding factors such as age, sex or differences between cohorts.

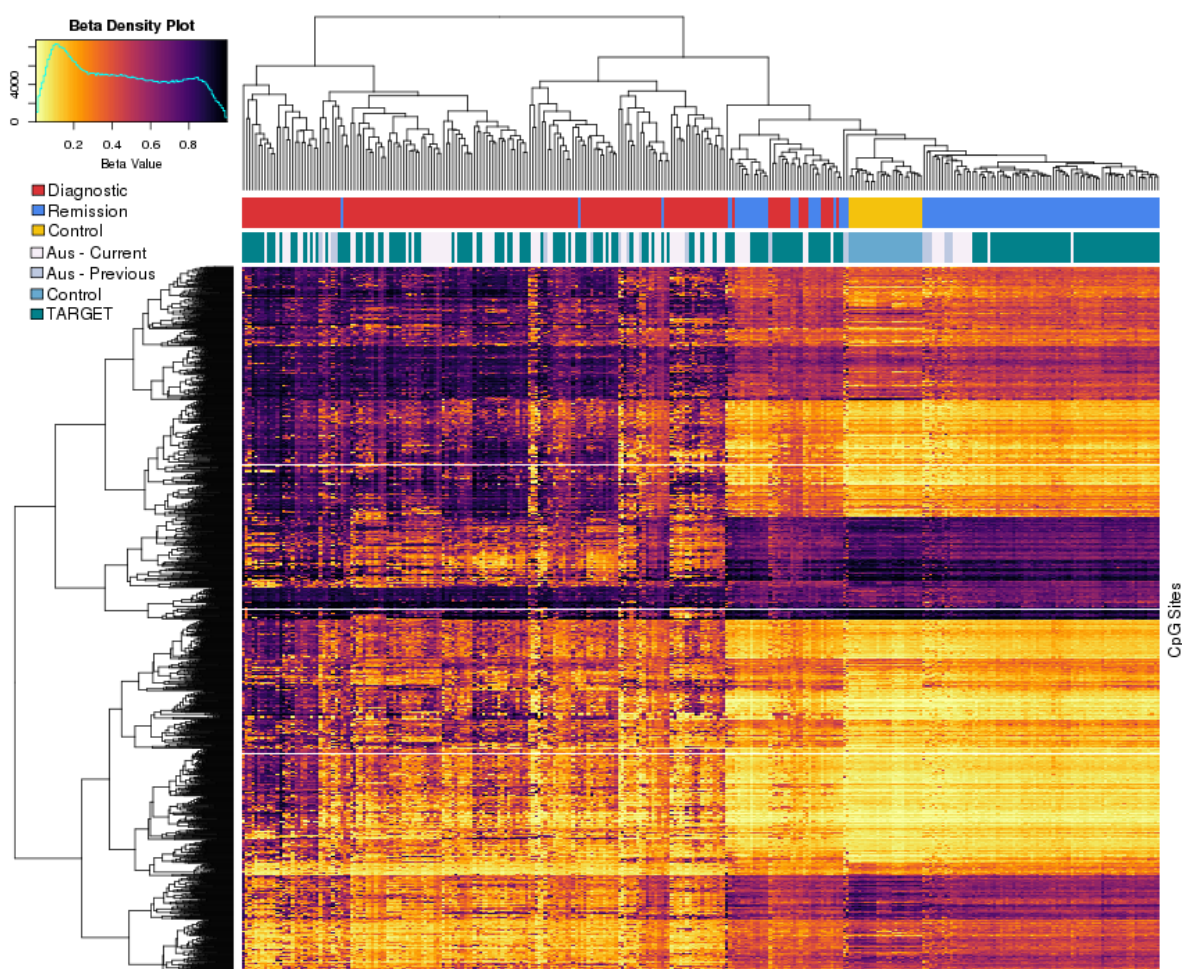


Figure 4.14 Heatmap of beta values across the 1622 highest ranked differentially methylated probes associated with sample time point in a pooled cohort

Heatmap with the y-axis representing 1622 differentially methylated CpG sites while the x-axis shows 298 patient samples where both are ordered by hierarchical clustering as depicted by the dendrograms. This heatmap shows that DNA methylation can somewhat accurately distinguish diagnostic AML bone marrow (red) from remission (blue) and healthy (yellow) adult bone marrow samples. Patient cohort has little influence upon the top differentially methylated probes.

4.5.2 Genes Differentially Methylated in Diagnostic AML Samples – A Pooled-Analysis

Validation of differential gene methylation as described in Section 4.3.2 was undertaken in the combined Australian and TARGET cohorts. The previously identified probes in *WT1*, *DGKG*, *ZSCAN1* and *ZNF876P* were again interrogated for DNAm differences. A total of 14 probes within *WT1* were validated and found to be statistically significant after correcting for multiple testing ($\Delta\beta$ 38.8%, Figure 4.15A, Benjamini-Hochberg adjusted $p < 0.05$). The two differentially methylated regions remained intact despite the loss of 9 significant probes when compared to previous analyses. Again, these DMRs were found to be in previously identified regions of differential methylation (Doi, Park et al. 2009) and also within enhancer regions of the gene.

The previously identified *DGKG* DMR was found to have four significantly differentially methylated probes in the combined analysis (Figure 4.15B, Benjamini-Hochberg adjusted $p < 0.05$) with an average $\Delta\beta$ of 49.8%, an effect size decrease of 5% from the previous analysis. Three probes in *ZSCAN1* validated from the original findings, with an average $\Delta\beta$ of 48.5%, an increase in effect size $\Delta\beta$ from the 43.2% previously found (Figure 4.15C, Benjamini-Hochberg adjusted $p < 0.05$). Only two out of four *ZNF876P* probes validated, however due to their distance from one another, the linked DMR was no longer present (Figure 4.15D, Benjamini-Hochberg adjusted $p < 0.05$). The average $\Delta\beta$ of the remaining significant probes was 41.4%, down 4.2%.

Novel genes were also identified in this analysis including Meiotic Recombination Protein (*REC8*) and Iroquois Homeobox 1 (*IRX1*). *REC8* is involved in the structural maintenance of chromosomes during meiosis, regulating sister chromatid cohesion and recombination (Garcia-Cruz, Brieno et al. 2010). A total of 8 probes in *REC8* were hypermethylated with an average $\Delta\beta$ of 38.5% (Figure 4.15E, Benjamini-Hochberg adjusted $p < 0.05$). One DMR was discovered localised to the transcription start site and 5'UTR region of the gene. Three of the associated probes were found to be in sites of DNase hypersensitivity and another two in an enhancer region.

IRX1 is a homeobox gene encoding for **IRX1**, involved in pattern formation in the embryo and previously implicated in tumour suppression in gastric, head and neck cancers (Guo, Liu et al. 2010). Five probes identified as differentially hypermethylated in diagnostic cases compared to non-AML samples, with an average $\Delta\beta$ of 25.9% (Figure 4.15F, Benjamini-Hochberg adjusted $p < 0.05$). Together, these probes created a single DMR with four adjacent probes being differentially methylated in a region of DNase hypersensitivity (Consortium 2012).

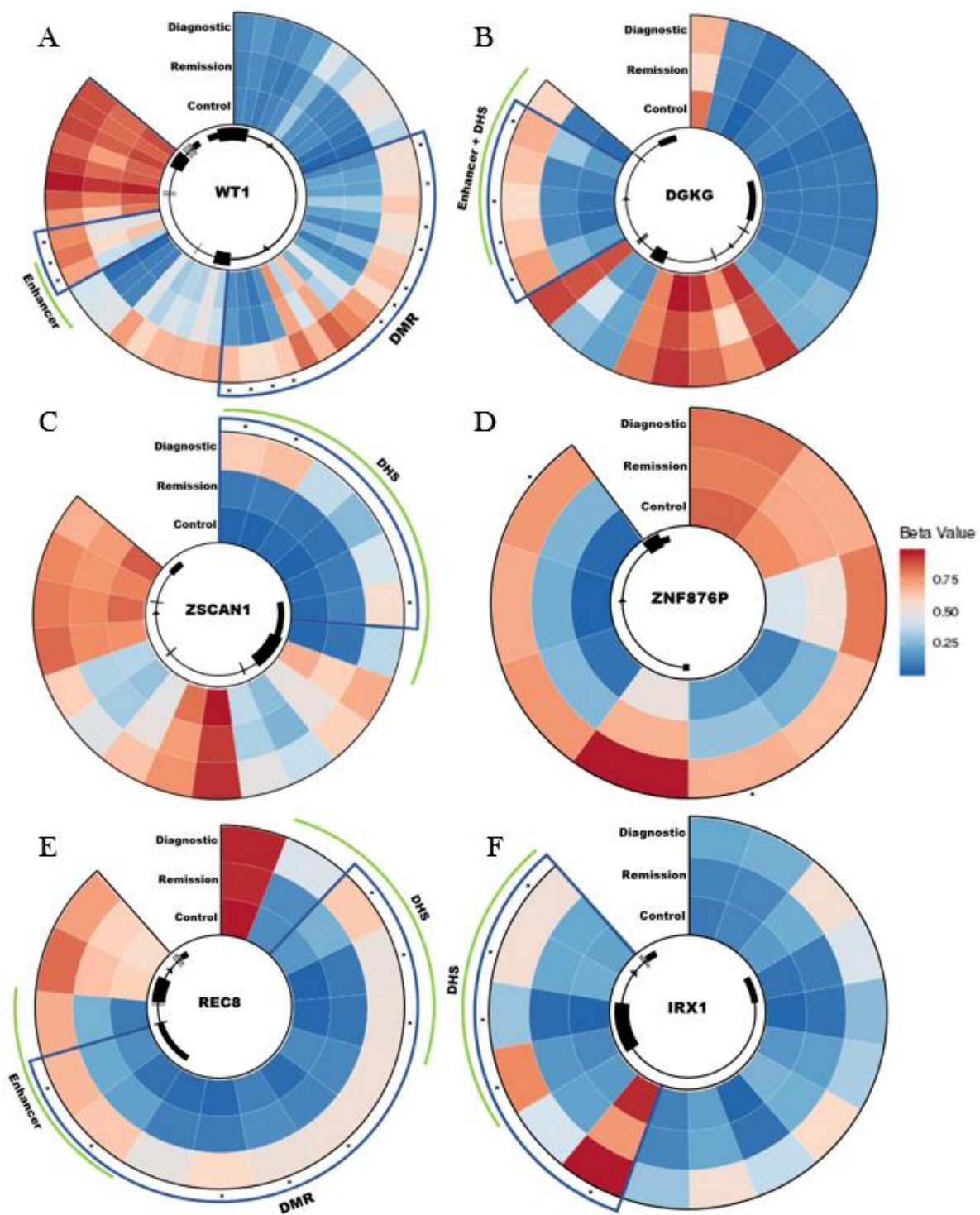


Figure 4.15 Circular heatmaps indicating DNA methylation levels in diagnostic, remission and control samples of all probes within identified genes of interest after pooled-analysis

Caption on next page.

Circular heatmaps shows DNA methylation at each probe that is found in the gene or within 1500 base pairs of the associated untranslated regions (n=131 samples). Inner circles identify the geographical position of any given probe in the gene (not to scale). Methylation levels are depicted by a spectrum of colour, with completely unmethylated probes dark blue while complete methylation is represented by dark red. Significantly differentially methylated probes are indicated by asterisk ($p < 0.05$, Student's t-test). Differentially methylated regions (DMRs), are indicated by blue wedges. Enhancer regions or regions of DNase hypersensitivity found within DMRs are indicated by the outer green line as defined by ENCODE and Doi *et al.* (Doi, Park et al. 2009; Consortium 2012).

A) Methylation levels at 57 *WT1* probes. 14 probes were identified to be significantly differentially methylated across two DMRs with an average $\Delta\beta$ of 38.8%. B) Methylation levels at 26 *DGKG* probes. Four probes were identified to be differentially methylated across one DMR with an average $\Delta\beta$ of 49.8% C) Methylation levels at 20 *ZSCAN1* probes. Three probes were identified to be significantly differentially methylated across one DMR with an average $\Delta\beta$ of 48.5%. D) Methylation levels at nine *ZNF876P* probes. Two probes were identified to be significantly differentially methylated with an average $\Delta\beta$ of 41.4%. The differentially methylated region was unable to be validated. E) Methylation levels at 15 *REC8* probes. Eight probes were identified to be differentially methylated across one DMR with an average $\Delta\beta$ of 38.5%. F) Methylation levels at 16 *IRX1* probes. Five probes were identified to be differentially methylated across a single DMR with an average $\Delta\beta$ of 25.9%.

The top individual probes are identified in Table 4.2 by regression analysis as defined by Benjamini-Hochberg adjusted p-value. These include *SERINC2* (Benjamini-Hochberg adjusted $p=1.32 \times 10^{-8}$, Figure 4.16A), *LPHN1* (Benjamini-Hochberg adjusted $p=2.48 \times 10^{-7}$, Figure 4.16B), *FBRSL1* (Benjamini-Hochberg adjusted $p=5.26 \times 10^{-7}$, Figure 4.16C), *TREX1* (Benjamini-Hochberg adjusted $p=5.26 \times 10^{-7}$, Figure 4.16D), *MAFG* (Benjamini-Hochberg adjusted $p=5.26 \times 10^{-7}$, Figure 4.16F), *XRCC3* (Benjamini-Hochberg adjusted $p=5.7 \times 10^{-7}$, Figure 4.16G), *B3GNTL1* (Benjamini-Hochberg adjusted $p=7.5 \times 10^{-7}$, Figure 4.16H), and *BTBD3* (Benjamini-Hochberg adjusted $p=1.07 \times 10^{-6}$, Figure 4.16I). Overall we find that only 17 probes remain within the top 100 significantly differentially methylated probes list when compared to discovery analysis. While findings were still inflated in this study ($\lambda = 1.31$), inflation was reduced, largely due to the increase in sample numbers.

Looking at the individual top differentially methylated probes illustrates two things. The first being, like the previous analysis, that the differences between diagnostic compared to non-AML bone marrow is the primary source of variation in the data. Secondly, we still find that not all diagnostic samples have similar behaviour at any given CpG site suggesting that there are site specific aberrant DNAm associated with particular subsets of diagnostic samples. This variance found in diagnostic samples is also found in remission and control samples, however the total variance of these samples is smaller than that of diagnostic samples. While remission and control bone marrow samples seem uniform in their methylation compared to diagnostic samples, there are examples such as Figure 4.16B and Figure 4.16G where we found that DNAm differences were driven by remission or control sample variation as opposed to diagnostic AML methylation. Comparing these probes to the discovery cohort analysis we find that only one probe (cg08173263) remains highly significant in pooled-analysis.

Table 4.2 Top 100 probes as defined by least squares regression analysis comparing AML to non-AML bone marrow samples in a pooled-analysis

Probe ID	Gene	Adjusted p-value [†]	$\Delta\beta$	Probe ID	Gene	Adjusted p-value [†]	$\Delta\beta$
cg17086398	SERINC2	1.32E-08	-0.363	cg13918150		1.32E-05	0.301
cg08173263	LPHN1	2.48E-07	-0.219	cg07766263	SLC7A14	1.35E-05	0.404
cg20454518	FBRSL1	5.26E-07	-0.366	cg09523782	WIPI2	1.35E-05	-0.199
cg09882118	TREX1	5.26E-07	-0.130	cg10549018	TLL2	1.38E-05	-0.174
cg07504763		5.26E-07	0.377	cg23181580		1.47E-05	0.465
cg09489844	MAFG	5.26E-07	-0.186	cg08813888		1.48E-05	-0.164
cg12798040	XRCC3	5.70E-07	-0.111	cg04304450	BIK	1.48E-05	-0.220
cg15727507	B3GNTL1	7.50E-07	-0.204	cg02286506		1.67E-05	0.216
cg05822633	BTBD3	1.07E-06	0.482	cg08310363	MIR181C	1.70E-05	0.239
cg18158306	FBRSL1	1.07E-06	-0.286	cg19637821	FBRSL1	1.91E-05	-0.246
cg04682802	BTBD3	1.18E-06	0.477	cg15131146	MAN1C1	2.51E-05	-0.222
cg09560590	HMHB1	1.37E-06	-0.255	cg17714703	UHRF1	2.87E-05	-0.301
cg22602105	SPON2	1.37E-06	0.232	cg09924848	FAM53B	2.96E-05	-0.330
cg04153882	WIPI2	1.37E-06	-0.133	cg24039042		2.98E-05	-0.225
cg00989249	N4BP3	1.39E-06	-0.185	cg01908551	METRNL	3.20E-05	0.148
cg17822325	SERINC2	1.53E-06	-0.344	cg21865657	PIP5K1C	3.33E-05	0.319
cg01766850		1.56E-06	-0.197	cg08466792		3.68E-05	0.280
cg01395047	TLR9	1.56E-06	-0.214	cg11254700		3.85E-05	0.629
cg06791426	FBRSL1	1.56E-06	-0.246	cg24488469		3.98E-05	-0.159
cg16145216	HIVEP3	2.14E-06	-0.315	cg11682724		4.02E-05	-0.273
cg24049493	HIVEP3	2.19E-06	-0.375	cg03741653	BMP1	4.07E-05	0.166
cg05251389	BIK	2.26E-06	-0.204	cg04252928	ACSF3	4.07E-05	-0.177
cg18482892	ARHGEF3	2.29E-06	0.220	cg13696706	DAB2IP	4.07E-05	-0.217
cg06585708		2.29E-06	0.230	cg20204986	WT1	4.36E-05	0.218
cg11084334	LHFPL4	2.29E-06	0.368	cg00161791		4.42E-05	-0.192
cg26606789		2.29E-06	-0.106	cg15084543	ELTD1	5.11E-05	0.620
cg11625488		2.29E-06	-0.263	cg02228160	HMHB1	5.11E-05	-0.299
cg13827209	TGFBR1	2.80E-06	-0.235	cg00715696	KIAA0182	5.11E-05	-0.165
cg06426416	EVC2	2.80E-06	0.269	cg22533573	WT1	5.19E-05	0.376
cg00346883	PLA2G15	3.00E-06	-0.252	cg26926221	UTRN	5.27E-05	-0.155
cg12761472		3.75E-06	0.169	cg09695430	WT1	5.64E-05	0.281
cg14012925	CUX1	3.98E-06	-0.192	cg17280346	ZIC1	5.70E-05	0.521
cg10312186		3.98E-06	-0.263	cg13544075	LCN6	6.86E-05	-0.266
cg02325250	CSF2	3.98E-06	0.238	cg21121843	HTT	8.28E-05	0.200
cg02158978	FBRSL1	4.23E-06	-0.311	cg18125510	WARS	8.29E-05	0.172
cg14192259		4.79E-06	-0.115	cg22108374	CCDC33	8.56E-05	0.297
cg06065019		4.79E-06	-0.222	cg12999267		8.74E-05	-0.147
cg26838949	KIF21B	4.79E-06	0.206	cg10678215	HMHB1	9.76E-05	-0.250
cg19900821		4.87E-06	-0.290	cg02505827	MARVELD2	9.80E-05	0.370
cg22539182		8.17E-06	-0.282	cg08996521	CISH	0.00010107	0.243
cg27630153	FAM38A	9.59E-06	-0.253	cg01294808	IRX1	0.00010107	0.485
cg17232357	SMAD6	1.03E-05	0.164	cg04225775	HMHB1	0.00010107	-0.230
cg01669161		1.06E-05	-0.216	cg16536330	LEPROTL1	0.00010107	0.177
cg05135521	RBMS1	1.11E-05	0.290	cg13437554	OAZ1	0.00010107	-0.118
cg18062140	RBM38	1.14E-05	-0.208	cg22576236	SPON2	0.00010107	0.274
cg15743533	FAM110A	1.17E-05	0.231	cg06729532		0.000101091	0.211
cg06699555		1.17E-05	0.249	cg03880642	SH3PXD2A	0.000111944	-0.199
cg03978682	ARL15	1.17E-05	0.290	cg03372852	LILRP2	0.000111944	-0.171
cg18686153		1.17E-05	-0.201	cg25002788	SPON2	0.000111944	0.319
cg23450038		1.17E-05	0.386	cg08404702	TSC2	0.000120421	-0.295

[†]Adjusted p-value calculated by Benjamini-Hochberg method.

$\Delta\beta$ = Difference in methylation calculated as AML average minus non-AML average

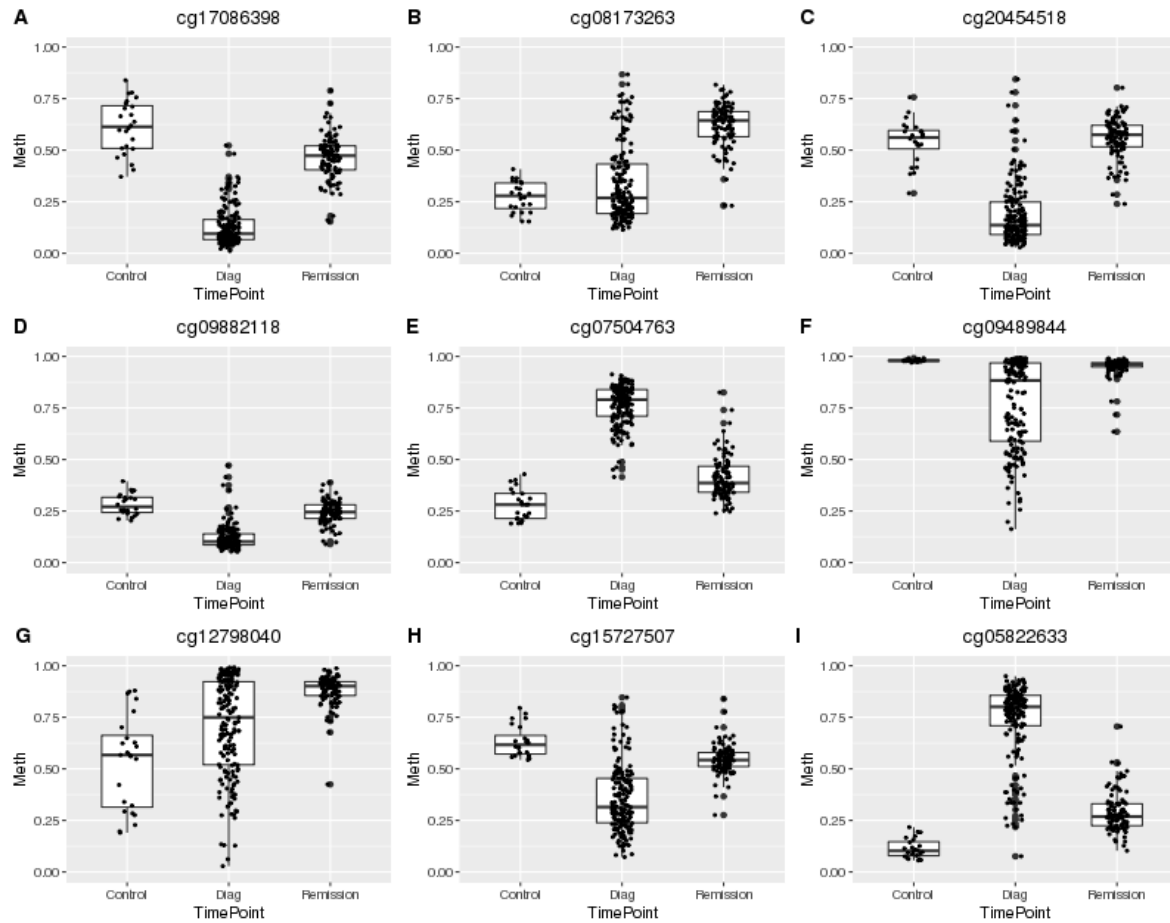


Figure 4.16 Boxplots of the most significantly differentially methylated probes between diagnostic and non-AML bone marrow samples after pooled-analysis

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between patient bone marrow biopsy time points (n=298). The median is represented by the line in the middle of the boxplot while outliers lay beyond the whiskers. Probes were identified as the most statistically significant DMPs (Benjamini-Hochberg adjusted $p \leq 1.07 \times 10^{-6}$). Probes identified can be attributed to the genes A) *SERINC2*, B) *LPHN1*, C) *FBRSL1*, D) *TREX1*, E) Intergenic region, F) *MAFG*, G) *XRCC3* H) *B3GNTL1*, I) *BTBD3*.

4.5.3 Pathway Analysis

Pathway analysis was undertaken on the previously identified differentially methylated probes found in the regression pooled-analysis. Again, utilising the GOMeth package in R, we identified 19 gene ontology terms that were associated with the 1622 probes in the analysis ($p < 0.05$, FDR adjusted). This was further refined using the Reduce + Visualise Gene Ontology (REViGO) software where a total of eight GO umbrella terms were identified. Cytoscape software identified that four of these terms were linked to one another, strongly centred on cell communication, signalling, transduction and development, all of which were highly significant in the discovery pathway analysis too. Other processes identified include biological adhesion, homophilic cell adhesion and developmental processes. This suggests that cell communication, adhesion and development are all systems that are targeted and aberrantly methylated in the development of paediatric AML. Adhesion and cell communication were the most commonly found GO terms in this analysis.

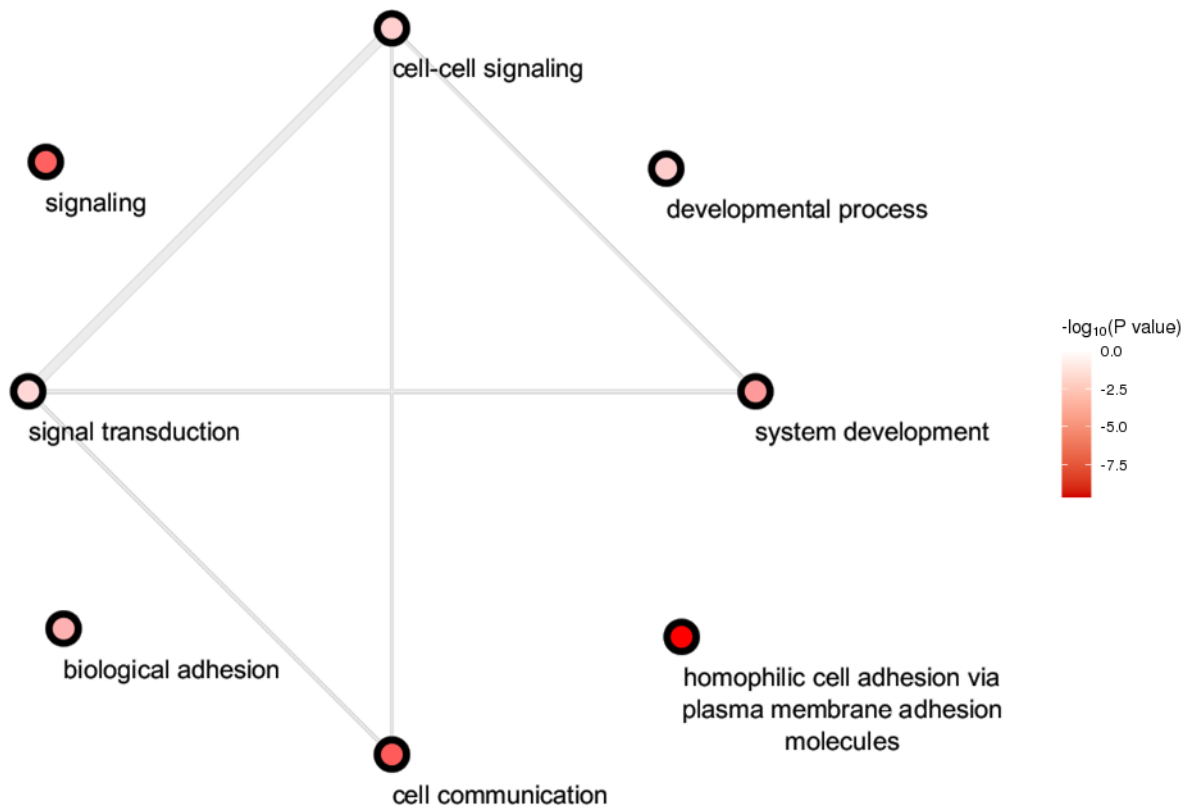


Figure 4.17 REViGO gene ontology network associated with 1622 differentially methylated probes identified in a diagnostic vs remission regression analysis

REViGO gene ontology reduction identified eight different GO terms associated with the 1622 significantly differentially methylated probes derived from regression analysis. $\log_{10}$ p values are shown for each GO term to indicate the significance of the number of probes that are associated with that term appearing by chance. Cell signalling and transduction, along with developmental processes dominate the GO terms associated with aberrantly methylated probes identified in this analysis.

4.6 Identification of Differential DNA Methylation Associated with Patient Relapse

4.6.1 Methylation Profiling of Diagnostic and Matched Relapsed Bone Marrow Samples

Several studies have revealed variation in DNAm patterns in matched diagnostic and relapse samples for a range of adult onset cancers, raising the possibility that epigenetic variation plays a role in the disease progression and relapse (Garcia-Manero, Bueso-Ramos et al. 2002; Kroeger, Jelinek et al. 2008; Jiang, Dunbar et al. 2009). In order to test this in paediatric AML the pooled-analysis sample cohort was refined to patients who had both a diagnostic and matched relapse sample. The resulting cohort had 53 patients at each disease stage totalling 106 samples. Unsupervised analyses including PCA was utilised to identify factors contributing to variance within the dataset. Plotting PC1 (15.4% of total variation) and PC2 (9% of total variation) against one another revealed little difference between diagnostic and relapse samples (Figure 4.18). Further analysis of other principle components also failed to identify relapse status as a major source of variation in the dataset. While the data does suggest little difference between diagnosis and relapse, regression analysis was still undertaken to elucidate any individual probes that may be driving subtle levels of currently undetected variation in the data.

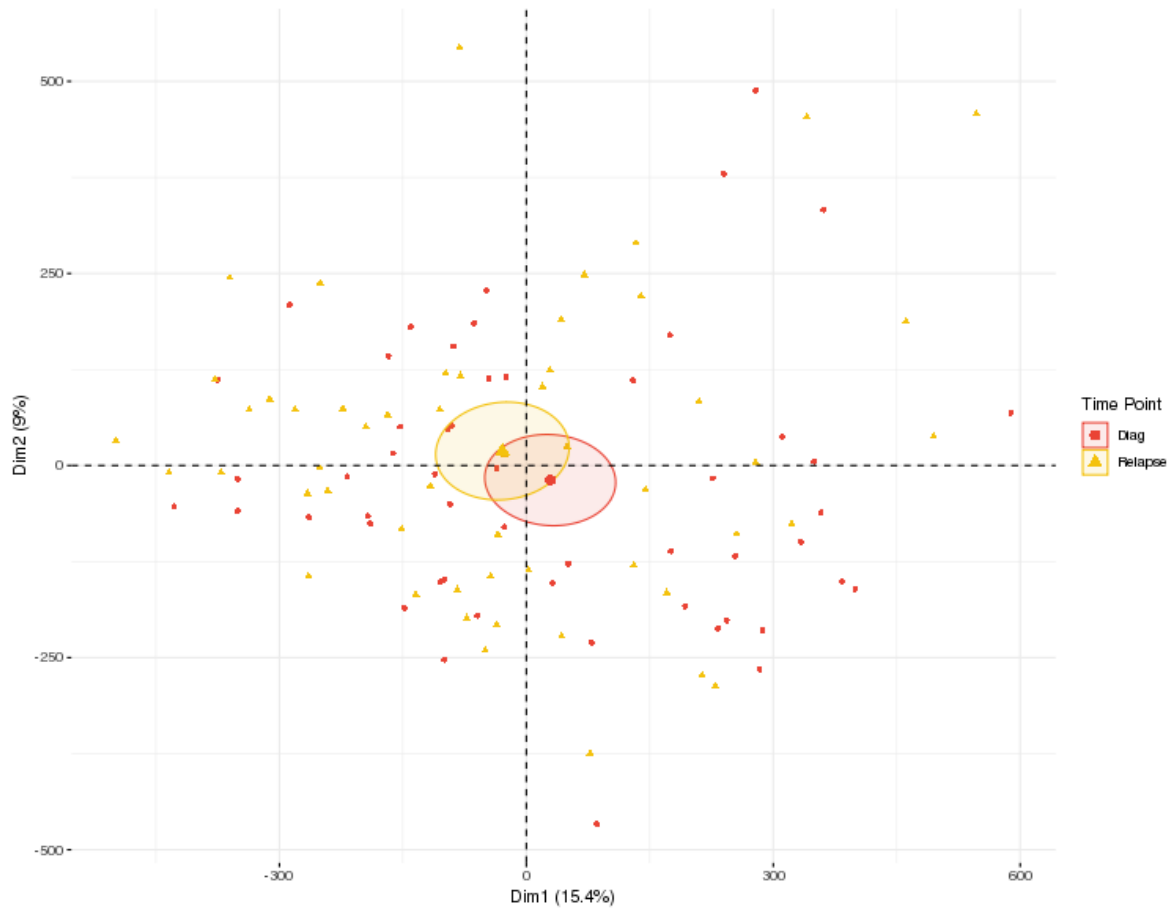


Figure 4.18 Unsupervised principal components analysis plot defining diagnostic from relapsed bone marrow samples in a pooled cohort

Beta values from all 373,253 probes from the HM450k across 106 bone marrow samples (53 matched diagnostic vs relapse AML) were used in a PCA to determine sources of variation in the dataset. The variation in PC1 (accounting for 15.4% total variation) and PC2 (accounting for 9% total variation) failed to distinguish any differences between matched samples. The larger circle, square and triangle represent their respective group means and ellipses drawn around these averages are defined by 0.95 confidence interval.

Regression analysis was undertaken on all 53-paired diagnostic and relapsed samples revealing 341 DMPs (unadjusted $p < 0.05$, Student's t-test). Of these, 191 remained significant once accounting for multiple testing (Benjamini-Hochberg adjusted $p < 0.05$). Utilising all probes in hierarchical clustering showed no capacity to distinguish diagnostic and relapsed samples. Hierarchical clustering of the 50 most significant probes (Figure 4.19) revealed some (albeit limited) capacity to separate diagnostic and relapse samples. Two major groups were identified, the first containing 27 diagnostic samples with 8 relapsed samples, and the second 45 relapsed samples with the 26 diagnostic samples.

Figure 4.20 shows that while there is a large amount of variation in the diagnostic and relapsed samples, there is still a notable disparity in the average methylation at several CpG sites.

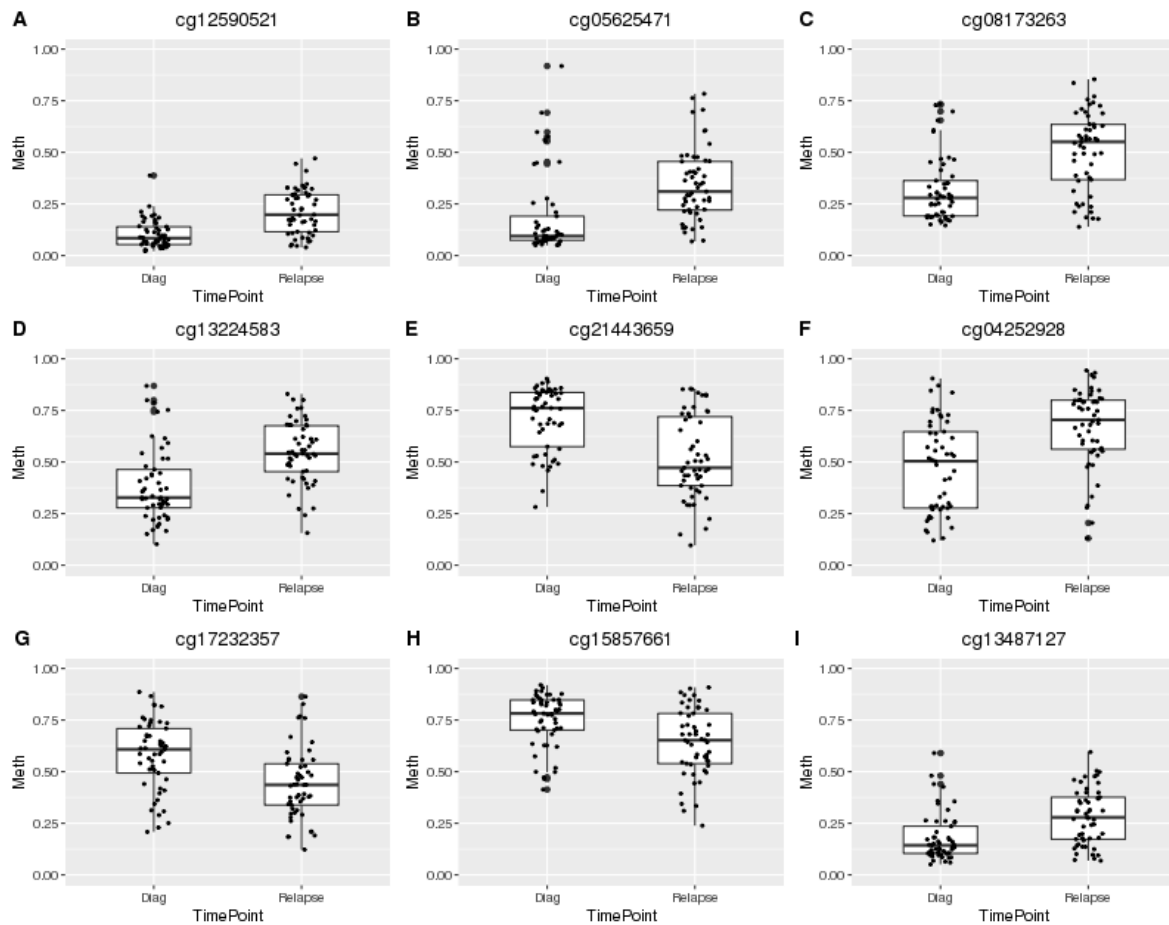


Figure 4.20 Boxplots of the most significantly differentially methylated probes between diagnostic and relapsed bone marrow samples in a pooled-analysis

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between patient bone marrow biopsy time points (n=106). The median is represented by the line in the middle of the boxplot while outliers lay beyond the whiskers. Probes were identified as the most statistically significant DMPs (Benjamini-Hochberg adjusted $p \leq 5.56 \times 10^{-4}$). Probes identified can be attributed to the genes A) *CAMKK2*, B) *ITGAM*, C) *LPHN1*, D) *PEAR1*, E) Intergenic region, F) *ACSF3*, G) *SMAD6*, H) *MIR146B*, I) Intergenic region.

4.6.2 Pathway Analyses

Pathway analysis utilising the GOMeth package in R was undertaken on the genes associated with 191 CpG sites identified to be differentially methylated. Of all possible terms available, 617 GO terms were found to be enriched in this analysis ($p < 0.05$), however, unsurprisingly none of these were significant after correction for multiple testing. REVIGO was utilised to reduce redundant GO terms and found that the 617 significant GO terms could be reduced to 56 ($p < 0.05$). Cytoscape visualised this network and found that each node was linked to at least six different alternative GO terms (Figure 4.21). Description of GO terms that were particularly enriched once again included signalling pathways. Negative regulation of insulin receptor signalling as well as cell surface receptor signalling were highly enriched. Maternal aggressive behaviour, cell migration and negative regulation of myotube differentiation round out the top five GO terms found.

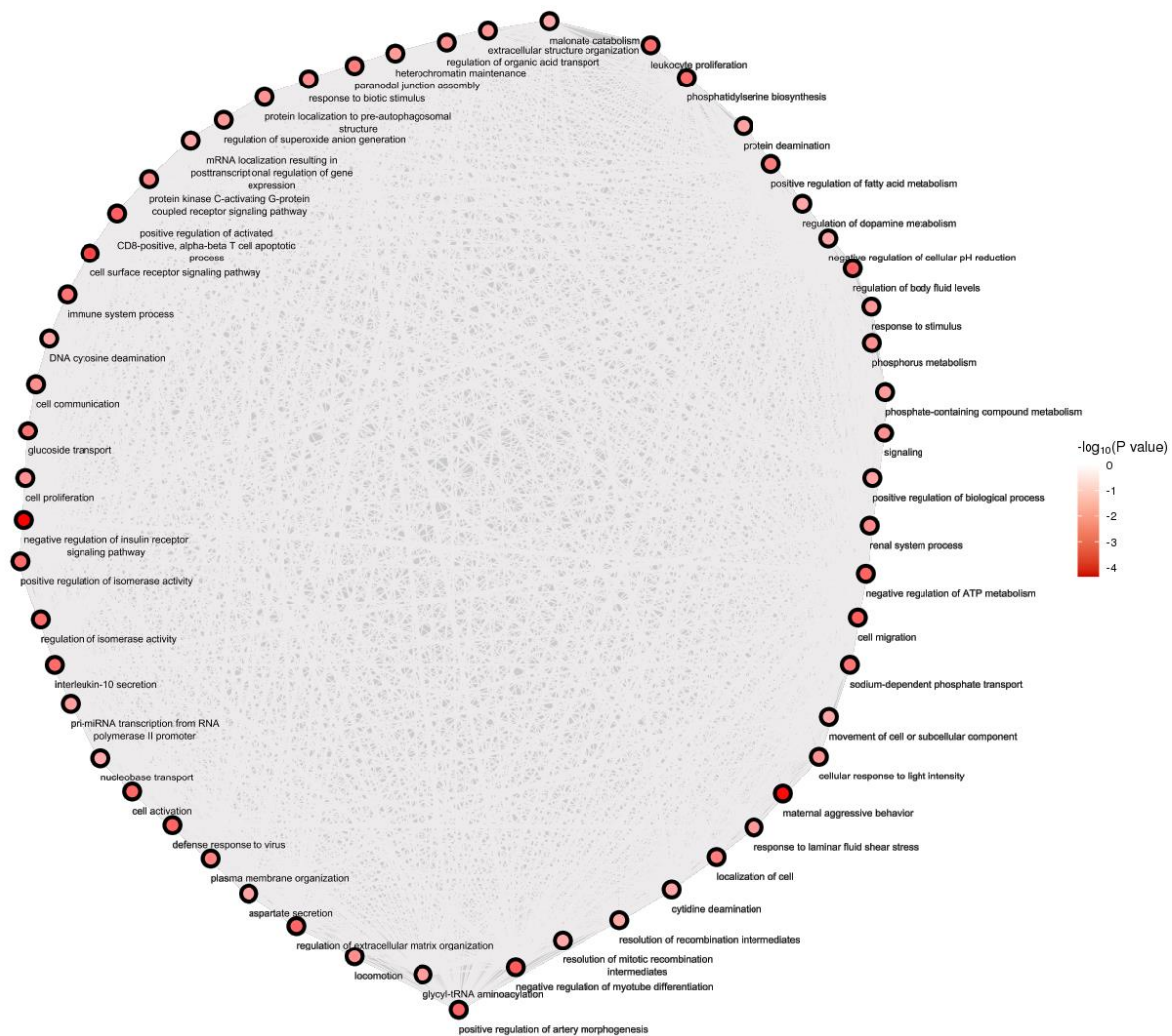


Figure 4.21 REViGO gene ontology network associated with 342 differentially methylated probes identified in a diagnostic vs relapse regression analysis

REVIGO gene ontology reduction identified eight different GO terms associated with the 342 significantly differentially methylated probes derived from regression analysis. $\log_{10}$ p values are shown for each GO term to indicate the significance of the number of probes that are associated with that term appearing by chance. Negative regulation of insulin receptor signalling, and cell migration dominate the GO terms associated with aberrantly methylated probes identified in this analysis

4.7 Discussion

One of the main challenges in the appropriate personalised treatment of paediatric AML is the array of pathogenic pathways that are disrupted in association with disease. While similar to adult AML in many ways, epigenetic alterations are largely specific to each age group, raising the possibility of pathogenic mechanisms that may be unique to paediatric AML.

The overarching aim of this study was to identify aberrant DNAm differences between diagnostic and remission samples of patients who have AML. We hypothesised that there would be a multi-probe DNAm signature associated with AML when compared to matched remission samples. Furthermore, we also hypothesised that of the probes within this signature, several would be in genes potentially involved in disease pathogenesis through either gene silencing or upregulation. Finally, we predicted that patients with a recurrence of AML would have subtle changes to their methylome, revealing novel genes associated with poor outcome.

A genome-wide approach was applied specifically to an Australian discovery cohort followed by a pooled-analysis of the combined Australian and TARGET study cohorts. We sought to identify aberrant DNAm signatures between diagnostic and remission AML samples so that these differences may be used to elucidate functional and targetable pathways for disease pathogenesis and treatment.

4.7.1 Interrogation of Genome-Wide DNA Methylation Reveals Differential Signature Associated with Paediatric Acute Myeloid Leukaemia

It is well known that aberrant DNAm plays an intrinsic role in AML pathogenesis, with several key pathways being implicated in progression or initiation of disease (Toyota, Kopecky et al. 2001). While the incidence of mutation in epigenetic modifying genes is well described, we are only beginning to understand the downstream effects that these mutations are having on both individual genes and the global methylome.

The Cancer Genome Atlas was one of the first comprehensive studies of DNAm in a pan-leukaemic manner. They discovered that DNAm patterning was dependent upon the profile of mutations in epigenetic modifying genes (The Cancer Genome Atlas Research 2013). Expression of these genes was highly concordant for mRNA, miRNA and was associated with methylation at 'open-sea' regions of the genome in acute promyelocytic leukaemia, suggesting that previously unknown regulatory relationships may exist between these molecular groups (The Cancer Genome Atlas Research 2013). More recent studies have continued looking into the effects of DNAm and have discovered that aberrant methylation is often found at regulatory elements of the genome in association with tumour suppressor gene silencing (Qu, Siggens et al. 2016). Furthermore, the functional relevance of these aberrant signatures in adults are beginning to emerge. For example, where DNAm at *HOXA4/5* genes is implicated in therapy resistance (Elias, Azlan et al. 2018).

The extent to which studies on adult AML apply to paediatrics is still largely unknown, primarily due to a lack of research into the field. However, it is known that the incidence of epigenetic modifier gene mutation is quite rare in the paediatric AML population (Ho, Kutny et al. 2011). While several studies have interrogated DNAm in paediatric AML, very few take a pan-leukaemic approach to their studies. One such study found that there are several CpG sites that are differentially methylated in diagnostic samples, compared to day 14, post-treatment remission bone marrow samples (Bolouri, Farrar et al. 2017). WNT pathway regulators such as *DKK1*, *TLE1*, *SFRP4/5*, *SMAD1/5*, *AXIN2*, *WIF1*, *FZD3* and *HES1* were all found to be deleted or aberrantly methylated in diagnostic samples, along with repression of other genes such as *ULBP1/2/3*, totalling 119 differentially methylated and expressed genes.

In our study, regression analysis identified 4219 differentially methylated probes (Benjamini-Hochberg adjusted $p < 0.05$) with donor and remission bone marrow clustering together. Hierarchical clustering also suggests that while remission samples are more closely related to healthy bone marrow than diagnostic samples, there is a distinct difference between these two groups, suggesting that remission samples may be in a state of repair before becoming truly healthy again. However, it is possible that this distinction may be due to the donor marrow being sourced from an adult, cell sorted population as opposed to a paediatric source.

Upon addition of the TARGET cohort, the original DNAm signature of 4219 probes was refined to 1622 probes, with around half of these probes found to be shared between both sample analyses. Furthermore, with the inclusion of around 800 new probes to the refined DNAm signature, we were able to identify new aberrantly methylated genes and regions that were previously undetected. While the revised signature does utilise less probes to separate samples, it was unable to differentiate samples as effectively as the discovery cohort alone. This inability to distinguish sample types may be due to the difference between day 14 and post-treatment bone marrow samples, likely pertaining to induction therapy induced changes.

Minimal evidence for any of the differential methylation previously described in the pan-leukaemic study of Bolouri *et al.* (Bolouri, Farrar et al. 2017) was observed. However, there were several analytical differences between studies, such as differentially methylated genes only being reported when in combination correlated expression differences. Six genes were identified as differentially methylated in both analyses, including *CDKN1C*, *RGS2*, *TALI*, *TEAD4*, *WDR35* and *ZFPM1*. Another study conducted by Lamba *et al.*, identified 50 genes that showed evidence for both differential methylation and mRNA expression change (Lamba, Cao et al. 2018). Of these, we identified four genes; *DNMT3B*, *LSP1*, *SOX15* and *FCGRT*, that are consistently differentially methylated.

A gene previously identified as a differentially methylated gene, *DLUE2* (Morenos 2014), did not replicate in the current study, however this may be due to the enrichment of M5 AML (FAB subtype), suggesting a subtype-specific effect occurring. However, *miR-663* (Yan-Fang, Jian et al. 2013) was also identified as differentially methylated in this study, with three individual CpG sites being identified as significantly differentially methylated. At present however, there is no specific CpG site information from the TCGA adult AML dataset allowing for comparison between studies. While this analysis validates several findings in other studies, we also challenge published findings in other datasets.

4.7.2 Pathways Showing Epigenetic Disruption Associated with Paediatric Acute Myeloid Leukaemia

To further explore the potential cellular pathways impacted by aberrant DNAm identified here, we firstly sought to validate candidate genes that were previously identified as differentially methylated. CpG sites and genes were selected for further testing if; they had multiple probes within a single gene, were found in regulatory regions such as enhancers or promoter regions as identified by the ENCODE consortium (Consortium 2012), had a large effect size and/or had previously been implicated in a pathway associated with AML pathogenesis. The genes fitting these criteria were *WT1*, *DGKG*, *ZSCAN1*, *IRX1* and *REC8*. *ZNF876P* was also originally identified as differentially methylated, however largely failed to replicate, thus it will not be discussed.

WT1

In our study we found 14 validated probes in *WT1* to be significantly hypermethylated in diagnostic samples, with several of these found in enhancer regions, however the majority were found to be within introns of the gene. This is in line with previous studies into adult AML and other forms of cancer that show *WT1* has a hypermethylated profile in the promoter region (Loeb, Evron et al. 2001; Melki and Clark 2002). Furthermore, this research shows that hypermethylation not only of *wt-WT1*, but of alternative transcripts (*AWT1*) also exists which has been shown to be a potent biomarker of disease (El Bordiny, Al-Ghandour et al. 2018). Studies in adults have shown that despite this hypermethylation, *WT1* is still highly expressed and is considered a marker of poor patient outcome (Trka, Kalinova et al. 2002; Yang, Han et al. 2007). Considering this trend mirrors research in the adult field of AML, it is reasonable to assume that expression is similar in paediatric AML, however further studies are required. While this may seem counter-intuitive, previous studies have shown that an anti-sense *WT1* (*WT1-AS*) transcript, sharing a promoter region with an alternative transcript of *WT1* (*AWT1*), is methylation-dependant and found within the 1st intron of *WT1* (Dallosso, Hancock et al. 2007). This transcript positively regulates wild-type *WT1*, therefore aberrant methylation of *WT1-AS* could have downstream effects upon *WT1* expression, ultimately leading to contribution to malignancy (Dallosso, Hancock et al. 2004). Furthermore, *WT1-AS*

has been shown to promote cell proliferation and invasion in other forms of cancer (Du, Zhang et al. 2016). Future studies would do well to research into *WT1-AS* and enhancers regions downstream of *WT1* would be necessary as previous findings suggests this truly regulates *WT1* expression (Furuhata, Murakami et al. 2009).

DGKG

In our study we found that four validated probes within *DGKG* were hypermethylated in diagnostic samples compared to controls, with an average $\Delta\beta$ of 49.8%. Five of these probes are found in an enhancer region of the gene and within a site of open chromatin. This suggests that while *DGKG* is in an open reading frame, it is being suppressed by DNAm in AML. This is the first report of *DGKG* being hypermethylated in paediatric AML, however it is commonly found to be hypermethylated in paediatric acute lymphoblastic leukaemia (Figuerola, Chen et al. 2013; Chatterton, Morenos et al. 2014). Ramos *et al.* found this gene to also be hypermethylated in many different stages of malignant haematopoietic differentiation suggesting that this is an early onset event and may play a role in initiation of leukaemogenesis opposed to progression of disease (Ramos, Ramos et al. 2018). In murine colorectal cancer, hypermethylation of *DGKG* has shown to also be a common finding and could be corrected through 5-azacitidine treatment (Kai, Yamamoto et al. 2017). This supports our discovery of aberrant methylation found in paediatric AML and further solidifies the notion of *DGKG* having therapeutically targetable tumour suppressor properties.

ZSCAN1

The zinc finger and SCAN domain containing 1 (*ZSCAN1*) gene was also identified due to the affected region, number and magnitude of differential methylation of the probes identified in regression analysis. This gene had an average $\Delta\beta$ of 43.2% across three adjacent probes within the promoter region of the gene, including one probe in the top 10 most significantly differentially methylated probes. *ZSCAN1* function is still being uncovered, however it has been suggested that it is involved in transcription factor activity and transcriptional regulation. Furthermore, this gene has been previously identified as differentially methylated and is a cancer-related DMR (Boers, Wang et al. 2016). Studies into adult AML have shown that

differential methylation does exist in this gene (Figuerola, Lugthart et al. 2010; Qu, Lennartsson et al. 2014) and potentially correlates with gene expression (Geyh, Rodriguez-Paredes et al. 2018), although numbers used in this study for expression analysis are low and may have several false positives. This supports the notion that *ZSCAN1* is involved in carcinogenesis, likely in an initiating role. This also supports the idea of zinc finger genes being integral to AML pathogenesis and are likely targets for aberrant DNAm. Further studies would need to be conducted to elucidate the true function of this gene in context of AML and larger sample sizes for correlation between aberrant methylation and expression.

REC8

Meiotic recombination protein (*REC8*) boasted an average $\Delta\beta$ 38.5% across eight probes, primarily localised around the TSS and enhancer region of the genes. Considering the localisation of the probes and that they are within a DNase hypersensitivity region it is likely that DNAm may play a role in maintenance of expression. *REC8* is an interesting target as the primary function of this gene, along with other genes in sister chromatid cohesion complex, however previous studies have shown that *REC8* does have tumour suppressor properties associated with clonogenicity, cell viability and cell cycle progression (Yu, Liang et al. 2017). Previous studies have also identified *REC8* as hypomethylated in leukaemic stem cells, and hypermethylated in blast cells (Jung, Dai et al. 2015). While this does suggest that there is epigenetic regulation involved with *REC8*, the functional effect of aberrant expression, particularly in a temporal context, remains to be discovered in AML.

IRX1

Iroquois Homeobox 1 (*IRX1*) is a tumour suppressor gene commonly found altered in head/neck and gastric cancers (Guo, Liu et al. 2010; Demokan, Chuang et al. 2014). *IRX1* was found to have an average $\Delta\beta$ of 25.9% across five probes creating a singular DMR within a region of DNase hypersensitivity. With several cancers reporting aberrant DNAm suppressing expression of *IRX1* it is likely that this homeobox gene is also being regulated by methylation in paediatric AML (Bennett, Karpenko et al. 2008). More recent studies have shown that *IRX1* induces transcription of *HOX* genes that are normally downstream targets of signalling

pathways such as c-KIT and Wnt in MLL-AF4 ALL (Kühn, Löscher et al. 2016). Conversely, they find that overexpression of *IRX1* is associated with poorer patient outcome. Further studies will need to be conducted to find the functional benefit of having this gene aberrantly expressed in AML.

~

These probes identified could be useful as a biomarker of disease in the paediatric setting, particularly *WT1*, as aforementioned studies have shown its utility in adults already. Unfortunately, individual probes may not be as useful to distinguish disease as several of these probes have large variance in their methylation, which is suggestive of subtype-specific effects (Figueroa, Lugthart et al. 2010; Qu, Lennartsson et al. 2014). These subtype-specific effects will be explored in more detail in the following chapter.

Finally, 191 probes were identified as significantly differentially methylated between diagnostic and matched relapse samples. Restricting results to the top 50 probes allowed for the greatest split between samples in unsupervised analysis. However, clustering showed a large amount of overlap between samples types as is to be expected when comparing between matched cancer samples. Top probes that were identified include *CAMKK2*, *SMAD6* and *miR146B* further evidencing the potential signal cascade and micro-RNA pathways involvement in AML progression (Teng, Racioppi et al. 2011; Wang, Chen et al. 2017). This analysis was fairly underpowered and therefore a massive increase in sample size would be necessary to elucidate the subtle difference between matched diagnostic/relapse samples in future studies.

In summary, we have demonstrated that there is a widespread DNAm disparity between diagnostic and healthy bone marrow samples in the largest paediatric AML cohort pooled-analysis to date. Signatures associated with disease as well as aberrant methylation in multiple genes have been identified and validated in a combined cohort as being potentially epigenetically regulated and serve as biomarkers of disease. Finally, we have also identified several aberrantly methylated probes that may be involved in the progression of disease from original diagnosis to relapse. All of these findings will necessitate further studies to corroborate

correlated expression and potential functional roles in paediatric AML before potentially being utilised as a therapeutic target.

4.7.3 Limitations and Future Directions

The major limitation for this study was the potential confounding factor of cell heterogeneity and lack of blast count data on many samples. This results in average methylation values of cancerous and non-cancerous cells in what is assumed to be a pure cell population. While many samples are purely AML (>90% blast count), missing blast-count data on some diagnostic samples may have impacted findings as the sample may be impure. Furthermore, due to the archival nature of the samples disallowing for cell sorting, downstream *in silico* reference-based, cellular deconvolution proved to be impossible. It may be worth applying a no reference deconvolution such as SVA (Teschendorff, Zhuang et al. 2011) or RefFreeEWAS (Houseman, Kile et al. 2016) in future, however it would be best that future studies have accurate blast count data, and that remission samples are sourced in a viable manner to ensure that either cell counts are available from clinical information or can be obtained through flow-cytometry.

While increases in power due to pooled-analyses is highly beneficial for high dimensionality studies, they bring their own set of potential confounding factors, not apparent in initial analyses on samples from a single site/study. Geographic differences in patient location raise the possibility of variable effects due to different treatment regimens, samples collection approaches and time points, or technical effects due to variation in location/approach to sample processing and methylation data generation. Of note, TARGET cohort patients had their remission samples taken at day 14, as opposed to day 30 for the Australian cohort. Patients were recruited to the TARGET study due to having severe or rare cases of AML, leading to an enrichment of particular subtypes, which potentially causes subtype-specific effects overriding pan-leukaemic results. While these limitations cannot be avoided when combining cohorts, it is important that detailed clinical, sample collection and processing data should be collected where possible to alleviate confounding factors where possible.

Another limitation for this study revolves around the lack of complementary data to DNAm measures. While methylation array data gives insights into genes that may be aberrantly regulated, until expression data is matched to these samples, it is not possible to directly assess functional relevance. Furthermore, previous studies assessing both gene expression and methylation for key genes implicated in AML pathogenesis (such as *WT1*), suggest DNAm alone cannot explain aberrant expression. This therefore suggests that several other forms of epigenetic data may be necessary to truly elucidate functional relevance. This is exemplified by studies conducted by Bernt *et al.* where they showed MLL-AF9 myeloid leukaemia is driven by aberrant H3K79 methylation induced by DOT1L dependence in leukaemic stem cells (Bernt, Zhu et al. 2011). Other studies have shown that aberrant methylation at enhancer regions may be insufficient to alter expression due to other epigenetic marks such as H3K27 acetylation disallowing transcription at the chromatin level. This suggests a complex interplay of epigenetic regulators necessary for transcriptional activation (Qu, Siggins et al. 2016). For future studies, appropriate use of fresh samples would allow for RNA and chromatin to be extracted for gene expression and ChIP analyses, to compliment DNAm array data.

This page is intentionally left blank.

Chapter 5 Defining DNA Methylation
Signatures of Cytogenetically
Distinct Subtypes of Paediatric
Acute Myeloid Leukaemia

5.1 Rationale

Paediatric acute myeloid leukaemia is one of the most heterogeneous cancers (Li, Mason et al. 2016). Characterised by multiple different chromosomal aberrations and mutations that contribute to disease, AML is highly heterogeneous genetically, despite being classified into a relatively small number of cytogenetic subtypes. Few epigenome-wide studies have been undertaken in paediatric AML and thus little is known about the aberrant methylation differences found in general, let alone in a subtype-specific manner. These few studies have been associated with several subtype-specific DNAm signatures (Figueroa, Lugthart et al. 2010; Bolouri, Farrar et al. 2017), potentially of use in further subtyping patients, identifying those with novel fusion genes and informing patient-specific treatment regimens (Milani, Lundmark et al. 2010; Nordlund, Bäcklin et al. 2015; Capper, Jones et al. 2018). Considering the genetic heterogeneity of paediatric AML, and previously described DNAm signatures associated to different subtypes in other cancers, it is plausible that subtype specific DNAm profiles may also exist in childhood AML. Furthermore, these signatures in other cancers have been further developed in a random forest machine-learning environment to predict subtype and even patient outcome (Nordlund, Bäcklin et al. 2015; Capper, Jones et al. 2018). However, little is known about DNAm-based signatures that accurately define subtype in a paediatric cohort beyond a single study conducted by Bolouri *et al.* that identified core-binding factor, *KMT2A* and *IDH1* mutant subgroups. (Bolouri, Farrar et al. 2017). As such, this study aims to explore and build upon DNAm-based signatures of specific paediatric AML patient subtypes.

The inv(16) (or *CBF β -MYH11*) subtype of paediatric AML is ideal for identifying DNAm-based signatures in paediatric AML as aberrant DNAm plays a role in the pathogenesis of disease and is thought to contribute to initiation (Larmonie, Arentsen-Peters et al. 2017; Lamba, Cao et al. 2018). However, few studies have examined, with genome-wide resolution, the contribution of DNAm to inv(16) pathogenesis in paediatric AML. Thus genome-wide profiling of DNAm in the inv(16) subtype presents an opportunity, not only to discover and validate the differential methylation in genes that may contribute to pathogenesis of disease, but to also identify a signature of DNAm that may accurately define inv(16) from other subtypes, independent of genetic analysis.

The overall aim of this study is to identify aberrant methylation associated with cytogenetic subtypes of paediatric AML. We hypothesis that;

- (i) DNAm signatures will differentiate specific subtypes;
- (ii) Unique DNAm signatures can be attributed to the inv(16) subtype, distinguishing them from other forms of paediatric AML, and;
- (iii) Gene-specific differential DNAm can be identified between inv(16) paediatric AML and other subtypes of the disease that may contribute to pathogenesis.

5.2 Study Cohort

This study utilised 171 primary AML patients across multiple cohorts with varying cytogenetic abnormalities and patient outcomes. All patients had samples derived from bone marrow taken at diagnosis and were characterised using fluorescence *in situ* hybridisation (FISH)(Table 3.1). These cohorts studied had a range of cytogenetic anomalies in an attempt to fully explore the link between specific aberrations and DNAm profile. Almost all cytogenetic subtypes were represented with enrichment for those with *KMT2A* and inv(16) rearranged samples. Samples with unknown cytogenetic data were not included in analyses. Overall there was a slight overrepresentation of male patients and patients over three years of age. However, both female and younger patients were represented in accordance with standard epidemiological studies of AML (Puumala, Ross et al. 2013). Clinical characteristics were obtained from the CLARA database for RCH (VIC) cases while other cohort information was received upon receipt of samples from their respective biobanking repositories. All data is correct as of April 2018.

Table 5.1 Clinical characteristics of diagnostic samples included in analyses

Clinical Characteristics	Australian Cohort (%) (n = 82)	TARGET Cohort (%) (n = 89)	Combined Cohort (%) (n=171)
Sex			
Male	49 (59.8%)	45 (50.6%)	94 (55.0%)
Female	33 (40.2%)	44 (49.4%)	77 (45.0%)
Age at diagnosis			
Infants (<3 y/o)	20 (24.4%)	7 (7.9%)	27 (15.8%)
Children (3-15 y/o)	28 (34.1%)	35 (39.3%)	63 (36.8%)
AYA (15+ y/o)	34 (41.5%)	47 (52.8%)	81 (47.4%)
Sample use			
450K	82 (100%)	89 (100%)	171 (100%)
Validation	16	0	16
Outcome[†]			
Good	30 (36.6%)	10 (11.2%)	40 (23.4%)
Poor	43 (52.4%)	79 (88.8%)	122 (71.3%)
Unknown	9 (11.0%)	0	9 (5.3%)
Cytogenetics			
inv(16)/t(16;16)	9 (11.0%)	6 (6.7%)	15 (8.8%)
t(8;21)	2 (2.4%)	7 (7.9%)	9 (5.3%)
<i>KMT2A</i> rearranged	23 (28.0%)	14 (15.7%)	37 (21.6%)
Trisomy 8	5 (6.1%)	<i>'Other'</i>	5 (2.9%)
CN	18 (22.0%)	25 (28.1%)	43 (25.2%)
Complex karyotype *	11 (13.4%)	<i>'Other'</i>	11 (6.4%)
Other [‡]	14 (17.1%)	37 (41.6%)	51 (29.8%)
Sample source			
RCH(VIC)	40 (48.8%)	0	40 (23.4%)
WM	37 (45.1%)	0	37 (21.6%)
RCH(QLD)	5 (6.1%)	0	5 (2.9%)
TARGET AML	0	89 (100%)	89 (52.1%)

Abbreviations: EFS - Event free survival; y/o - Years old; AYA - Adolescent/Young adult; inv - Inversion; t - Translocation; CN - Cytogenetically normal; RCH - Royal Children's Hospital; VIC - Victoria; WM - Westmead; QLD - Queensland; TARGET AML - Therapeutically Applicable Research To Generate Effective Treatments in Acute Myeloid Leukaemia.

* Complex karyotypes are defined by patients having two or more cytogenetic abnormalities or abnormalities that are otherwise unrelated to AML pathogenesis.

[‡] Other cytogenetic abnormalities are defined as Trisomy21, del5/7/9/13/20q, inv(5q), -7, t(5;6), t(6;9), t(7;14), t(15;17), t(16;21). Trisomy 8 and Complex were considered 'Other' by TARGET cohort standards. Genetic abnormalities are mostly undefined in the Australian cohort.

[†] Patients achieved a 5-year EFS however recent data is unavailable to ensure that they are currently still alive. Patients that achieved a 5-year EFS are considered 'Good outcome'. 'Unknown' has not yet achieved 5 years.

5.3 Distinct Methylation Signatures of AML Cytogenetic Subtypes

5.3.1 Defining Patient Subtype through Unsupervised Analysis

Considering paediatric AML is a disease defined by its' genetic heterogeneity, there is great value in being able to accurately subtype patients by finding additional information that helps further improve diagnosis and prognosis. Previously generated DNAm data for Australian-based cases (discovery cohort) and TARGET cohort (pooled-analysis) were included in this analysis and was used to calculate the variance across all probes, independent of patient subtype. The top 10,000 probes with the largest variance were retained for use in subsequent analyses including unsupervised clustering, t-SNE and machine-based learning approaches. Each analysis was conducted with various arbitrary probe-cut-offs (10,000, 1,000, 500, 250, 100, 50) to find the optimal number of probes for distinguishing patient subtype.

Unsupervised Hierarchical Clustering

Optimal separation of samples utilising the top 250 probes in an unsupervised hierarchical clustering analysis, three main branches of cases were defined (Figure 5.1A). The first is characterised primarily by a hypomethylated signature with enrichment of the *KMT2A* subtype (17/23 *KMT2A* defined cases). Several complex and rare ('Other') cases were also found within this cluster. The second cluster is defined by an equal assortment of methylated and unmethylated probes. A small sub-cluster primarily classifies the inv(16) subtype, with all samples clustering together along with a single Trisomy 8 case. The final cluster is defined by a hypermethylated/hypersilenced signature with no true subtype signature. There are several cytogenetically normal cases found which suggests that gene silencing via DNA hypermethylation is a major factor in disease pathogenesis. We also find that 75% of all Trisomy 8 cases also have a hypermethylated profile.

A subsequent analysis was also carried out with inclusion of TARGET sample data. Utilising the top 250 variably methylated probes produced similar results to the discovery analysis whereby samples clustered primarily into four main categories; hypermethylated,

hypomethylated and two sections of variable intermediate methylation (Figure 5.1B). Similarly, the *KMT2A* subtype was defined by hypomethylation while inv(16) samples were intermediately methylated. Again, several cytogenetically normal, rare and trisomy 8 cases were found to be hypermethylated, while others were hypo or intermediately methylated.

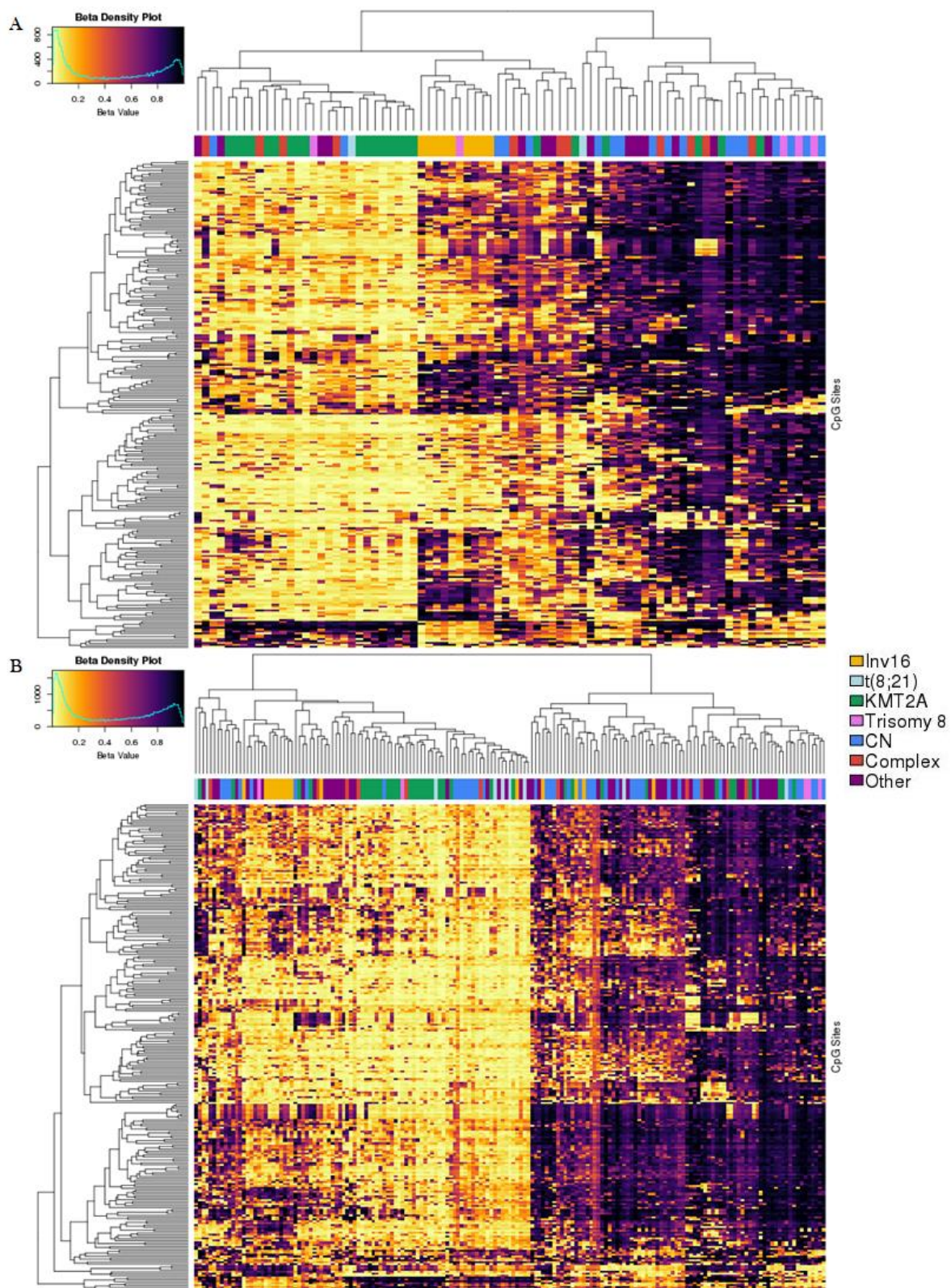


Figure 5.1 Heatmap of beta values across the top 250 most variably methylated probes in paediatric AML: Australian and Pooled-Analyses

Heatmap where the y-axis is representing the top 250 most variably methylated CpG sites while the x-axis shows diagnostic patient samples. A) Heatmap of AUS cohort (n=82). B) Heatmap of pooled cohort (n=171). Hypomethylation is depicted in yellow and a gradual increase to hypermethylation is indicated by purple. Both are ordered by hierarchical clustering as depicted by the dendrograms. This heatmap shows that through variability alone we can distinguish DNA methylation patterns associated with the inv(16) and *KMT2A* subtypes of paediatric AML. While unclear of the link, we also find that several rare, complex and cytogenetically normal samples seem to be hypermethylated. Abbreviations: Inv - Inversion; t - Translocation; CN - Cytogenetically normal.

t-SNE Analysis

Further unsupervised analyses were also undertaken to determine if alternative methods can elucidate different cytogenetic subgroups based solely on methylation data. Utilising t-Distributed Stochastic Neighbour Embedding (t-SNE) along with multivariate t-distributed ellipses we found that the variance of the 250 most variable probes most reliably distinguished specific subtypes from others (Figure 5.2), however most subtypes did not actively cluster together. Despite this, there is much less variation apparent within the inv(16) and slightly less in the *KMT2A* subtypes samples as defined by the smaller ellipses compared to other subtypes. In combination, these two unsupervised approaches suggest a wide degree of methylation variation in most cytogenetic subtypes with the exception of inv(16) and *KMT2A* which may have a more defined epigenetic signature.

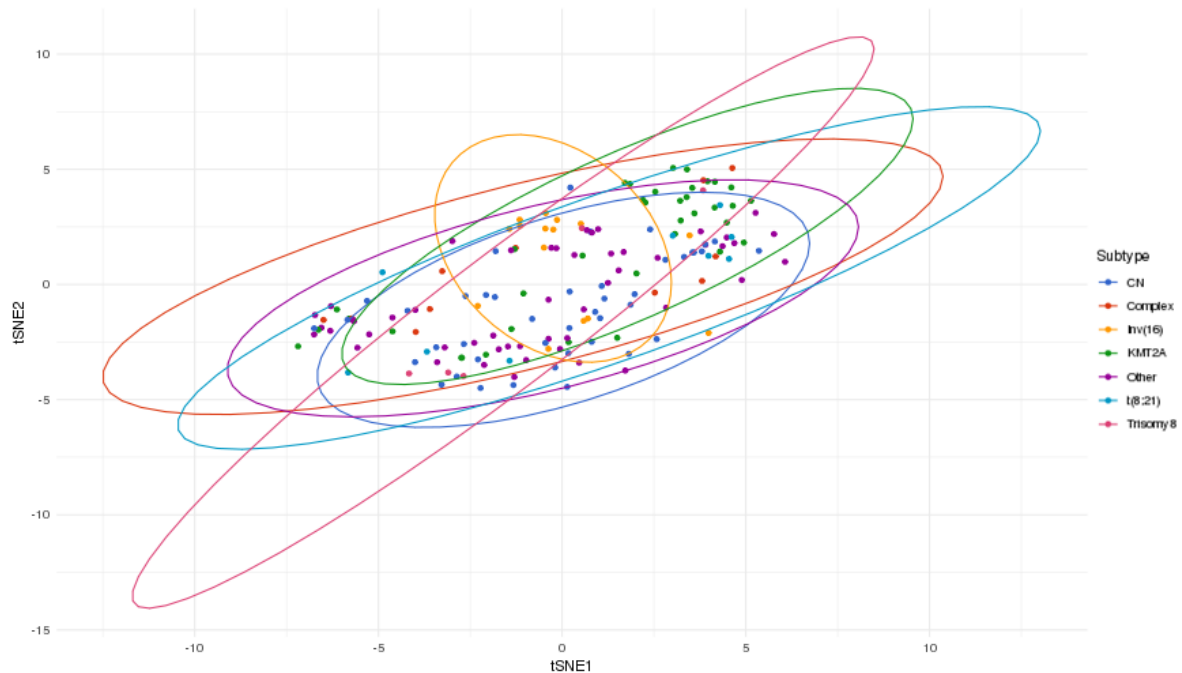


Figure 5.2 t-Distributed Stochastic Neighbour plot of the 250 most variable probes in a pooled cohort to define sample subtype

t-Distributed stochastic neighbour plot showing all samples divided by patient subtype (n=171). The top 250 most variably methylated probes were used in an attempt to define patient subtype. Through multivariate t-distributed ellipses for each group it was clear that overall methylation at these probes had little utility to distinguish specific subtypes, with the possible exception of inv(16) and *KMT2A* samples.

Machine-Based Learning Approach

As a novel approach to uncover the optimal probes to distinguish patient subtype, we employed the random forest machine-based learning algorithm, like other studies conducted by Capper *et al* and Nordlund *et al*, along with the pre-defined 10,000 most variably methylated probes (Nordlund, Bäcklin *et al.* 2015; Capper, Jones *et al.* 2018). Feature importance, a ranking of probes based on utility to the machine-based learning (MBL) model, was initially determined by running all 10,000 probes through a random forest analysis. Initial MBL analysis failed to predict patient subtype due to high levels of feature redundancy. Utilising ranked feature importance (varImp function in R), further analyses were undertaken using the aforementioned cut-offs. We identified the top 100 probes as being the ideal cut-off for identification and prediction of patient subtype (Figure 5.3). Of these 100 probes, only two were previously identified in unsupervised analyses to identify patient subtype (cg18559579 and cg07435445).

In order to best predict patient subtype these top 100 probes, as previously defined by variable importance, were used as the pool of probes for a new random forest model. This random forest model was partly able to predict inv(16) and *KMT2A* subtypes with a balanced accuracy (harmonic mean of precision and recall) of 68.3% and 72.5% respectively (Table 5.2). The model was also somewhat effective at predicting cytogenetically normal and ‘Other’ subtypes with a 63.0% and 59.7% balanced accuracy respectively. Complex, t(8;21) and Trisomy 8 cases could not be predicted using the model. While these results do suggest that there is some utility to using DNAm as a biomarker of patient subtype, sample numbers and overall accuracy would need to be vastly improved before this could ever be truly translated into a test that could identify, not only novel translocations, but standard ones too.

Table 5.2 Final model evaluation of random forest algorithm using 5-fold, 3-times repeated cross validation to predict paediatric AML subtype at diagnosis (n=171)

Metric	CN	Inv(16)	KMT2A	t(8;21)	Trisomy 8	Complex	Other
Precision	0.437	0.739	0.573	0	0	0	0.384
Recall	0.457	0.3778	0.568	0	0	0	0.607
Balanced Accuracy	0.629	0.682	0.725	0.497	0.5	0.5	0.597

Precision: Positive predictive value; Recall: True positive rate (sensitivity); Balanced Accuracy: True accuracy accounting for imbalanced sample set

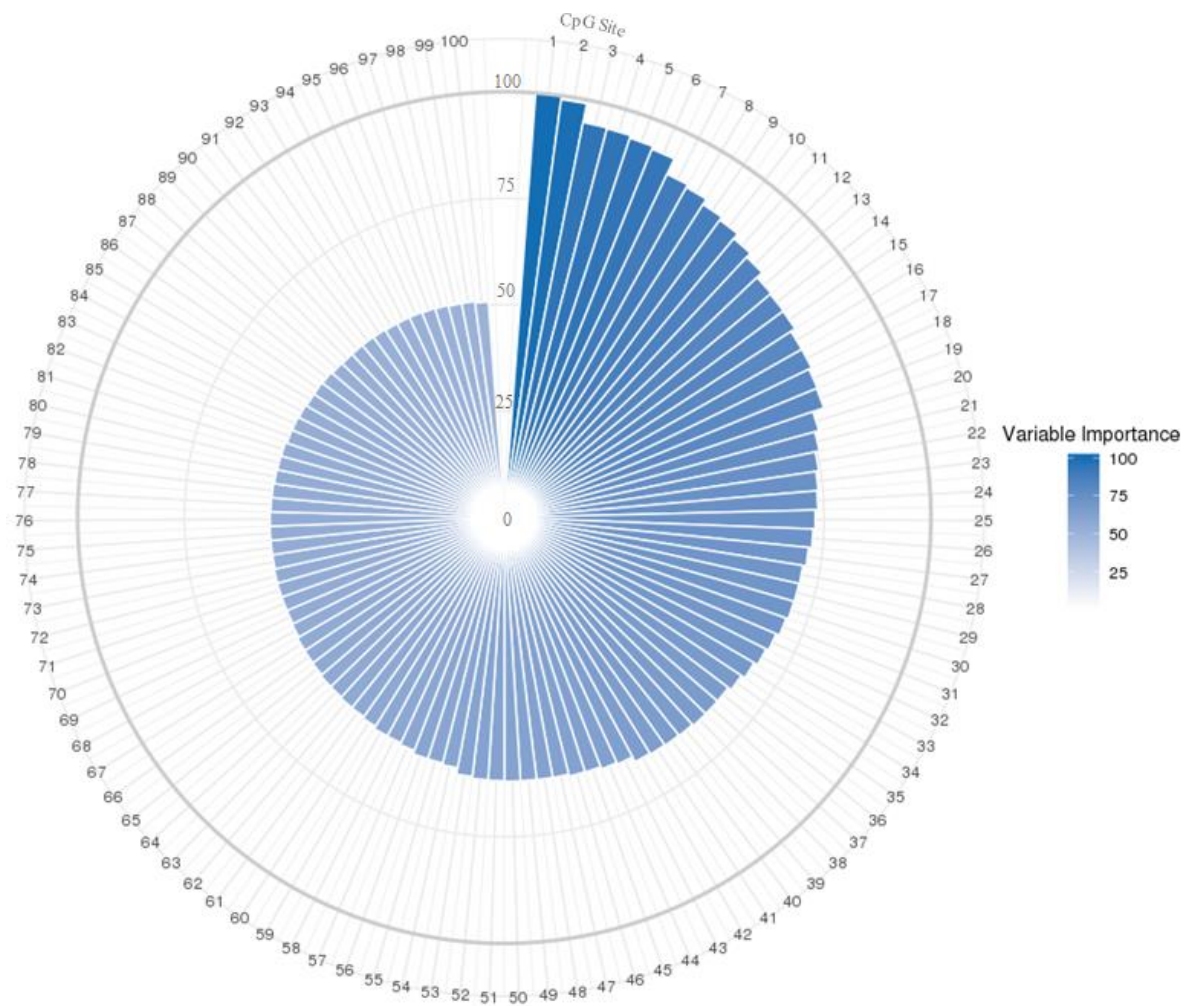


Figure 5.3 Variable importance plot of top 100 probes contributing to random forest algorithm determining paediatric AML patient subtype

Utilising a random forest algorithm using the top 10,000 most variably methylated probes to predict patient subtype variable importance was calculated using the varImp function in the caret package in R. This calculation is defined by the sum of decreased error to a model divided by the highest variable importance (thus the most important probe is always 100). Probes were ranked and the top 100 probes were visualised. These top probes were taken as the new pool of probes for the next random forest to remove redundancy within the data.

Utilising these same 100 probes that were previously determined to be the most important to the random forest model's ability to predict subtype, we undertook hierarchical clustering analysis to determine if this could better illustrate patient subtype prediction when compared to using the most variable probes (Figure 5.1). As with previous analyses, inv(16) and *KMT2A* subtypes once again clustered with the first major split in hierarchical clustering containing almost all the *KMT2A* samples alone, based primarily on hypomethylated probes (Figure 5.4). Inv(16) samples also clustered together however, in contrast to the previous hierarchical

approach based on probe variability, this appears due to a combination of both hypo- and hypermethylated probes. This suggests that it is likely a balance of over- and under expressed genes contributing to the inv(16) pathogenesis and warrants further investigation.

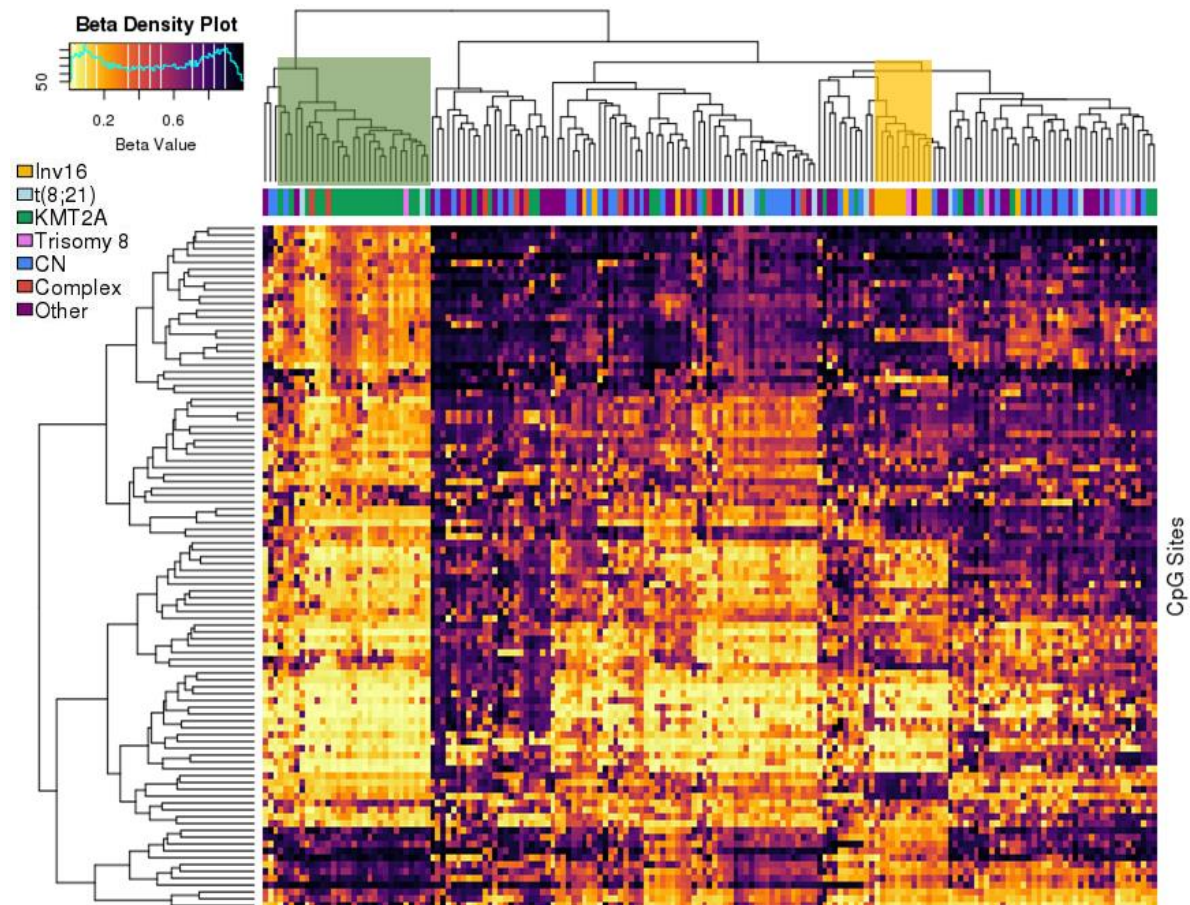


Figure 5.4 Heatmap of beta values across the top 100 probes as defined by variable importance in a random forest algorithm in paediatric AML

Heatmap of the top 100 methylation probes (y-axis) as defined by variable importance in a random forest model. Patient samples of the combined paediatric AML cohort (n=171) are separated on the x-axis. Both are ordered by hierarchical clustering as depicted by the dendrograms. As with previous analysis, only inv(16) and *KMT2A* subtypes show any degree of clustering according to methylation profile.

5.4 Differential DNA Methylation Associated with Chromosome 16 Rearranged Paediatric AML

5.4.1 Principal Components Analysis

Unsupervised clustering through principal components analysis (PCA) was undertaken using methylation data for all diagnostic samples in the Australian AML cohort (n=82). Diagnostic samples were separated into two groups, namely inv(16) (n=10) and all other subtypes (n=72). Combination of principal components 1 and 2, accounting for 19.6% of the variation found between these two groups shows little separation of inv16 from other subtypes (Figure 5.5). Most inv(16) cases tend to cluster together in the bottom left quadrant of Figure 5.5 suggesting a similar DNAm profile, potentially useful in elucidating DNAm unique to the inv(16) subtype.

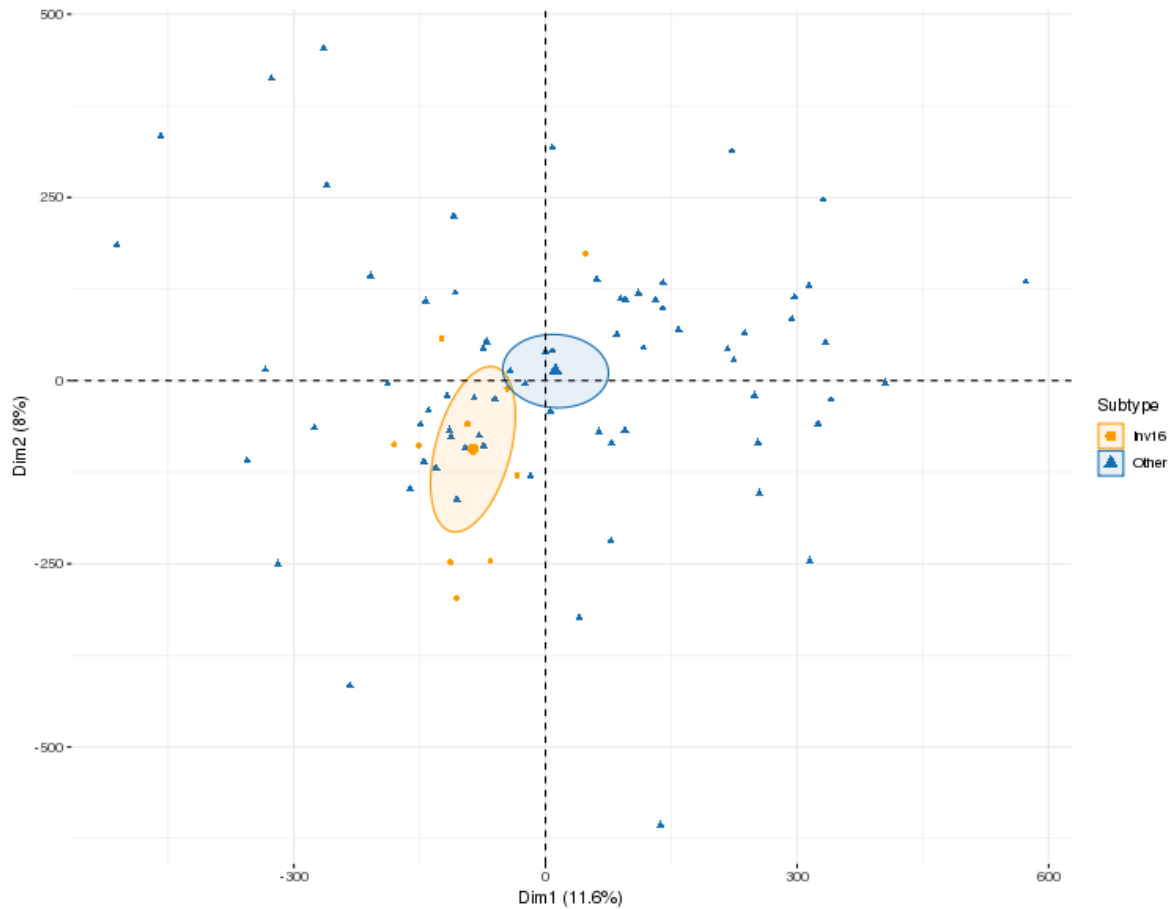


Figure 5.5 Unsupervised principal components analysis plot defining inv(16) paediatric AML from other subtypes in diagnostic bone marrow samples

Methylation beta values of 373,253 probes from the HM450k across 82 samples (inv(16) = 10, Other = 72) were used in a PCA to determine major sources of variation in the dataset. The variation in PC1 (accounting for 11.6% total variation) can be partially attributed to differences between inv(16) AML and other forms of the disease. Variation in PC2 (accounting for 8% total variation) is largely unexplained in this analysis. The larger circle and triangle represent their respective group means and ellipses are 95% confidence interval.

5.4.2 Least Squares Regression Analysis

Least squares regression analysis was undertaken on all methylation data to identify significant differential methylation specifically in the inv(16) subtype relative to all other subtypes. A total of 12,003 differentially methylated probes (DMPs) were identified using this approach following adjustment for multiple testing (Benjamini-Hochberg adjusted $p < 0.05$). Utilising all significant DMPs and applying unsupervised hierarchical clustering showed little ability to distinguish inv(16) samples from other AML subtypes. However, paring this down to the top 100 DMPs revealed a clear DNAm signature specific to the inv(16) subtype (Figure 5.6). This

signature is dominated primarily by hypomethylation however there are a few probes that are hypermethylated in comparison to other subtypes. There is an uncommon uniformity of DNAm in all other subtypes further suggesting that the mechanism underlying differential methylation of these probes is unique to inv(16).

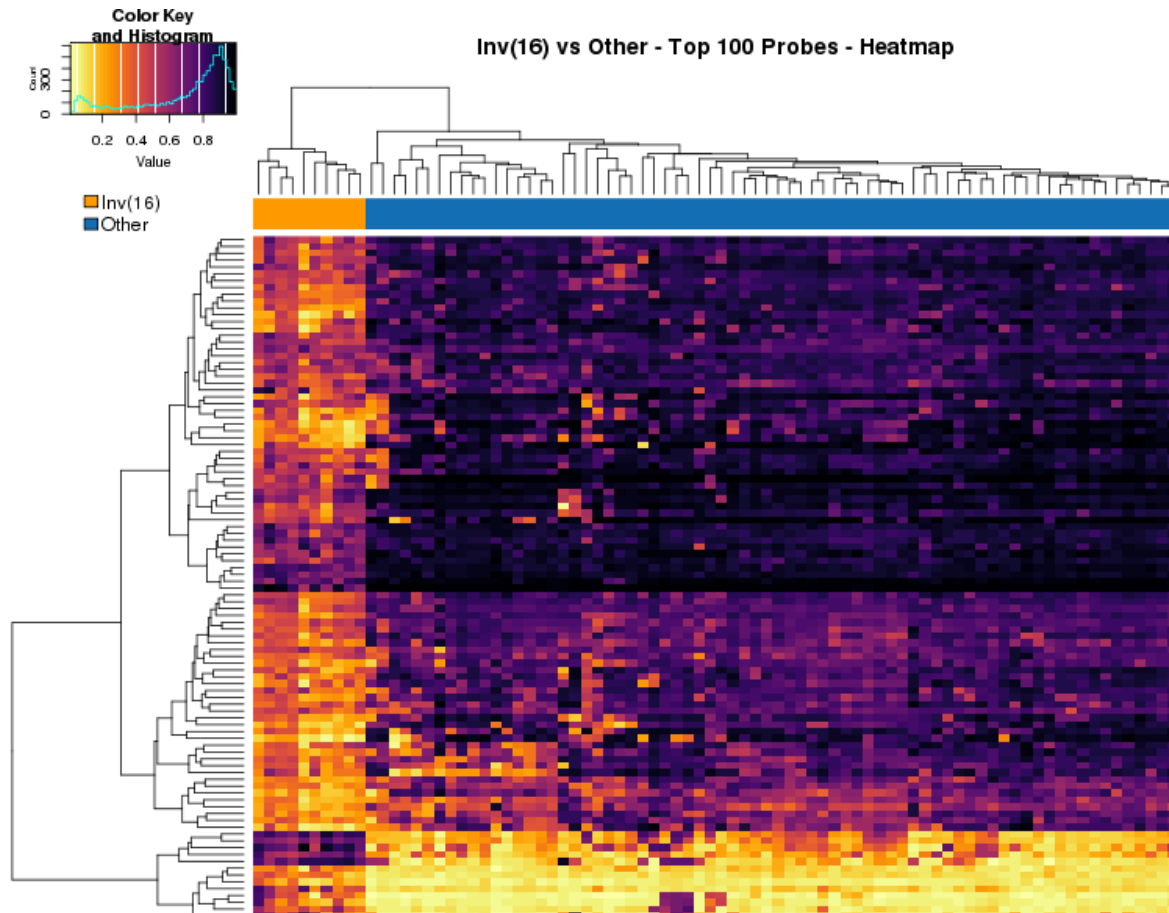


Figure 5.6 Heatmap of beta values across the top 100 differentially methylated probes associated with the inv(16) subtype of paediatric AML

Heatmap where the y-axis is representing the top 100 differentially methylated probes while the x-axis shows 82 patient samples (inv(16) = 10, other subtypes = 72) where both are ordered by hierarchical clustering as depicted by the dendrograms. This heatmap shows that DNA methylation can accurately distinguish the inv(16) subtype of paediatric AML from all other forms of the disease, primarily driven by a hypomethylation event unique to inv(16) samples.

Pathway analysis was undertaken on all 12,003 probes identified to be differentially methylated between inv(16) cases and all other subtypes of paediatric AML by regression analysis. Utilising the GOMeth package we identified 741 gene ontology terms that were significantly overrepresented. This was refined using REVIGO which summarises and removes redundant GO terms. This narrowed the list of terms down to 51 (Figure 5.7). The top five processes identified include: anatomical structural development, cell surface receptor signalling, cell communication, signalling and development processes. The themes identified primarily revolve around typical hallmarks of cancer such as cell signalling and localisation.

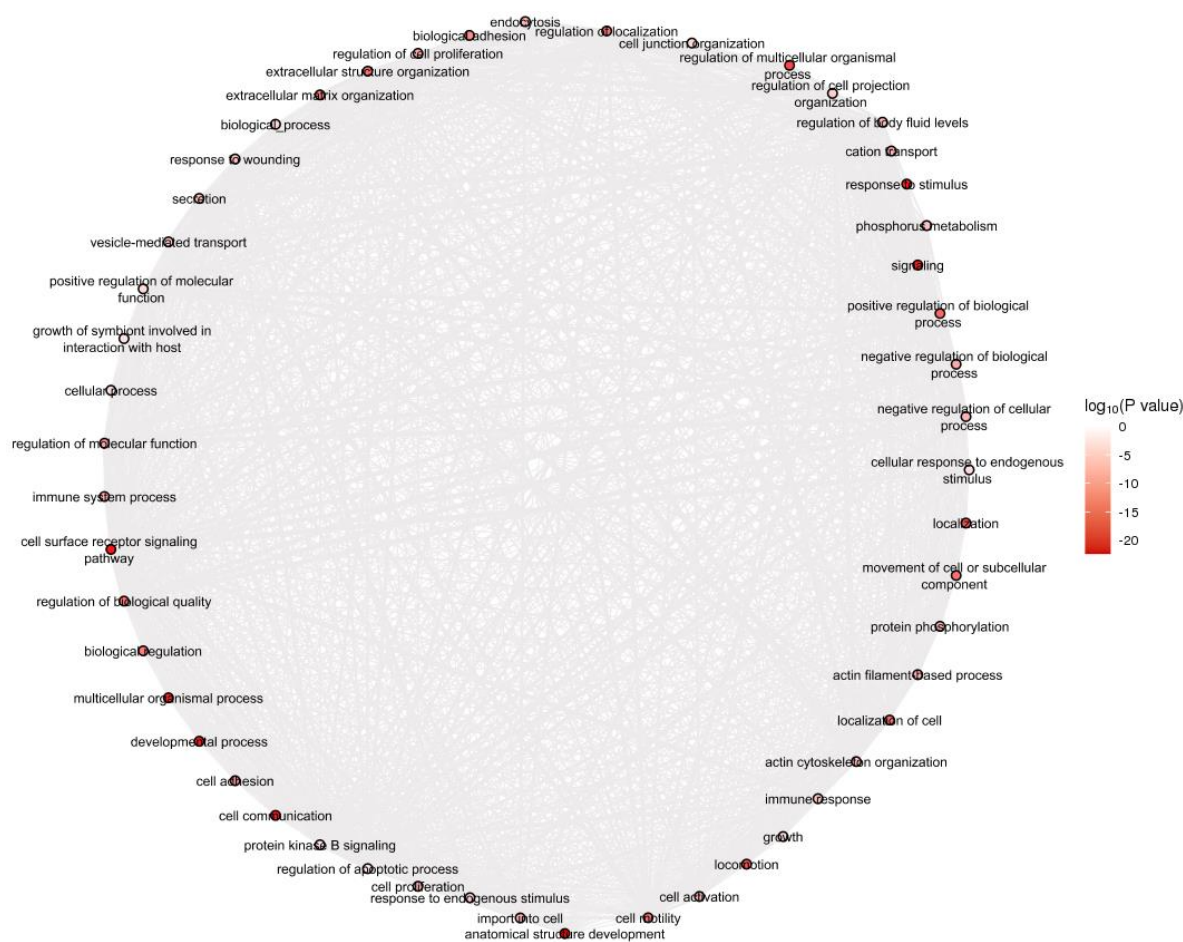


Figure 5.7 REVIGO gene ontology network associated with 12003 differentially methylated probes

REVIGO gene ontology reduction identified 51 different GO terms associated with the 12003 significantly differentially methylated probes derived from regression analysis. $\log_{10}$ p values are shown for each GO term to indicate the significance of the number of probes that are associated with that term appearing by chance.

5.4.3 Probe and Gene Specific Differential DNA Methylation Associated with Chromosome 16 Rearranged Paediatric AML

To further explore the DNAm signature specific to inv(16), we interrogated individual probe and gene methylation differences for potential functional effects on gene expression and AML pathogenesis. Criteria for focusing on specific regions included: (i) the number of DMPs within the a gene; (ii) potential for functional effect based on genomic location; (iii) the effect size of observed $\Delta\beta$ with larger changes being preferable, and; (iv) any previous evidence for a gene in cancer pathogenesis. Genes that met all these criteria included Meningioma 1 (*MNI*) and Meis Homeobox 1 (*MEIS1*).

Some of the most significant differentially methylated probes were found within *MNI*. Disruption of this gene is associated with meningioma (Lekanne, Riegman et al. 1995). The top DMPs identified here included cg00514241, cg00187327, cg06004017, cg05061886 and were all found within the top 20 significant DMPs (Benjamini-Hochberg adjusted $p < 1.0 \times 10^{-10}$). Overall, ten probes in *MNI* were identified as significantly differentially methylated with an average $\Delta\beta$ of 37.9% (Figure 5.8A). Seven out of the ten were within the first exon of the gene and all significant probes were adjacent, suggesting a wide-spread demethylation event in inv(16) samples.

The gene with the most significant DMPs following regression analysis was *MEIS1*. This gene is considered crucial to normal development (as most HOX genes are) and has previously been implicated in adult AML disease progression (Lawrence and Largman 2005; Wang and Kamps 2007). Furthermore *MEIS1* has been implicated in haematopoiesis and works as a cofactor with *HOXA7* and *HOXA9* for the induction of myeloid leukaemia (Thorsteinsdottir, Kroon et al. 2001). Overall, 21 individual probes were identified as differentially methylated between inv(16) and other subtypes of AML. All probes were hypermethylated in inv(16) samples with an average $\Delta\beta$ of 48.1% between groups (Figure 5.8B, Benjamini-Hochberg adjusted $p < 0.05$). Most of the significant probes were found within the gene body, 3 within exons and 18 within introns. Of the 21 probes, 19 were adjacent in the genome, with five located within a known enhancer region.

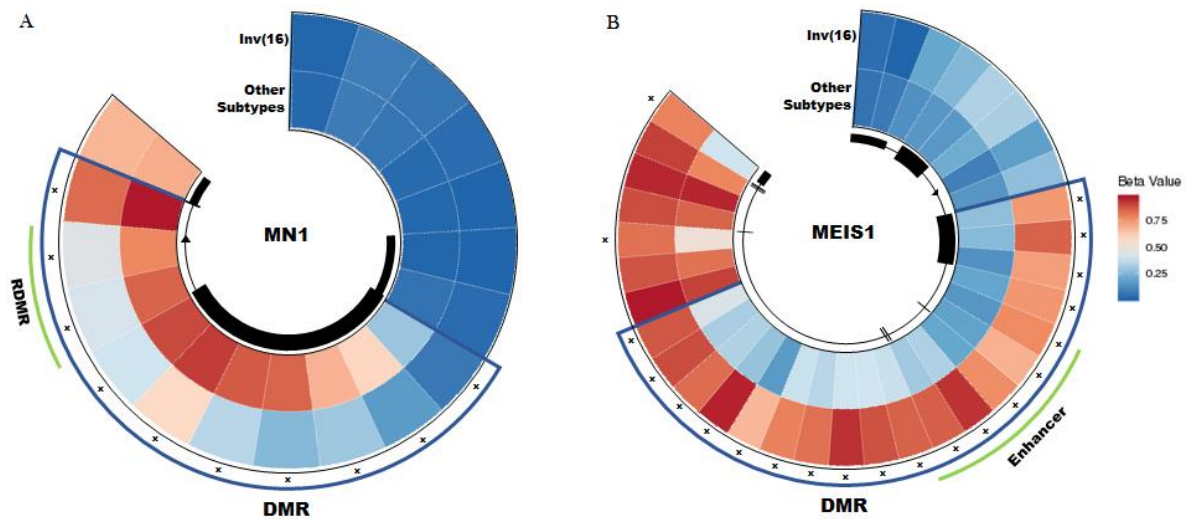


Figure 5.8 Circular heatmaps of differential DNA methylation within genes of interest in *inv(16)* relative to all other subtypes of paediatric AML

Circular heatmaps show mean DNA methylation at each probe that is found in the gene or within 1500 base pairs of the gene's untranslated regions ($n=82$). Inner circles identify the relative position of any given probe in the gene (not to scale). Mean methylation of *inv(16)* (outer circle) and all other subtypes (middle circle) is shown. Completely unmethylated probes are represented by dark blue, while complete methylation is represented by dark red. Significantly differentially methylated probes are indicated by asterisk (Benjamini-Hochberg adjusted $p<0.05$, Student's t -test). Differentially methylated regions (DMR), as indicated by blue wedges, span adjacent significantly differentially methylated probes within 1000 base pairs of each other. Enhancer regions or regions of DNase hypersensitivity found within DMRs are indicated by the outer green line. **A)** Methylation levels at 18 probes. Ten probes were identified to be significantly differentially methylated across a single DMR with an average $\Delta\beta$ of 37.9%. **B)** Methylation levels at 34 *MEIS1* probes. Overall, 21 probes were identified to be differentially methylated across one DMR with an average $\Delta\beta$ of 48.1%.

Figure 5.9 shows the nine most significant probes associated with inv(16) subtype. These probes were attributed to *ULK1* (Benjamini-Hochberg adjusted $p=2.75 \times 10^{-18}$, Figure 5.9A), *MNI* (Benjamini-Hochberg adjusted $p=2.97 \times 10^{-13}$, Figure 5.9B), *CLIP3* (Benjamini-Hochberg adjusted $p=2.97 \times 10^{-13}$, Figure 5.9C), *SLC27A3* (Benjamini-Hochberg adjusted $p=2.97 \times 10^{-13}$, Figure 5.9D), *GEMIN7* (Benjamini-Hochberg adjusted $p=6.49 \times 10^{-13}$, Figure 5.9E), *KIRREL* (Benjamini-Hochberg adjusted $p=1.06 \times 10^{-12}$, Figure 5.9F), *MNI* (Benjamini-Hochberg adjusted $p=1.06 \times 10^{-12}$, Figure 5.9G), *KLHL6* (Benjamini-Hochberg adjusted $p=3.29 \times 10^{-12}$, Figure 5.9H) & *CLIP3* (Benjamini-Hochberg adjusted $p=1.04 \times 10^{-11}$, Figure 5.9I). Both *MNI* and *CLIP3* were represented multiple times in the top 9 probes. Findings were marginally inflated ($\lambda = 1.38$), again suggesting the need for validation, as discussed in section 5.5.1.

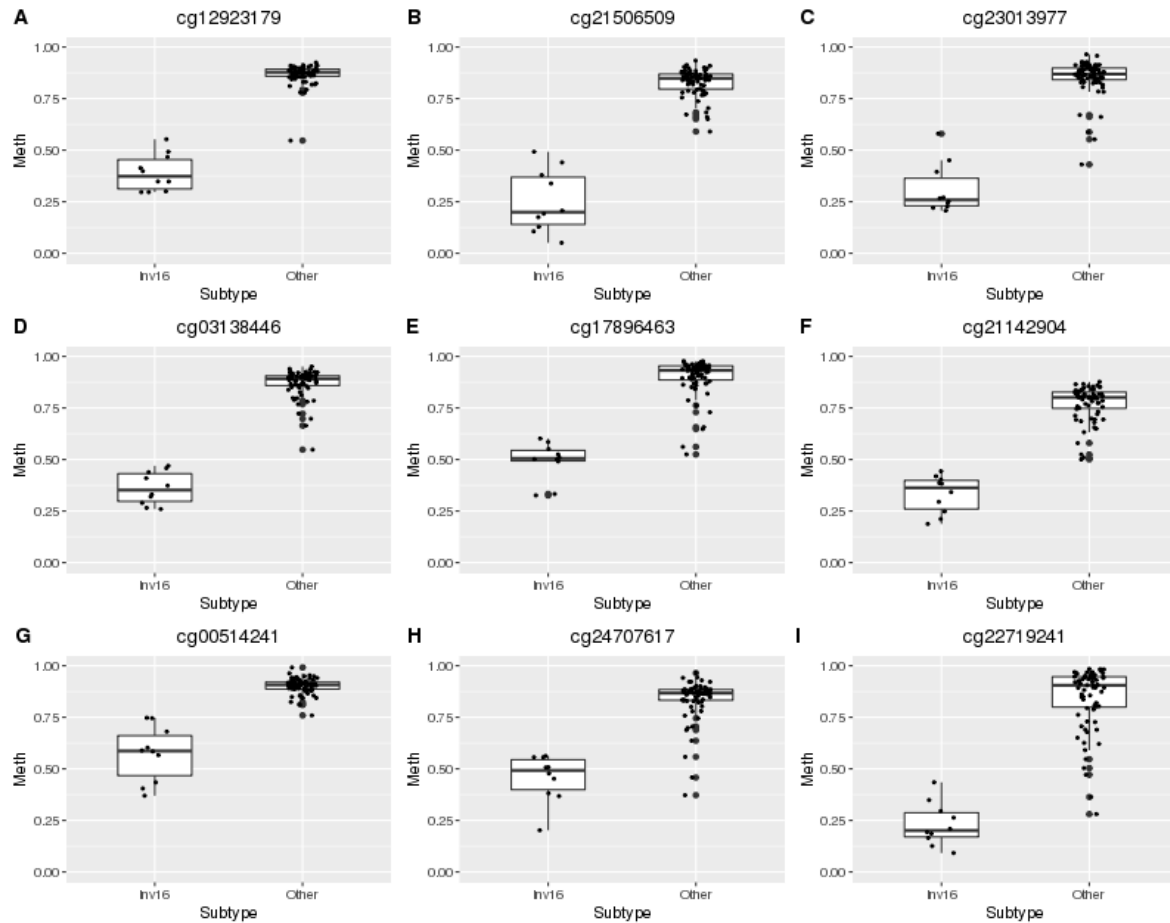


Figure 5.9 Boxplots of the most significantly differentially methylated probes between inv(16) and other subtype diagnostic bone marrow samples

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between patient bone marrow biopsy time points (n=82). The median is represented by the line in the middle of the boxplot while outliers lie beyond the whiskers. Probes were identified as the most statistically significant DMPs (Benjamini-Hochberg adjusted $p \leq 1.04 \times 10^{-11}$). Probes identified can be attributed to the genes A) *ULK1*, B) *MNI*, C) *CLIP3*, D) *SLC27A3*, E) *GEMIN7*, F) *KIRREL*, G) *MNI*, H) *KLHL6* & I) *CLIP3*.

5.5 Locus Specific Analysis of Genes Implicated in Chromosome 16 Rearranged AML: *MEIS1* & *MN1*

5.5.1 Validation and Replication of Methylation between HM450k Array and SEQUENOM Assayed CpG Sites

Utilising MALDI-TOF MS, region specific methylation analysis was undertaken on diagnostic samples to validate and replicate original findings of regression analysis. Assays covering a region within *MEIS1* and *MN1* were designed to test multiple HM450K CpG sites previously identified as significantly differentially methylated between inv(16) and all other subtypes. Each assay also allowed for the interrogation of multiple new CpG sites that were not tested on the HM450K, allowing for a more detailed view of methylation variation across the region of interest.

MEIS1

The MALDI-TOF MS assay designed to interrogate methylation in *MEIS1* spanned two HM450K CpG sites, namely cg04288999 and cg05877497 which were captured within SEQUENOM analytical units as CpG 5.6 and CpG 11.12 respectively. Original findings showed a large and consistent disparity of methylation at both CpG sites (Figure 5.10A/D, $p < 0.0001$, Student's t-test) between inv(16) and other subtypes. These findings were replicated in a separate cohort, showing similar results at both CpG sites (Figure 5.10C/F, $p < 0.0001$, Student's t-test). Methylation was highly correlated between platforms ($n=14$, Figure 5.10B/E, $R > 0.93$, Pearson's correlation coefficient). This suggests that not only can these results be validated, but there is a high level of correlation of methylation between adjacent CpG sites assayed.

Within *MEIS1* a total of 24 CpG sites were assayed using SEQUENOM analysis, however after quality control (See Section 2.5.7), data for only 12 CpG sites remained. A total of 9 CpG units were assayed, encompassing all 12 CpG sites, as several CpG units contained multiple CpG sites within single fragments. All 9 CpG units assayed were significantly differentially methylated between inv(16) and other subtypes of AML using this approach, with an average

$\Delta\beta$ of 67.1% (Figure 5.11, $p < 0.05$, Student's t-test). We identified a 19% increase in methylation in this region compared to the average significant DMPs defined by HM450K analysis. This suggests that this region within *MEIS1* could be a part of a larger DMR within the area. Several CpG sites in the 'Other' subtyping group showed variable levels of methylation, suggesting that while inv(16) samples have consistent hypermethylation, these methylation levels may not be entirely unique to this subtype.

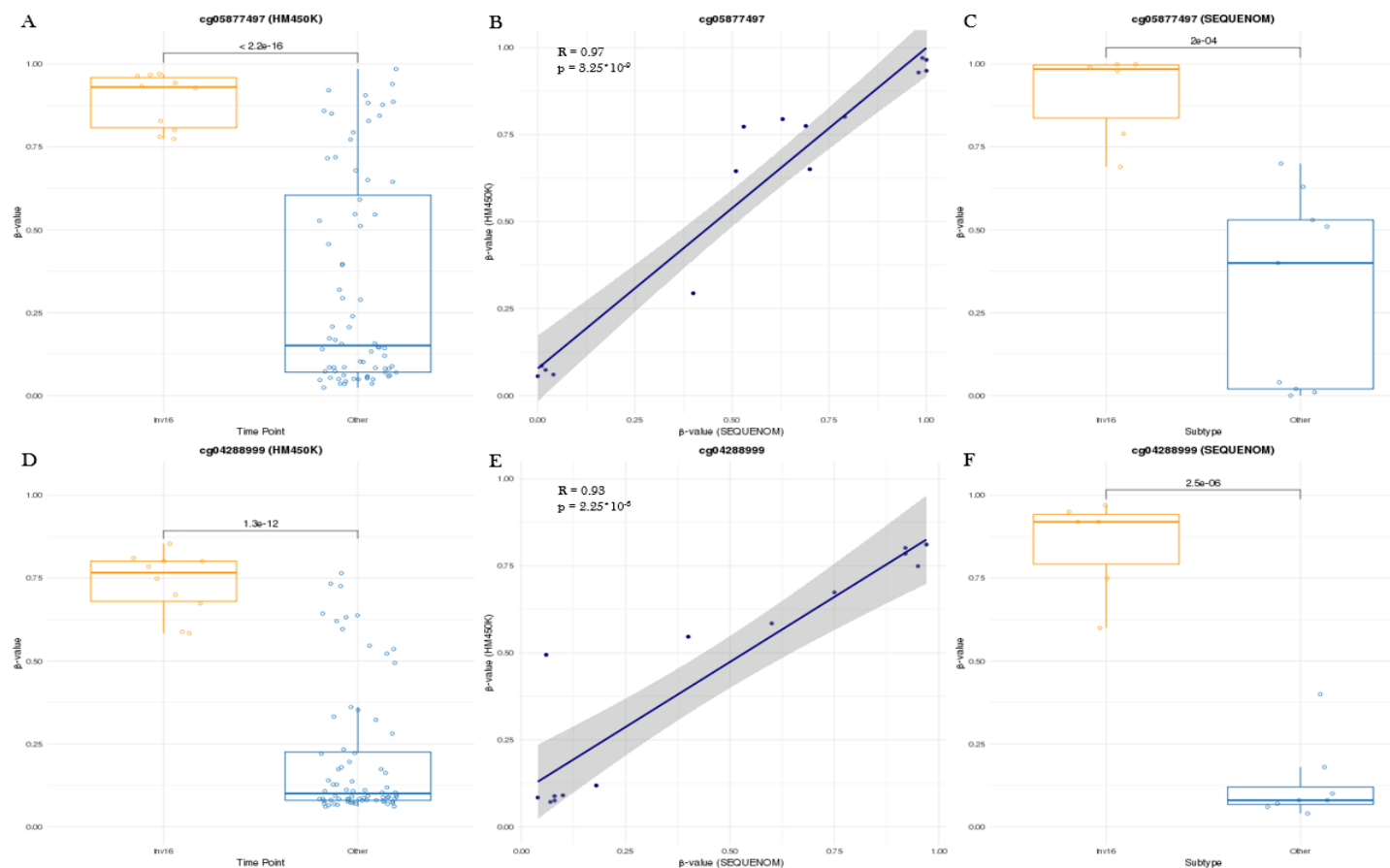


Figure 5.10 Validation of HM450K methylation using MALDI-TOF MS – *MEIS1*

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between inv(16) and all other subtypes of paediatric AML. Outliers lay beyond the whiskers of each boxplot. Scatterplot show the original HM450K methylation on the y-axis and the validation (SEQUENOM) methylation on the x-axis. Line of best fit ($\pm 95\%$ CI) has been added and statistical analysis was undertaken on all samples (Pearson's R). Figures A/B/C show a boxplot of HM450K methylation between inv(16) and other AML subtypes, a scatterplot and correlation between HM450K/SEQUENOM validation samples, and a SEQUENOM derived methylation boxplot between previously mentioned subtypes at cg05877497 within *MEIS1*. Figures D/E/F are as above for cg04288999. Both CpG sites were shown to be significantly differentially methylated between inv(16) and other subtypes of AML in both the HM450K analysis (n=82) and in validation (n=16, $p < 0.001$, Student's t-test). Both sites were highly correlated between HM450K β -value and validated SEQUENOM derived methylation value (n=14, $R > 0.9$ Pearson's correlation coefficient).

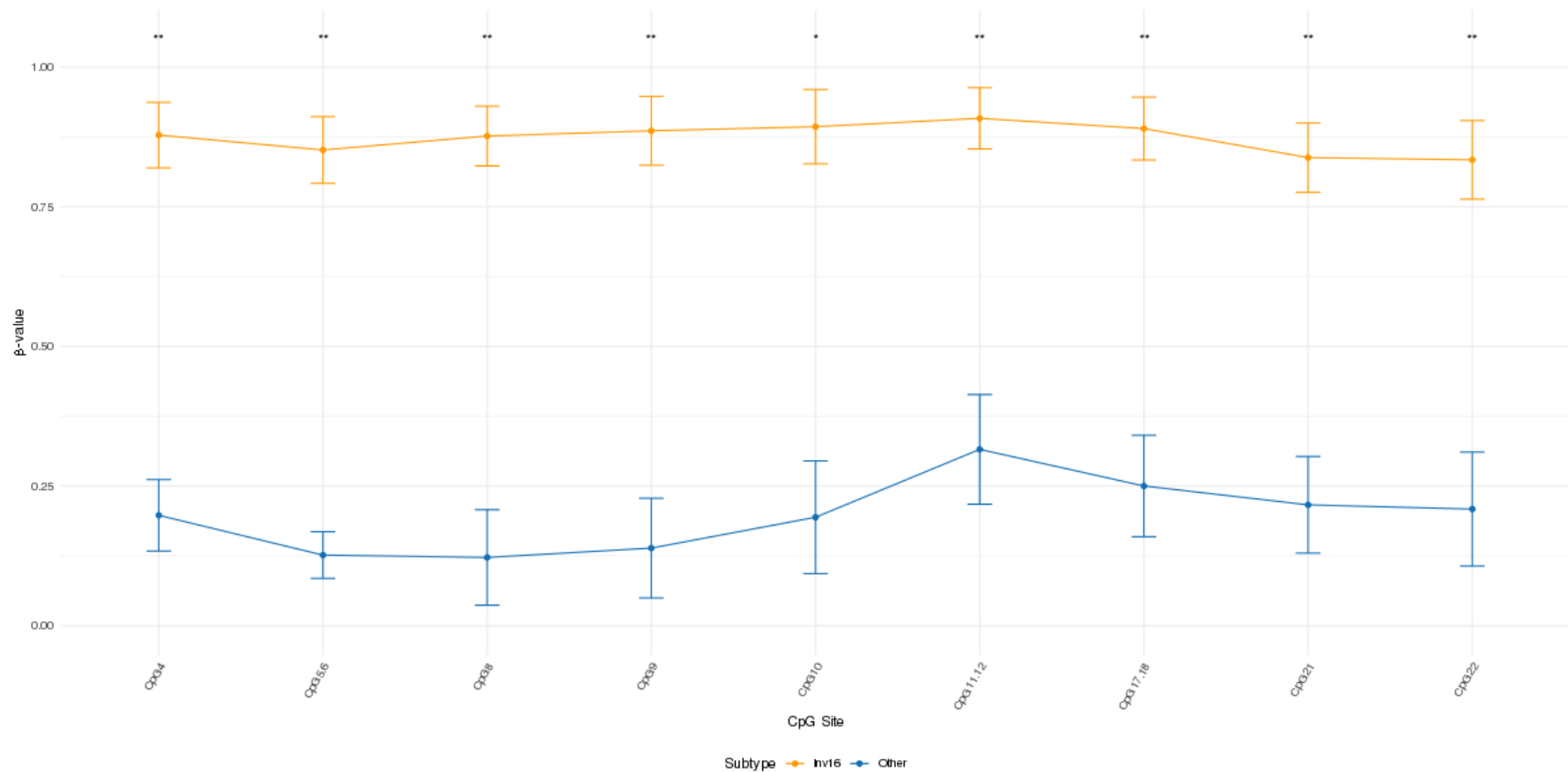


Figure 5.11 Methylation of all CpG units measured by MALDI-TOF MS – *MEIS1*

Line graph denoting the level of methylation of each CpG site assayed. Groups are split between diagnostic samples of inv(16) and all other AML subtype samples and Student's t-test was used to test for significant differences between groups. Significance is shown as; ns= not significant, *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$. All points are stated as the mean of the group with SEM for error bars. Methylation in *MEIS1* (n= 16) across 10 CpG sites. All sites were significantly differentially methylated ($p < 0.05$, Student's t-test).

MNI

MNI was also previously identified as a gene of interest that may have expression influenced by DNAm. The assay designed to interrogate *MNI* also replicated two individual CpG sites found on HM450K, namely cg12872311 and cg02919572, which relate to CpG1.2 and CpG17 respectively. HM450K β -values at both CpG sites were found to be significantly differentially methylated between inv(16) and other subtypes of paediatric AML (Figure 5.12A/D, $p < 0.0001$, Student's t-test). After quality control measures were taken, we found a similar methylation level existed in our replication cohort for both CpG sites (Figure 5.12C/F, $p < 0.001$, Student's t-test), validating our previous findings. Common samples between both analyses showed HM450K β -value and SEQUENOM derived methylation values were highly correlated at both sites (Figure 5.12B/E, $R > 0.94$, Pearson's correlation coefficient).

For *MNI*, a total of 19 CpG sites were assayed, however after quality control only 15 CpG sites, encompassed by 9 CpG units could be utilised. Our data shows that all CpG units were significantly differentially methylated with an average $\Delta\beta$ of 60.2% between inv(16) and other subtypes of AML (Figure 5.13, $p < 0.05$, Student's t-test). This shows a 22.7% decrease in methylation in this particular region compared to the average significantly different methylation of *MNI* identified by all sites but CpG18.19 showed marked hypomethylation suggesting that *MNI* may be upregulated in the inv(16) subtype of AML.

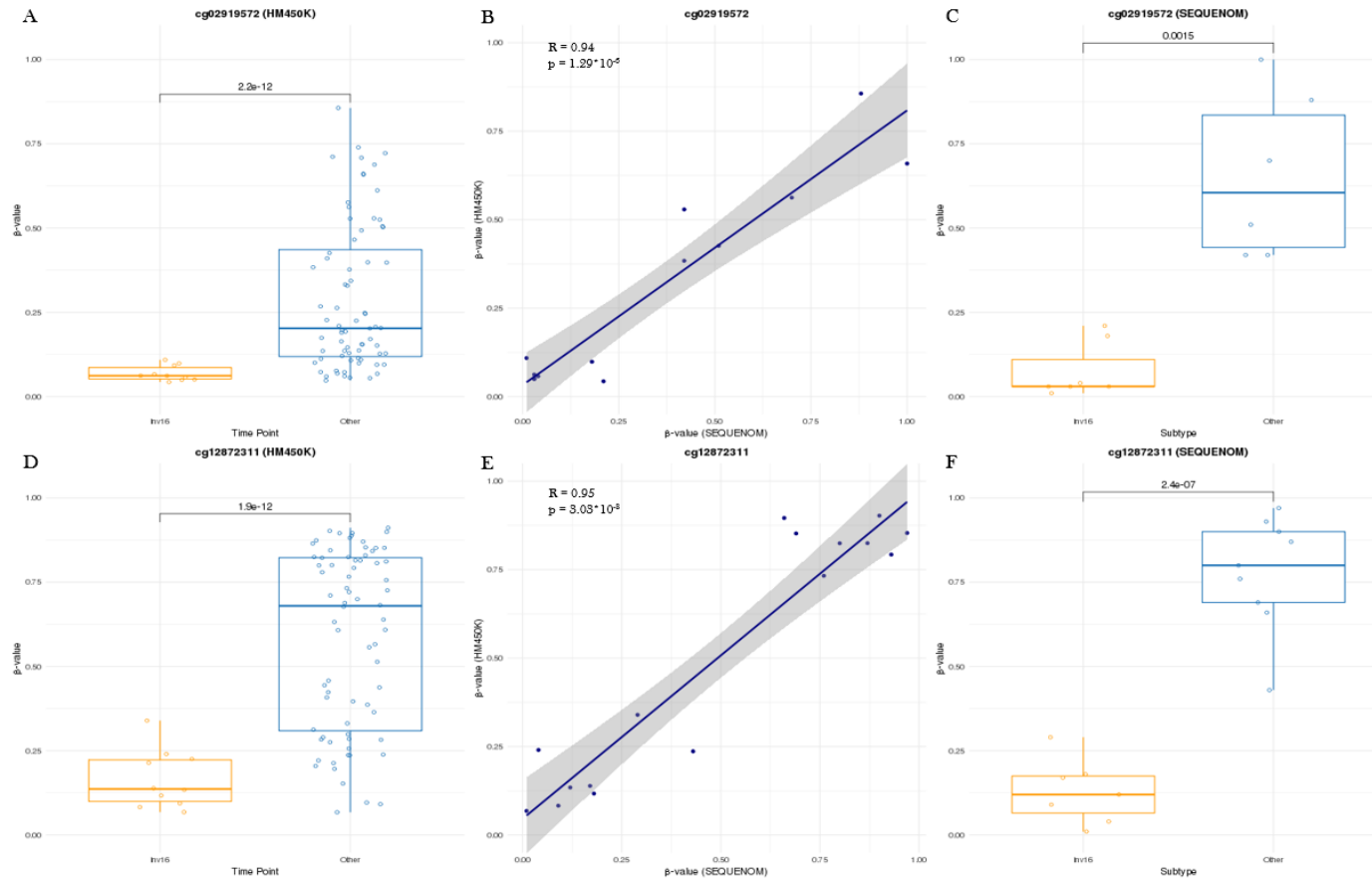


Figure 5.12 Validation of HM450K methylation using MALDI-TOF MS – MN1

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between inv(16) and all other subtypes of paediatric AML. Outliers lay beyond the whiskers of each boxplot. Scatterplot show the original HM450K methylation on the y-axis and the validation (SEQUENOM) methylation on the x-axis. Line of best fit ($\pm 95\%$ CI) is added and statistical analysis is undertaken on all samples (Pearson's R). Figures A/B/C show a boxplot of HM450K methylation between inv(16) and other AML subtypes, a scatterplot and correlation between HM450K/SEQUENOM validation samples, and a SEQUENOM derived methylation boxplot between previously mentioned subtypes at cg05877497 within *MN1*. Figures D/E/F are as above for cg04288999. Both CpG sites were shown to be significantly differentially methylated between inv(16) and other subtypes of AML in both the HM450K analysis ($n=82$) and in validation ($n=16$, $p < 0.001$, Student's t-test). Both sites were highly correlated between HM450K β -value and validated SEQUENOM derived methylation value ($n=14$, $R > 0.9$ Pearson's correlation coefficient).

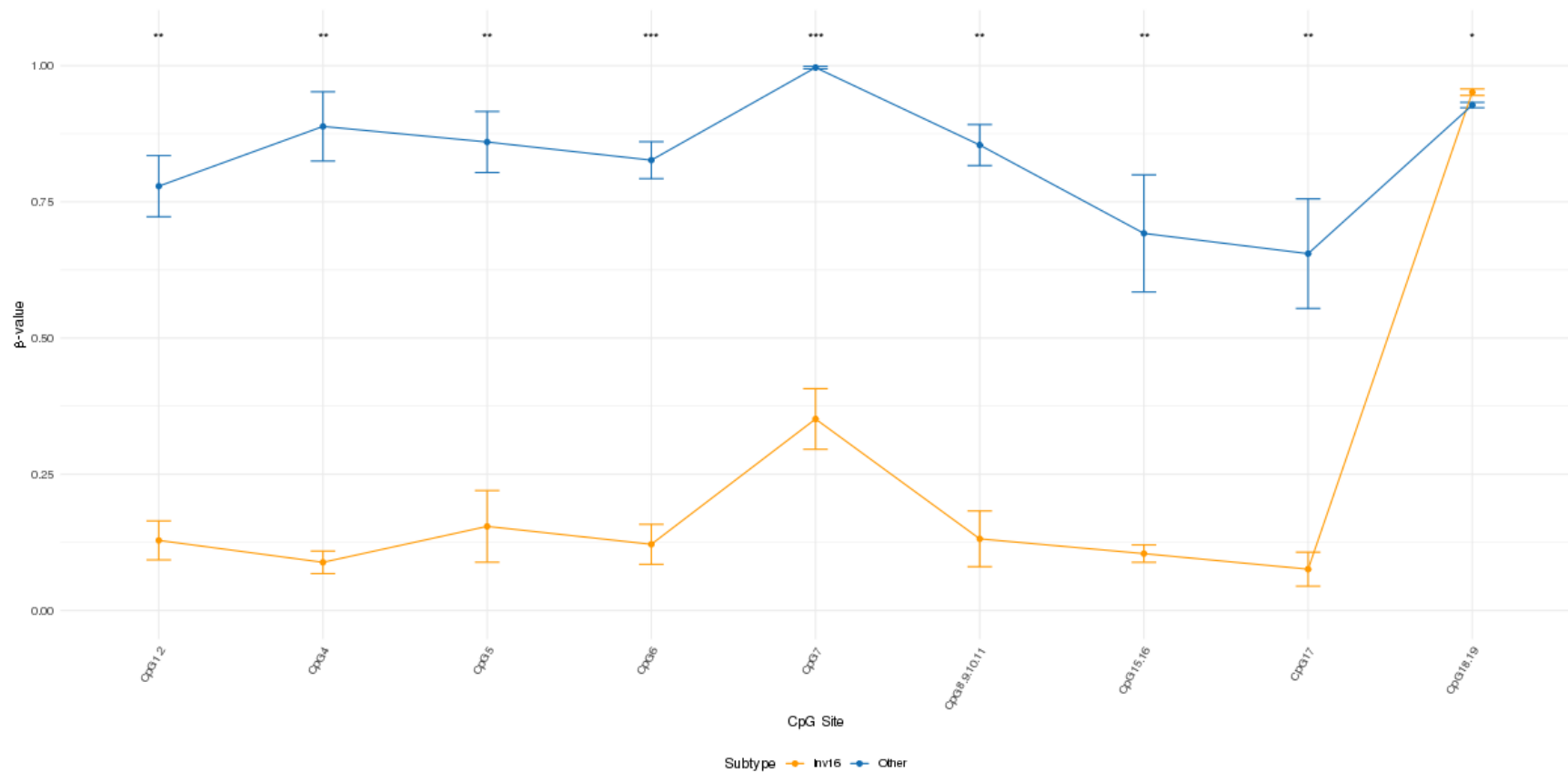


Figure 5.13 Methylation of all CpG units measured by MALDI-TOF MS – *MN1*

Line graph denoting the level of methylation of each CpG site assayed. Groups are split between diagnostic samples of inv(16) and all other AML subtype samples and Student's t-test was used to test for significant differences between groups. Significance is shown as; ns= not significant, *=p<0.05, **=p<0.01, ***=p<0.001. All points are stated as the mean of the group with SEM for error bars. Methylation in *MN1* (n=16) across 9 CpG sites. All sites were significantly differentially methylated (p<0.05, Student's t-test). This denotes a differentially methylated region in their respective areas.

5.5.2 Unsupervised Clustering Analysis of *MEIS1* & *MN1* Methylation to Distinguish Chromosome 16 Rearranged Paediatric AML

Using the MALDI-TOF MS derived methylation data for both *MN1* and *MEIS1*, further unsupervised analyses were undertaken to visualise the degree to which these data, in isolation, could distinguish inv(16) from other subtypes. Multi-dimensional scaling plots were firstly employed to reduce data dimensionality and demonstrate the ability of *MN1* and *MEIS1* to distinguish inv(16) samples (Figure 5.14A and Figure 5.14B respectively). The combination of data from both genes showed the greatest utility to distinguish subtypes (Figure 5.14C), primarily due to variation in one dimension (x-axis). An unknown source of variation (y-axis) also lead to separation of samples and this may warrant further investigation.

Further unsupervised analysis through unsupervised hierarchical clustering validated our findings using MDS analysis. Across 15 samples (6 inv(16), 9 other subtypes) and 19 probes, we obtain an excellent separation of subtypes at the first major split of hierarchical clustering.

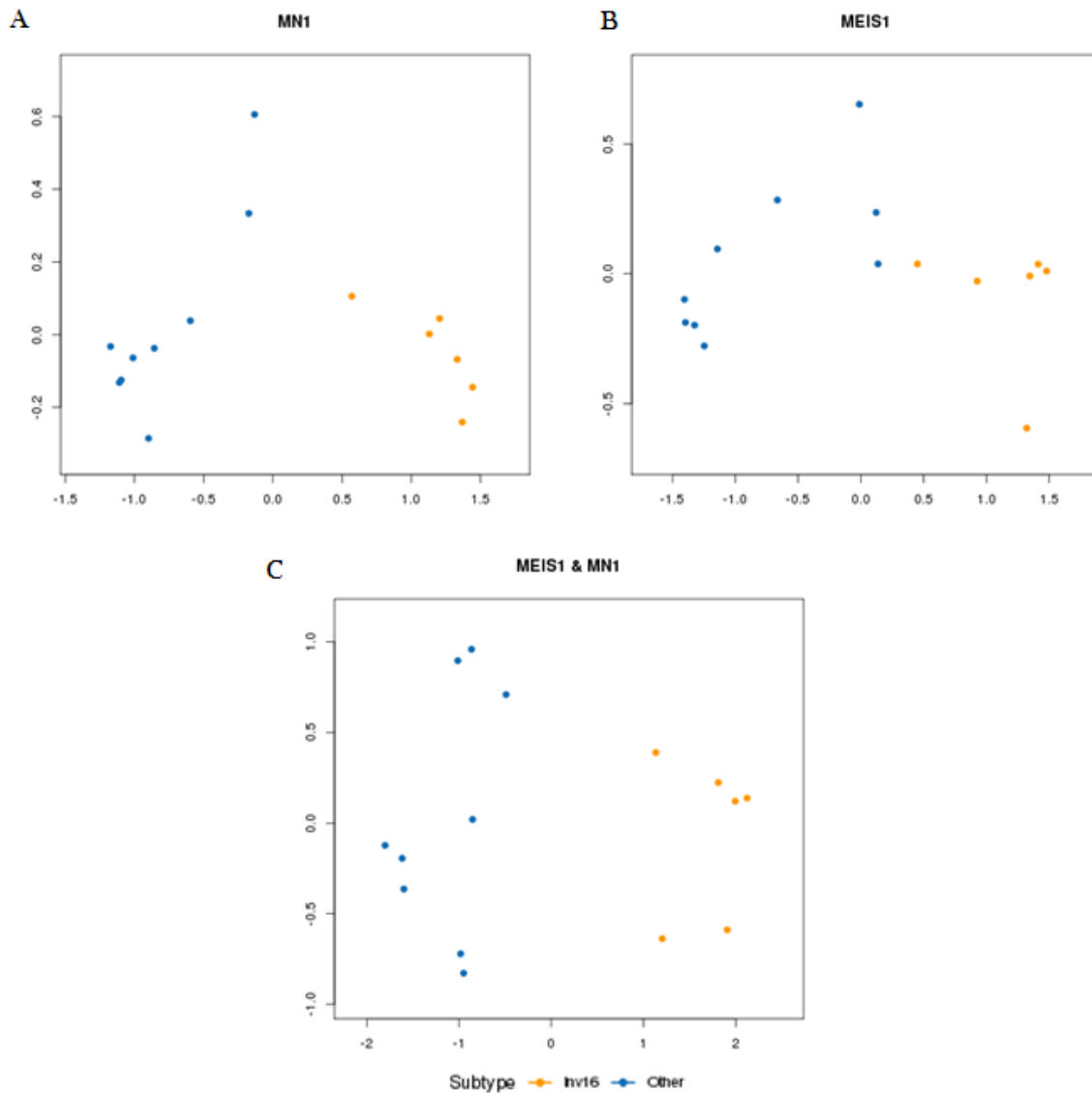


Figure 5.14 Multi-dimensional scaling plot for both *MN1*, *MEIS1* and combination of genes to separate *inv(16)* diagnostic samples from other subtypes of AML

Multi-dimensional scaling plot used to differentiate *inv(16)* and other subtypes of paediatric AML using MALDI-TOF MS derived methylation values of A) *MN1*, B) *MEIS1* & C) Combination of both genes. In all cases, data were able to distinguish *inv(16)* from other subtypes of AML, though the best results were obtained when data were combined.

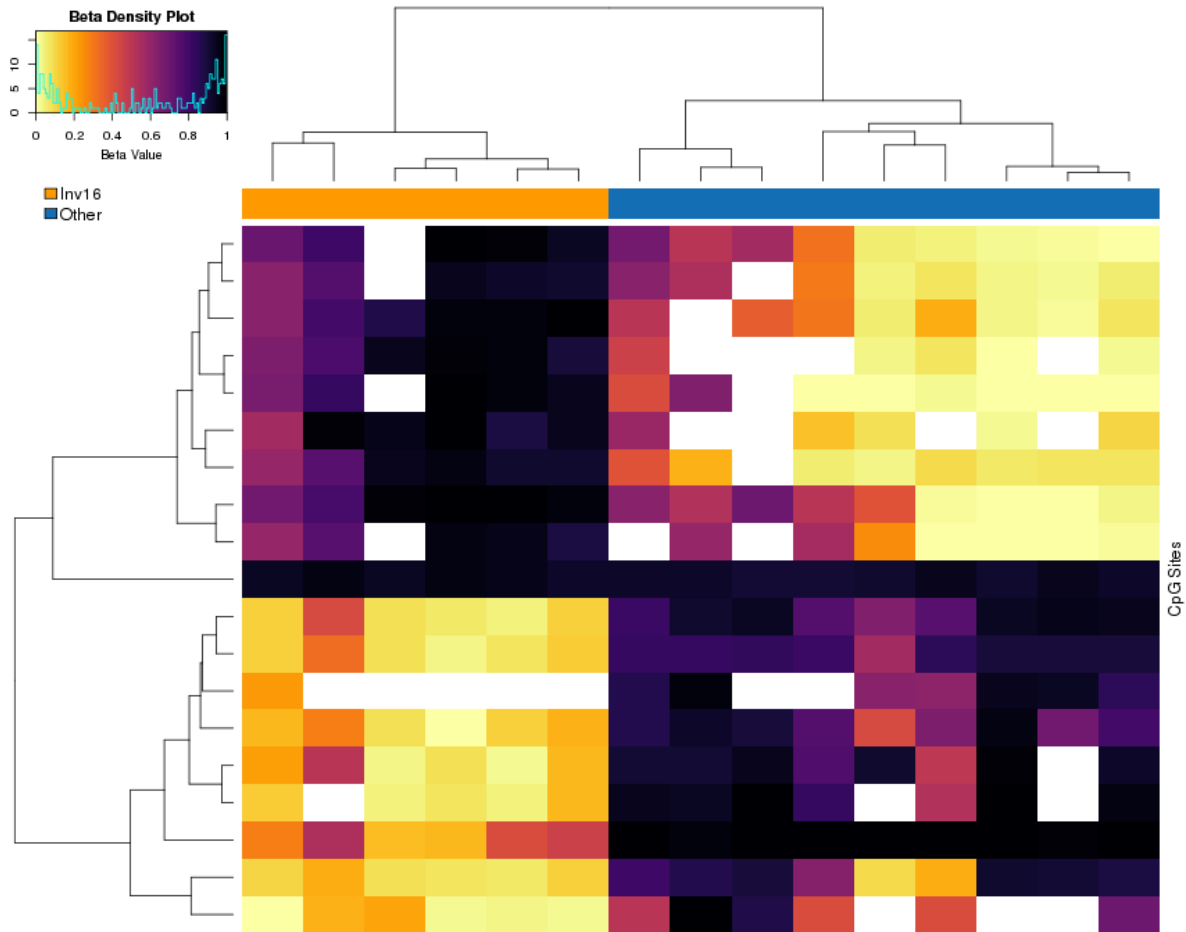


Figure 5.15 Unsupervised hierarchical clustering and heatmap analysis of 19 CpG sites across *MEIS1* and *MN1* to define inv(16) diagnostic samples from other subtypes of AML

Heatmap with the y-axis representing 19 differentially methylated CpG sites, the x-axis shows 15 patient samples where both axes are ordered by hierarchical clustering as depicted by the dendrograms. Hypomethylation is depicted as light yellow while hypermethylation is depicted as dark purple. Missing methylation data is indicated in white. Samples are split into two groups, with one group being inv(16) diagnostic samples (yellow) and the other consisting of all other subtypes of AML (blue).

5.6 A Combined Analysis of DNA Methylation Signatures in Chromosome 16 Rearranged Paediatric AML across Two Independent Cohorts

5.6.1 Principal Components Analysis

With the addition of the TARGET cohort, consisting of 89 new diagnostic samples, pooled-analysis with the Australian discovery cohort was undertaken. Overall, the combined epigenome-wide dataset comprised 16 inv(16) samples and 155 diagnostic samples from other subtypes. Unsupervised PC analysis in isolation again highlighted that DNAm profile alone cannot fully distinguish inv(16) samples from all other subtypes (Figure 5.16). Combination of PC1 and PC2 accounted for 25.9% of the variation found in the dataset and showed that inv(16) samples do tend to cluster together. However, most inv(16) samples also cluster with individual cases of other subtypes using this approach.

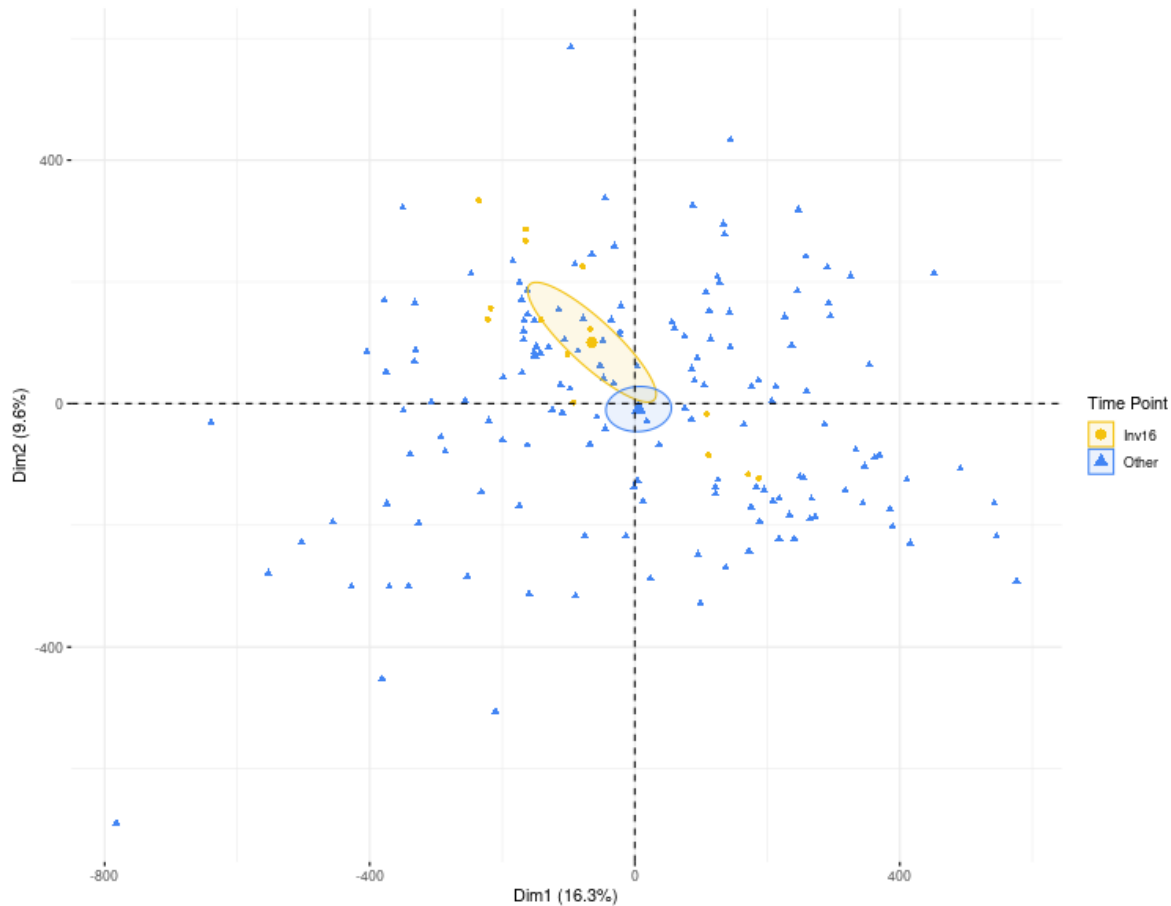


Figure 5.16 Unsupervised principal components analysis plot defining inv(16) paediatric AML from other subtypes in diagnostic bone marrow samples: A pooled-analysis

Methylation beta values of 373,253 probes from the HM450k across 171 samples (inv(16) = 16, Other = 155) were used in a PCA to determine major sources of variation in the dataset. The variation in PC1 and PC2 combined, accounting for 25.9% total variation, cannot actively distinguish inv(16) samples from other subtypes of paediatric AML. Considering the clustering of most inv(16) samples, this suggests that they are similarly methylated. The larger circle and triangle represent their respective group means and ellipses drawn around these averages are defined by 0.95 confidence interval.

5.6.2 Least Squares Regression Analysis

Regression analysis was undertaken in the combined dataset to identify probes that were previously discovered in the Australian cohort. Overall, 5942 probes were differentially methylated between inv(16) samples and all other subtypes in paediatric AML (adjusted $p < 0.0001$). Again, utilising all significant probes showed little ability to distinguish patient subtype, however, restricting the analysis to the top 50 differentially methylated probes allowed for the greatest distinction between inv(16) diagnostic samples and other subtypes of paediatric

AML (Figure 5.17). All inv(16) samples clustered together after the first split in hierarchical clustering, defining itself from the remaining 171 samples with other various genetic abnormalities.

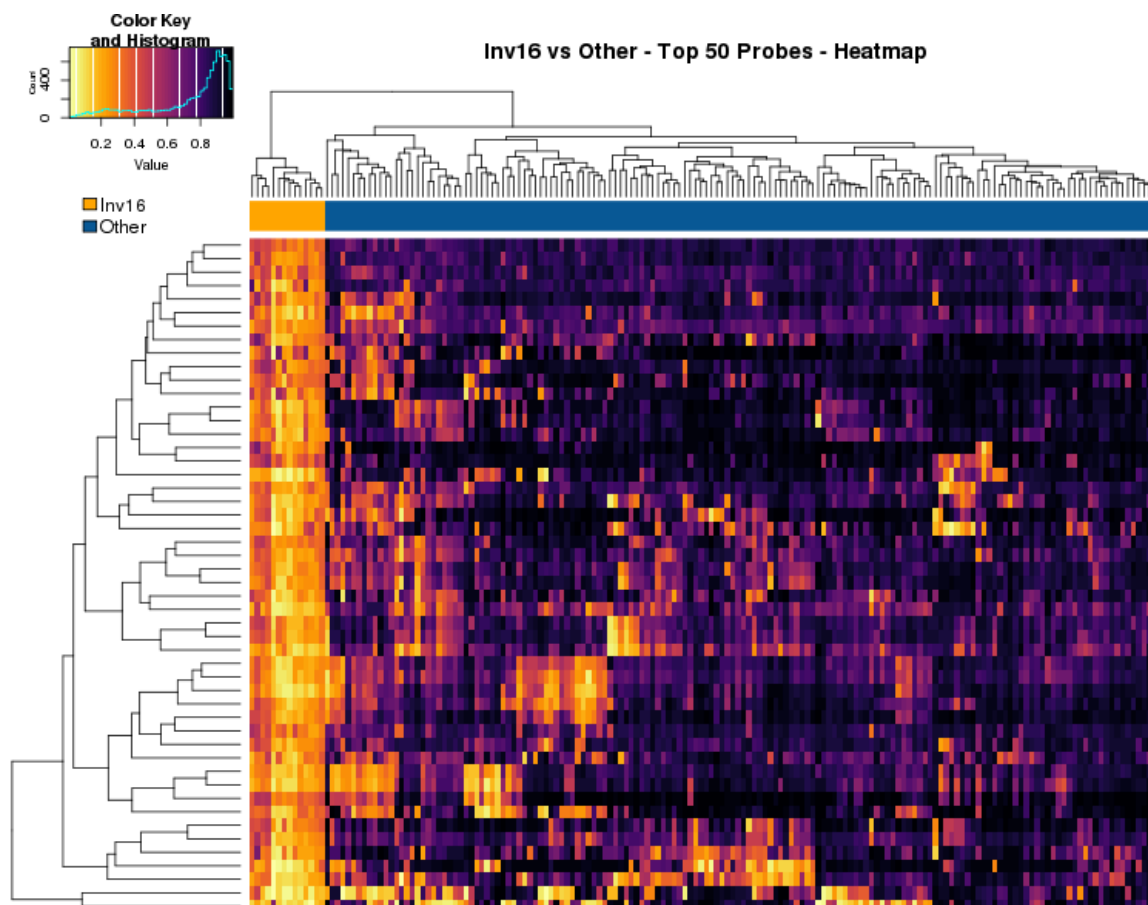


Figure 5.17 Heatmap of beta values across the top 50 differentially methylated probes associated with the inv(16) subtype of paediatric AML

Heatmap where the y-axis is representing the top 50 differentially methylated CpG sites while the x-axis shows 171 patient samples (inv(16) = 16, other subtypes = 155). Both are ordered by hierarchical clustering as depicted by the dendrograms. This heatmap shows that DNA methylation can accurately distinguish the inv(16) subtype of paediatric AML from all other forms of the disease.

Comparing this data to the Australian discovery cohort (at the same p-value threshold of <0.0001) revealed 1317 concurrently identified probes. This accounted for 95.57% of the discovery cohorts probes originally identified and 22.14% of probes found following the pooled-analysis. Probes that were replicated in the pooled-analysis were run through pathway analysis to identify GO terms that may underpin functional consequences in inv(16) diagnostic samples. Originally 161 GO terms were identified however they were further refined using

REViGO software down to 21 umbrella terms. The most significant individual terms multicellular organismal processes, developmental processes and anatomical structural development. The consistent themes between terms circled around developmental processes, cell signalling/communication and stimulus response.

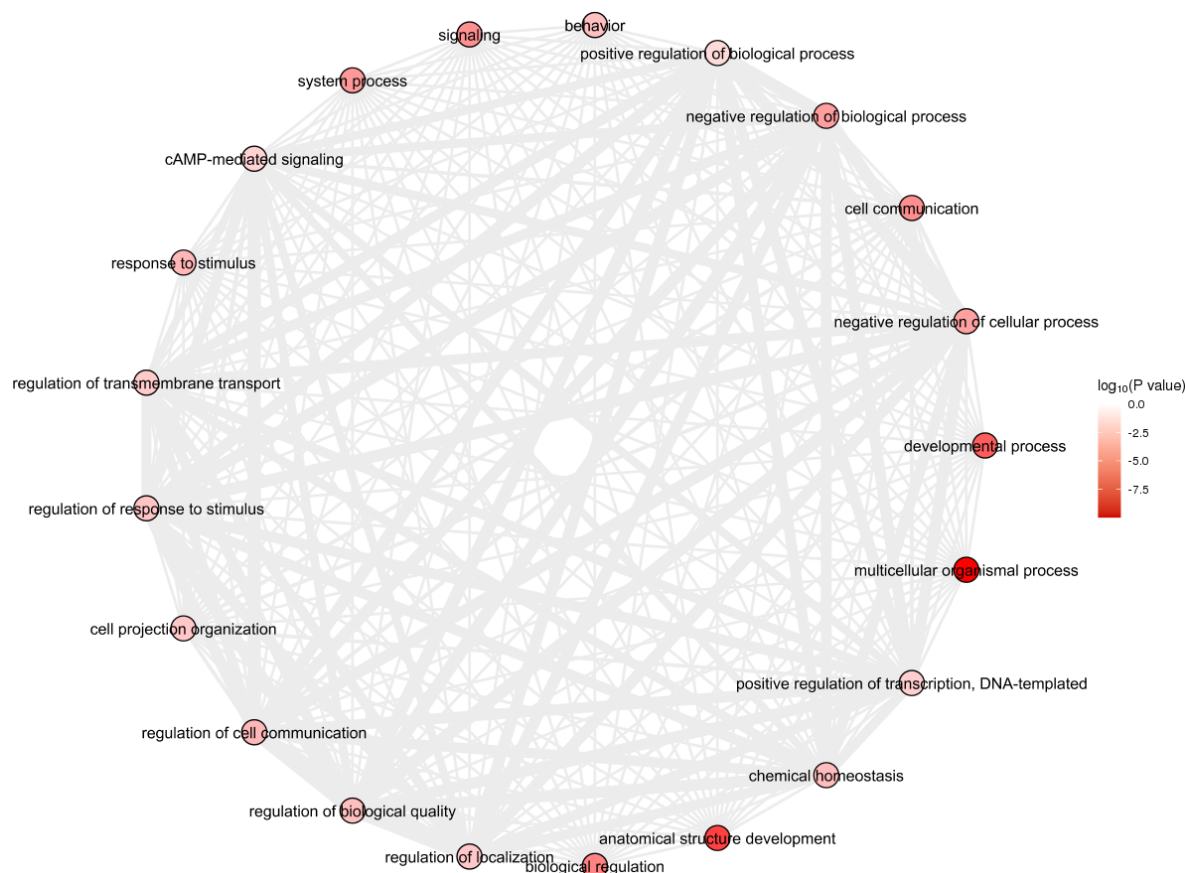


Figure 5.18 REViGO gene ontology network associated with 1744 differentially methylated probes replicated in both discovery and pooled-analyses

REViGO gene ontology reduction identified 21 different GO terms associated with the 1317 significantly differentially methylated probes derived from regression analysis common to both discovery and pooled-analyses. $\log_{10}$ p values are shown for each GO term to indicate the significance of the number of probes that are associated with that term appearing by chance.

5.6.3 Probe and Gene Specific Differential DNA Methylation between Chromosome 16 Rearranged Paediatric AML and Other Paediatric AML Subtypes

Utilising this pooled-analysis, we aimed to validate our previous findings from the Australian discovery cohort. Regression analysis determined that all of the previously identified *MNI* probes remained individually significant with a gain of one probe following the pooled-analysis as well. The 11 probes that were significantly differentially methylated between inv(16) and other subtypes of paediatric AML were found to have an average $\Delta\beta$ of -31.02% (Figure 5.19A, Student's t-test). All significant probes were adjacent to one another including the two probes previously identified as a reprogramming-specific DMR (Doi, Park et al. 2009). As for *MEIS1*, almost all probes remained differentially methylated between inv(16) and other subtypes. These probes were found to have an average $\Delta\beta$ of 35.4% (Figure 5.19B, Student's t-test), however this was down 12.7% from the previous analysis. Significant probes still formed a DMR including five probes that were validated within a known enhancer region.

The newly identified differentially methylated gene with this pooled-analysis was Runt Related Transcription Factor 3 (*RUNX3*). *RUNX3* has been identified as a tumour suppressor gene and it has previously been shown that hypermethylation of this gene can lead to downregulation in AML and other myeloid cell lines (Estecio, Maddipoti et al. 2014). In this study we found evidence of hypermethylation in the promoter region, rather, along with spurious non-promoter hypomethylation occurring across 26 probes. The average $\Delta\beta$ between inv(16) and other subtypes of AML was found to be 13.49% (Figure 5.19C, Student's t-test).

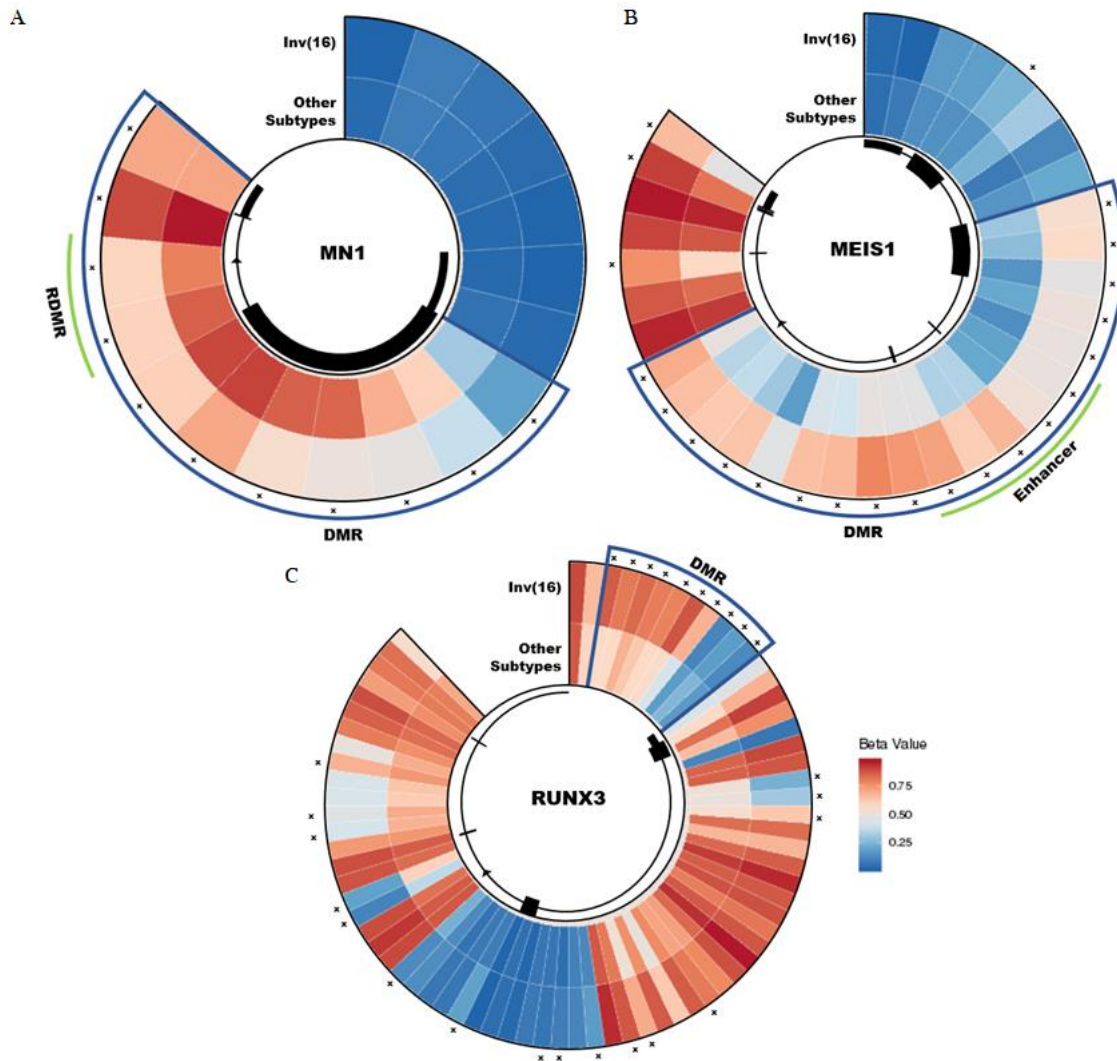


Figure 5.19 Circular heatmaps of differential DNA methylation within genes of interest in *inv(16)* relative to all other subtypes of paediatric AML after pooled-analysis

Circular heatmaps shows mean DNA methylation at each probe that is found in the gene or within 1500 base pairs of the gene's untranslated regions (n=171). Inner circles identify the relative position of any given probe in the gene (not to scale). Mean methylation of *inv(16)* (outer circle) and all other subtypes of AML (middle circle). Completely unmethylated probes are represented by dark blue while complete methylation is represented by dark red. Significantly differentially methylated probes are indicated by asterisk (Benjamini-Hochberg adjusted $p < 0.05$, Student's t-test). Differentially methylated regions (DMR), as indicated by blue wedges, span adjacent significantly differentially methylated probes within 1000 base pairs of each other. Enhancer regions or regions of DNase hypersensitivity found within DMRs are indicated by the outer green line. A) Methylation levels at 18 probes within *MN1*. Eleven probes were identified to be significantly differentially methylated across a single DMR with an average $\Delta\beta$ of -31.02%. B) Methylation levels at 34 *MEIS1* probes. Overall, 23 probes were identified to be differentially methylated across one DMR with an average $\Delta\beta$ of 35.4%. C) Methylation levels at 71 *RUNX3* probes. Overall, 26 probes were identified as differentially methylated with an average $\Delta\beta$ of 13.49%.

Figure 5.20 details the nine most significant probes found in this analysis. These probes were attributed to *ULK1* ($p=1.39 \times 10^{-23}$, Figure 5.20A), *SLC27A3* ($p=1.03 \times 10^{-21}$, Figure 5.20B), *CLIP3* ($p=2.00 \times 10^{-20}$, Figure 5.20C), *C1orf177* ($p=1.64 \times 10^{-19}$, Figure 5.20D), *GEMIN7* ($p=4.32 \times 10^{-19}$, Figure 5.20E), *CST6* ($p=6.47 \times 10^{-19}$, Figure 5.20F), *TMCC2* ($p=9.71 \times 10^{-19}$, Figure 5.20G), *CLIP3* ($p=1.04 \times 10^{-18}$, Figure 5.20H) & *ATP2B2* ($p=1.04 \times 10^{-18}$, Figure 5.20I). Five sites (cg12923179, cg03138446, cg23013977, cg17896463 and cg22719241) were found to be within the top 9 probes in both discovery and pooled-analyses. Thorough inflation of findings were identified ($\lambda = 1.68$), largely due to increased sample numbers. While sample sized drastically increased overall, the amount of inv(16) samples compared to others subtypes created a vastly imbalanced analysis which is reflected in the inflation of findings.

Top probes that were individually identified by least squares regression analysis show a fairly even distribution of hyper and hypomethylated probes (Figure 5.20). The majority of samples tend to cluster together for inv(16) samples at each probe, but we do see that there is a large heterogeneity between all other subtypes, suggesting that the probes identified, while relevant for inv(16), may hold relevance for other subtypes as well. This is reflective of previous analyses in the ability to somewhat cluster this subtype too.

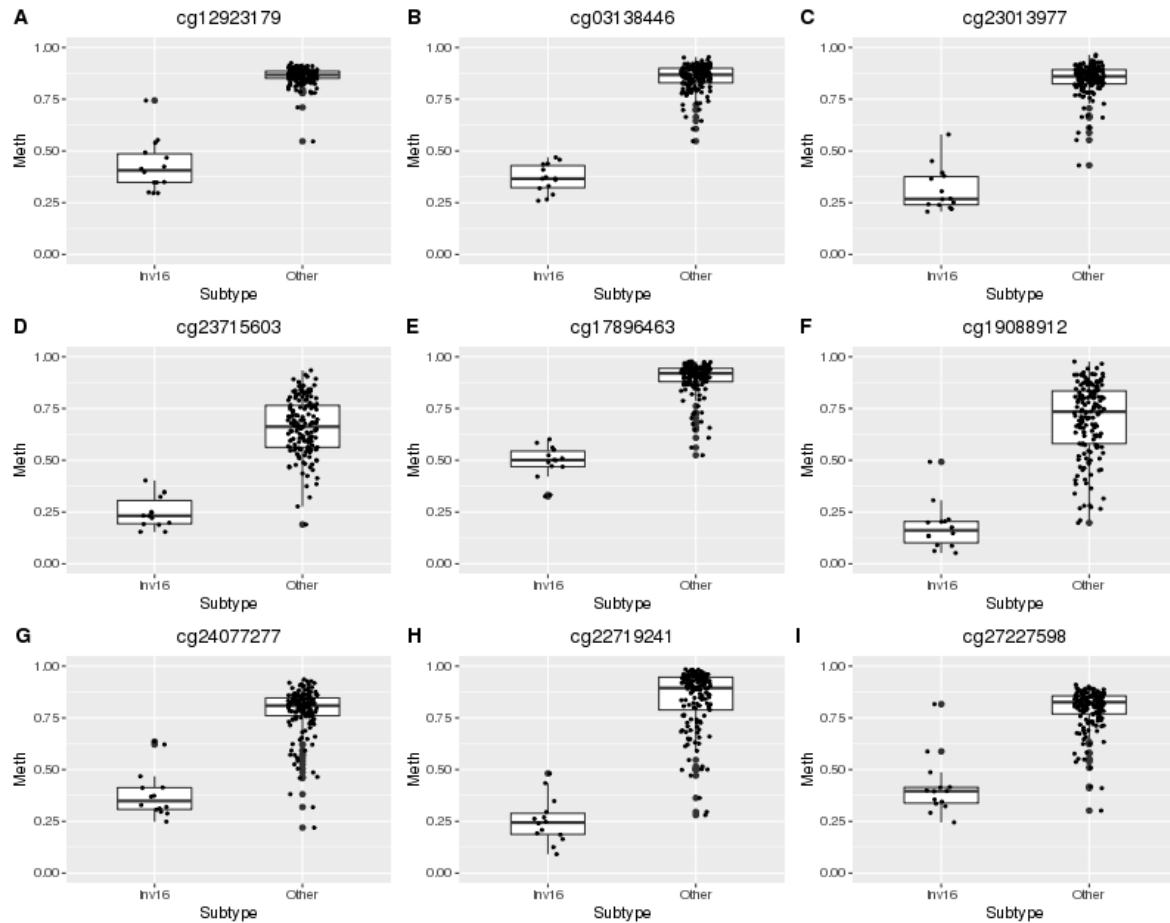


Figure 5.20 Boxplots of the most significantly differentially methylated probes between inv(16) and other subtypes diagnostic bone marrow samples in a pooled-analysis

Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation between patient bone marrow biopsy time points (n=171). The median is represented by the line in the middle of the boxplot while outliers lay beyond the whiskers. Probes were identified as the most statistically significant DMPs ($p \leq 5.52 \times 10^{-6}$, Student's t-test). Probes identified can be attributed to the genes A) *ULK1*, B) *SLC27A3*, C) *CLIP3*, D) *C1orf177*, E) *GEMIN7*, /F) *CST6*, G) *TMCC2*, H) *CLIP3* & I) *ATP2B2*.

5.7 Discussion

It remains challenging to predict outcome and subsequently offer precision treatment in paediatric AML. Although cytogenetic subtyping offers some prognostic utility, many cases with subtype-defined favourable prognosis still succumb to recurrence, while others with poor prognosis may achieve full remission.

The aims of this study were to identify aberrant DNAm specific to cytogenetic subtypes of paediatric acute myeloid leukaemia in an effort to help further categorise cases and define potentially novel subtype-specific functional pathways to disease based on methylation profile. We hypothesised that the most variably methylated probes within the entire (discovery) methylation dataset would suffice for effective patient subtyping in the absence of cytogenetic data. The identification of these subtype-specific biomarkers may have potential for the more precise categorisation of patient samples for personalised treatment.

Using unsupervised analyses and machine-based learning algorithms we determined that specific subtypes, namely inv(16) and *KMT2A*, could be identified based on DNAm in isolation. Using least squares regression analysis we were also able to identify specific probes and genes that are associated with the inv(16) subtype of paediatric AML. These probes were later validated in a larger cohort using locus specific analysis that also generated additional specific methylation data in the region of interest.

5.7.1 Improved Prognostication in AML by Genome-Wide DNA Methylation Profiling

Molecular subtyping (as opposed to karyotype) using DNA methylation profiling has been carried out in several adult and paediatric cancers, with the goal of more accurately predicting patient response and informing treatment. Furthermore, this approach has also been successfully used to understand the functional pathways disrupted in subtypes of acute lymphoblastic leukaemia, potentially clustering patients with complex cytogenetic abnormalities with more common subtypes at the molecular level (Milani, Lundmark et al. 2010; Nordlund, Bäcklin et al. 2015).

Studies have previously investigated the utility of DNA methylation subtyping AML with varying results. Figueroa *et al.* showed that DNA methylation can accurately classify *AML1-ETO*, inv(16) and *PML-RARA* rearranged adult AML as distinct groups, but also noted that there was difficulty in subtyping cytogenetically normal patients by methylation alone (Figueroa, Lugthart et al. 2010). Bolouri *et al.* similarly showed that *IDH1* mutants, *KMT2A* and *CBFB* rearranged AML can also be subtyped effectively in both adult and paediatric forms suggesting a large overlap in differential methylation between the two age groups (Bolouri, Farrar et al. 2017). However, a large proportion of the cohort was unable to be subtyped in accordance with mutation or cytogenetic aberration.

We found clear evidence of a DNA methylation signature (primarily hypomethylation) associated with inv(16) and *KMT2A* subtypes of paediatric AML. Hierarchical clustering identified two major groups being dominated by divergent methylation levels. This is a validation in line with findings from Bolouri *et al.* who also identified low levels of methylation in these cytogenetic subtypes in both children and adults. Considering the fairly consistent trend of hypomethylation in the *KMT2A* subtype, this suggests a dysregulation of an upstream epigenetic mechanism associated with the demethylation of DNA. However, without paired expression and functional data the sequence of events and mechanistic relevance is yet to be determined.

Additionally, a group of hypermethylated samples, largely consisting of cytogenetically normal and novel rearranged forms of AML, was also present in our analysis. This group, was also previously described by Bolouri *et al.* and found to be primarily made up of paediatric cases (Bolouri, Farrar et al. 2017). This group of samples warrant further investigation to elucidate the mechanism of such wide-spread hypermethylation. Despite the genetic heterogeneity of this group, each appears totally hypermethylated suggesting a potential common underlying mechanism between patients that may have arisen in a progenitor cell early in life (Deneberg, Guardiola et al. 2011). Future functional studies (e.g. iPSC) would do well to identify changes over time from pluripotency to haematopoiesis and matched expression data would be useful to identify if there are aberrant changes in gene expression exclusive to this heterogeneous group of samples.

Machine-based learning algorithms were utilised to determine, not only if patients subtypes can be predicted based off methylation profiling, but if feature importance can be used to decrease the number of probes necessary to delineate samples. The clear distinction of *KMT2A* through clustering analysis is likely suggestive of a biological mechanism causing targeted, yet widespread hypomethylation unique to this subtype. Interestingly, we find a slight ability to predict cytogenetically normal samples which was largely unexpected due to the heterogeneous nature of this subtype. Despite this prognostic ability, hierarchical clustering again failed to cluster cytogenetically normal samples together. This does suggest that there may be some small commonality between cytogenetically normal samples that define it from cytogenetically abnormal samples and warrants further study.

5.7.2 The Methylation Signature of Chromosome 16 Rearranged Acute Myeloid Leukaemia

The aim of this study was to define a DNAm based signature associated with the inv(16) subtype of paediatric AML, and to identify regions of differential methylation that may have functional consequence.

The Australian discovery cohort identified 12,003 probes significantly associated with inv(16) through least squares regression and when restricted to the top 100 most associated probes we find a signature that entirely distinguishes inv(16) paediatric AML from other forms of the disease. Upon pooled-analysis, this signature did not validate, however a new signature utilising the top 50 most associated probes achieved the best level of separation between inv(16) samples and all others. Clustering analysis did not manage to cluster all samples, suggesting that there is a level of intra-subtype heterogeneity likely caused by the multitude of secondary mutations or chromosomal aberrations (Marlton, Keating et al. 1995; Paschka, Du et al. 2013). As discussed in the review by Kumar *et al*; this heterogeneity (in all AML subtypes) has potential to modulate patient outcomes and treatment resistance (Kumar 2011). This further solidifies the need for accurate identification of patient samples, both between and within subtypes, to inform targeted treatment response. Larger studies would need to be conducted to identify the intra-subtype methylation differences and how this may be affecting pathogenesis. Between analyses, 42% of probes identified in the pooled-analysis were found in

the discovery analysis, highlighting that there is a fairly substantial DNAm signature associated with inv(16) AML, while also showing the necessity for validation cohorts.

5.7.3 Pathways Showing Epigenetic Disruption in Chromosome 16 Rearranged AML

Current and previous studies have shown that differential methylation exists at a subtype-specific level in AML (Figueroa, Lugthart et al. 2010). This suggests that epigenetic aberrations have an effect upon the pathogenesis of each individual subtype of AML. Goldman *et al.* discuss in their review the lack of insight into epigenetic alteration in inv(16) AML (Goldman, Hassan et al. 2019) with the exception of differential methylation, and subsequent expression, in *MNI* discovered by Larmonie *et al.* (Larmonie, Arentsen-Peters et al. 2017). To address this we explored individual probes clustered by gene in our least squares regression analysis and revealed that several clusters of probes were concurrently differentially methylated in inv(16) samples compared to other forms of AML. Discovery analysis identified *MEIS1* as prime a candidate for potential altered expression based on changes to DNAm while also validating the finding of Larmonie *et al.* with the identification of aberrant methylation of *MNI* (Larmonie, van den Heuvel-Eibrink et al. 2014).

MEIS1

MEIS1 is a homeobox gene and is known to play a key role in early life development and dysregulation in cancer. Several studies have shown that *MEIS1* is involved in chemotherapy resistance and maintenance of cell survival in AML (Kühnl, Kaiser et al. 2011; Liu, Qin et al. 2017). Furthermore, murine studies have shown that *MEIS1*, in the context of *HOX* gene overexpression, causes a feedback loop upregulating Syk signalling and driving subsequent disease (Mohr, Doebele et al. 2017). It has previously been demonstrated to be silenced by DNA hypermethylation in several adult cancers, including AML (Lasa, Carnicer et al. 2004; Zangenberg, Grubach et al. 2009; Musialik, Bujko et al. 2015), and can be modulated by *DNMT3A* mutation status (Ferreira, Heyn et al. 2015; Larmonie, Arentsen-Peters et al. 2017).

MEIS1 was clearly differentially methylated in a way that may prove useful as a biomarker of disease in paediatric inv(16) AML. We have validated findings of previous studies showing aberrant methylation in the promoter region of *MEIS1* (Larmonie, van den Heuvel-Eibrink *et al.* 2014). While absent in this study, previous studies suggest this aberrant methylation affects expression levels and is indicative of patient outcome.

MNI

MNI has been identified as a primary candidate for contribution to AML disease pathogenesis due to several studies previously describing tumour suppressor activity and disruption contributing to other cancers such as meningioma and colorectal cancer (Lekanne, Riegman *et al.* 1995; Jo, Nam *et al.* 2012). On the other hand, high levels of expression of *MNI* has been associated with poor patient outcome in AML suggesting it plays a dual role in oncogenic activity and a delicate balance of expression is necessary to keep it in check (Heuser, Beutel *et al.* 2006). Pooled-analysis confirmed the finding of an average 22.8% reduction in methylation in inv(16) samples compared to all others across eight probes in potentially functionally relevant regions of the gene (1st exon). We identified a DMR within *MNI* showing an average reduction in methylation of 60.2% in inv(16) samples while simultaneously validating findings from two probes from the HM450K array. While these differences strongly associate with inv(16) AML, there is a large level of variability within the other subtypes, suggesting that *MNI* methylation may be a contributing, or causal factor to many subtypes. Heuser *et al.* showed that *MNI* is highly expressed in cytogenetically normal adult AML patients (Heuser, Beutel *et al.* 2006) and is therefore a likely key contributor to the subtype variability observed in our data. Grosveld *et al.* suggests that the differentiation inhibition of the *CBF β -MYH11* fusion in combination with the growth induction of *MNI* create a perfect environment for AML (Grosveld 2007). However, changes in DNAm require a mechanistic input and Larmonie *et al.* showed that changes in *MNI* DNAm is induced by reduced function of *DNMT3 β* (Larmonie, Arentsen-Peters *et al.* 2017). Thus, it is likely a combination of the three genes that are leading to the initiation of inv(16) AML, and this study builds evidence that it may occur in the paediatric setting as well. Future studies should focus on measuring the capacity of *DNMT3 β* and *MNI* expression to truly elucidate these genes functionality in paediatric inv(16) AML.

RUNX3

Finally, pooled-analysis identified the promoter region of the gene *RUNX3* as both hypo- and hypermethylated in inv(16) samples compared to all other subtypes. This finding is of particular interest as high expression of *RUNX3* correlates with poor prognosis to patients. This finding being solely localised to a promoter region of the gene suggests that there may be some control upon gene expression induced by DNAm. Interestingly, multiple studies have shown that hypermethylation of the promoter region of *RUNX3* and subsequent repression of gene expression is associated with the inv(16) subtype (Estecio, Maddipoti et al. 2014; Estécio, Maddipoti et al. 2015), although this change has been shown to be caused by the fusion protein itself rather than the change in methylation (Cheng, Li et al. 2008). While we did identify similar hypermethylation patterns, the closest probes to the transcription start site were hypomethylated in inv(16) samples, therefore prompting necessary further investigation to identify how this might be affecting gene expression.

5.7.4 Future Directions & Limitations

Overall this study suggests that there is unique aberrant methylation associated with the inv(16) subtype of paediatric AML in *MEIS1*, *MN1* and *RUNX3*. Our study validates findings in the paediatric setting of previous work in adult AML for both *MEIS1* and *MN1*, however it challenges the current literature in the methylation status and association of *RUNX3* in the inv(16) samples. Future studies should assess combining methylation data with the expression and mutational profiling of these genes to discover downstream functional consequences. Such studies would also do well to look at patient treatments and outcomes in relation to apparent subtype-specific methylation patterns to determine if this correlates with improved accuracy of prognosis or outcome over traditional measures. For now, we have shown that there are contrasting methylation levels between inv(16) and other forms of paediatric AML that serve well as a biomarker of disease and show potential prognostic implications for both treatment and overall outcomes. However, it is clear that considerably larger sample sizes, with multiple ‘layers’ of genetic, epigenetic and expression data will likely be needed to more fully explore the potential of molecular subtyping for prognosis and informing treatment.

Limitations for this study mainly revolve around sample acquisition and data availability. Data from international studies only became available after analysis of local samples was complete, which was unideal. In an already relatively small patient cohort, further dividing into subtypes severely diminished power to identify probes with predictive ability, especially when using variance as the main determinant. In future studies, it would be ideal to increase sample size so that each subtype is well represented, with sufficient power to detect subtype-specific methylation signatures of modest-to-small size. This would also allow regression analysis to be performed on each individual subtype against the remainder of the cohort as was performed here specifically for *inv(16)* cases. Not only would this allow for greater separation between subtypes, but it may also provide valuable insights into the underlying mechanisms associated with each, thereby potentially informing novel treatment approaches (Newcombe, Gibson et al. 2018).

An additional issue in the current study was the extent of genetic data acquisition of each sample, with the majority lacking any information related to point mutation status. Previous studies have found DNAm signatures of adult AML samples associated with common SNPs, also linked to the disease (Griffiths, Gore et al. 2010). Therefore, it is likely that some samples may be clustering based on specific point mutations rather than chromosomal rearrangement in isolation and this cannot be detected without appropriate genetic profiling. It is also likely that aberrant DNAm signatures result from a combination of effects of multiple genetic abnormalities. As previously mentioned, this can create an intra-subtype heterogeneity causing clustering to be ineffective and indicative of a SNP based signature rather than a chromosomal lesion signature. For future predictive studies, genetic profiling of common SNPs in AML (such as *FLT3*, *TP53*, *RUNX1*, *ASXL1*, *NPM1*, *KIT*, *DNMT3A*, *IDH1*, *IDH2* and *TET2*) is paramount to ensure appropriate clustering of samples.

Finally, as previously mentioned, using variance as the definitive form of differential methylation to subtype all forms of paediatric AML is rudimentary. With the advent of MBL we can explore other unsupervised methods of subtyping, however input data is largely the defining factor for performance of both the model creation and prediction. With limited computational power we were only able to use the 10,000 most variable probes as our input data from a total pool of 373,253 probes. While many of these probes were likely redundant

in terms of their subtyping capability, MBL can identify the most useful probes to determine correct subtype and therefore should be able to identify the top probes from the whole dataset rather than just a subset of the most variable probes. This lack of computational power could be corrected using either a combination of multiple individual models best performing probes, or preferably using a super-computer to run all probes simultaneously in a single model and removing redundancy manually.

Addressing these issues may reveal a marked improvement in both unsupervised and MBL based predictive analyses. Furthermore, we may even be able to breakdown the data into smaller categories and find unique DNAm signatures associated with combinations of genetic and chromosomal abnormalities. In reference to the data at hand, we may also be able to identify why particular samples do not cluster with their respective subtypes despite discovery analyses suggesting that these are fairly similar to one another.

This page is intentionally left blank.

Chapter 6 The Potential of DNA Methylation as a Predictor of Outcome in Paediatric Acute Myeloid Leukaemia

6.1 Rationale

Paediatric acute myeloid leukaemia may be one of the rarer forms, however the rate of mortality far exceeds that of others (Dohner, Estey et al. 2010). The reasons behind this are two-fold. Firstly, despite an induction therapy success rate of 80-90%, the rate of relapse in AML is estimated to be around 40%. Relapsed patients have an exacerbated disease state, become more difficult to treat due to chemotherapy resistance, and often have clonally evolved and diverse leukaemia which complicates targeted therapy (Ding, Ley et al. 2012). The mortality rate of AML jumps from between 10-20% to upwards of 60% upon recurrence (Puumala, Ross et al. 2013). Secondly, with paediatric AML being a rare disease, the body of knowledge into the pathogenesis and prognosis is lacking relative to other types of leukaemia. Due to this lack of knowledge, classification and current prognostic measures have been designed with adult leukaemia in mind and therefore may not adequately translate to the paediatric setting (Mrozek, Marcucci et al. 2012).

The classification system currently in place for AML is defined by the World Health Organisation guidelines (Arber, Orazi et al. 2016). Incorporated into this is a risk stratification standardised system as suggested by the European LeukaemiaNet (ELN) study (Dohner, Estey et al. 2010) and recently updated by Estey *et al* to account for new research in the field (Estey 2016). The primary outcome of risk stratification is to be able to accurately identify those that are most at risk and act accordingly should they not respond to induction therapy. This enables some targeting of treatment. Potential novel trials are available to patients who are unresponsive. However, studies have shown that co-mutation of genes that are not incorporated into current stratification can inform treatment and potentially change patient outcomes (Bolouri, Farrar et al. 2017). Some common chromosomal aberration are not included at all. While the current stratification model has shown to be replicable and clinically useful (Mrozek, Marcucci et al. 2012), it is becoming increasingly clear that expansion and subtype specific co-mutation needs to be incorporated to adequately define patient risk.

Genetics-based prognosis has been the standard for AML, however many alternatives to prognosis have been suggested with varying outcomes. Gene expression panels, utilising an assortment of genes to predict patient outcome has been widely studied (Metzeler, Hummel

et al. 2008; Li, Herold et al. 2013; Niavarani, Herold et al. 2015), however these studies have often been unable to validate previous findings and are rarely utilised in the clinical setting despite improving upon the ELN stratification. The feasibility of these analyses are hindered by availability of RNA from fresh patient samples and therefore rarely translate clinically. Unfortunately, many of these other prognostic tests (largely genetic in nature) are very specific to subtype and do not take into account the variability of AML, making their prognostic scope too narrow to be effectively translated (Zhou, Yin et al. 2015; Candoni, De Marchi et al. 2017; Zhang, Lin et al. 2017). While these aforementioned tests are useful for refining the current ELN system, one could argue that incorporating other mechanisms of disease beyond genetics would alleviate some heterogeneity that comes with standard classification by cytogenetics.

DNAm has been suggested as an additional tool to be utilised alongside genetic and expression based prognostic markers. It is very dynamic in early development in association with cell/tissue specification, yet generally stable once established. DNAm is also fairly easily measured relative to other epigenetic marks implicated in tumorigenesis and has been widely studied in a range of cancers, both for functional insights and potential prognostic use. Examples of DNAm as a prognostic biomarker include studies by Nordlund *et al.* who utilised DNAm in 764 paediatric patients to accurately distinguish subtype in ALL patients (Nordlund, Bäcklin et al. 2013) and Capper *et al.* who collated the largest collection of DNAm array data across all ages and forms of CNS tumours (n=2801) to create a classifier that is purported to be more accurate than histopathological diagnosis (Capper, Jones et al. 2018).

In the current study we aimed to use a similar approach to that of Capper *et al.*, encompassing machine-based learning using HM450k data to predict paediatric AML patient outcome from diagnostic bone marrow aspirate samples. Furthermore, we also attempted to predict AML patient outcome independent of underlying genetic variation, therefore distinguishing it from the standard ELN classification system which is lacking in extended genetic alterations and does not translate as well in the paediatric setting.

The overall aim of this study was to predict patient outcome using DNAm array data in both a discovery and pooled combined cohort. We hypothesised that;

- (i) There is a distinct DNAm profile at diagnosis that can distinguish patients that relapse or do not respond to induction therapy, relative to those that achieve an uninterrupted 5-year EFS;
- (ii) DNAm differences associated with elevated risk of poor outcome can elucidate aberrant functional genes or pathways that may play a role in the AML recurrence, and;
- (iii) Identified DNAm differences can be utilised in a machine-based learning environment to predict patient outcome at diagnosis.

6.2 Study Cohort

The total patient cohort was compromised of a total 158 AML samples with varying cytogenetics, sex and patient outcomes. All samples are patient derived bone marrow taken at diagnosis and a characterised breakdown of these samples can be found in Table 6.1. Good outcome patients (n=36) achieved a 5-year EFS without induction failure or relapse, while poor outcome patients did not (n=122). Males were slightly overrepresented (ratio 1.2), however this is to be expected as males are slightly overrepresented in the paediatric AML population (Puumala, Ross et al. 2013). Age of patients at diagnosis primarily consists of children aged 3-14 but is well represented by other age groups as well. The average age of patients was 8 with all patients between the ages of <1 to 18 at the time of diagnosis, with the exception of three TARGET patients at 19, 21 and 23 years of age. Clinical patient characteristics were obtained from the CLARA database for RCH (VIC) cases while other cohort information was received upon receipt of samples from their respective biobanking repositories. All data is correct as of April 2018.

Current patient status again closely reflects the general population, with approximately 30% of patients having a poor outcome. Most abnormalities found in AML are covered with 60% of the cohort being represented by the most common abnormalities such as *KMT2A* rearrangements, inv(16), t(8;21) and CN patients. The replication cohort (TARGET) consists of predominantly poor outcome samples due to cohort inclusion criteria favouring those with poor risk outcomes and patients that are in clinical trials which is heavily biased towards patients with rare or notoriously difficult to treat forms of AML.

Table 6.1 Clinical characteristics of good and poor outcome patients from both the Australian (discovery) and TARGET (replication) AML cohorts

Clinical Characteristics	Good Outcome (%) (n = 36)	Poor Outcome (%) (n = 122)
Outcome		
Initial 5-year EFS achieved	36 (100%)	0 (0%)
Relapse	0 (0%)	86 (70.5%)
Induction failure	0 (0%)	36 (29.5%)
Sex		
Male	23 (63.9%)	63 (51.6%)
Female	13 (26.1%)	59 (48.4%)
Age at diagnosis		
Infants (<3 y/o)	14 (38.9%)	29 (23.8%)
Children (3-14 y/o)	19 (52.8%)	60 (49.2%)
AYA (15+ y/o)	3 (8.3%)	33 (27%)
Status		
Alive	34 (94.4%)	36 (29.5%)
Deceased	0 (0%)	86 (70.5%)
Unknown [†]	2 (5.6%)	0 (0%)
Cytogenetics		
inv(16)/t(16;16)	5 (13.9%)	6 (4.9%)
t(8;21)	1 (2.8%)	6 (4.9%)
<i>KMT2A</i> rearranged	10 (27.8%)	23 (18.9%)
Trisomy 8	3 (8.3%)	3 (2.5%)
Trisomy 21	1 (2.8%)	1 (<1%)
CN	6 (16.7%)	37 (30.3%)
Complex karyotype*	2 (5.5%)	5 (4.1%)
Other [‡]	8 (22.2%)	30 (24.6%)
Unknown	0 (0%)	6 (9.0%)
Sample Source		
RCH(VIC) } Discovery	20 (55.5%)	22 (18.1%)
WM } Cohort	10 (27.8%)	17 (13.9%)
RCH(QLD)	0 (0%)	4 (3.2%)
TARGET AML	6 (16.7%)	79 (64.8%)

Abbreviations: EFS - Event free survival; y/o - Years old; AYA - Adolescent/Young adult; inv - Inversion; t - Translocation; CN - Cytogenetically normal; RCH - Royal Children's Hospital; VIC - Victoria; WM - Westmead; QLD - Queensland; TARGET AML - Therapeutically Applicable Research To Generate Effective Treatments in Acute Myeloid Leukaemia.

*Complex karyotypes are defined by patients having two or more cytogenetic abnormalities or abnormalities that are otherwise unrelated to AML pathogenesis. [‡]Other cytogenetic abnormalities are defined as del5/7/9/13/20q, inv(5q), -7, t(5;6), t(6;9), t(7;14), t(15;17), t(16;21). Genetic abnormalities are mostly undefined.

[†]Patients achieved a 5-year EFS however recent data is unavailable to ensure that they are currently still alive.

6.3 DNA Methylation Profiling to Predict Patient Outcome in an Australian AML Cohort

6.3.1 Using DNA Methylation to Define Patient Outcome

As previously outlined, patients in our Australian AML cohort (n=47) that were defined as ‘poor outcome’ went on to relapse (Figure 6.1A, $p < 0.0001$, log rank test). This occurred as late as 900 days after original diagnosis.

Prior to implementing DNAm into any analyses, we interrogated standard epidemiological data for their prognostic ability within the cohort. Fisher’s exact test showed that age, sex and patient cytogenetics had no significant correlation with patient outcome (Table 6.2). However, log rank test of survival analysis revealed that while patient cytogenetics did not influence EFS (Figure 6.1C,) it did influence patient survival (Figure 6.1B, $p = 0.046$). Furthermore, receiver operating characteristic (ROC) analysis confirmed that these clinical characteristics are unable to accurately predict patient outcome in this cohort, both separately and in combination. Therefore, while the current standard in mortality prognosis (ELN classification) does serve well to predict patient mortality, it does not adequately predict patient relapse.

Table 6.2 Fisher’s exact test and ROC analysis predictive utility of clinical characteristics for patient outcome

Clinical Characteristics	p-value	AUC (95% CI)
Sex (male, female)	0.467	0.5543 (0.440 - 0.668)
Age at diagnosis (infant, child, AYA)	0.326	0.5922 (0.475 - 0.710)
Patient cytogenetics*	0.509	0.6140 (0.481 - 0.748)
All combined	N/A	0.6461 (0.515 - 0.777)

Abbreviations: AUC – Area under the curve

*Patient cytogenetics was split into inv(16)/t(16;16), t(8;21), *KMT2A* rearranged, trisomy 8, CN, complex and all others fell into category ‘other’.

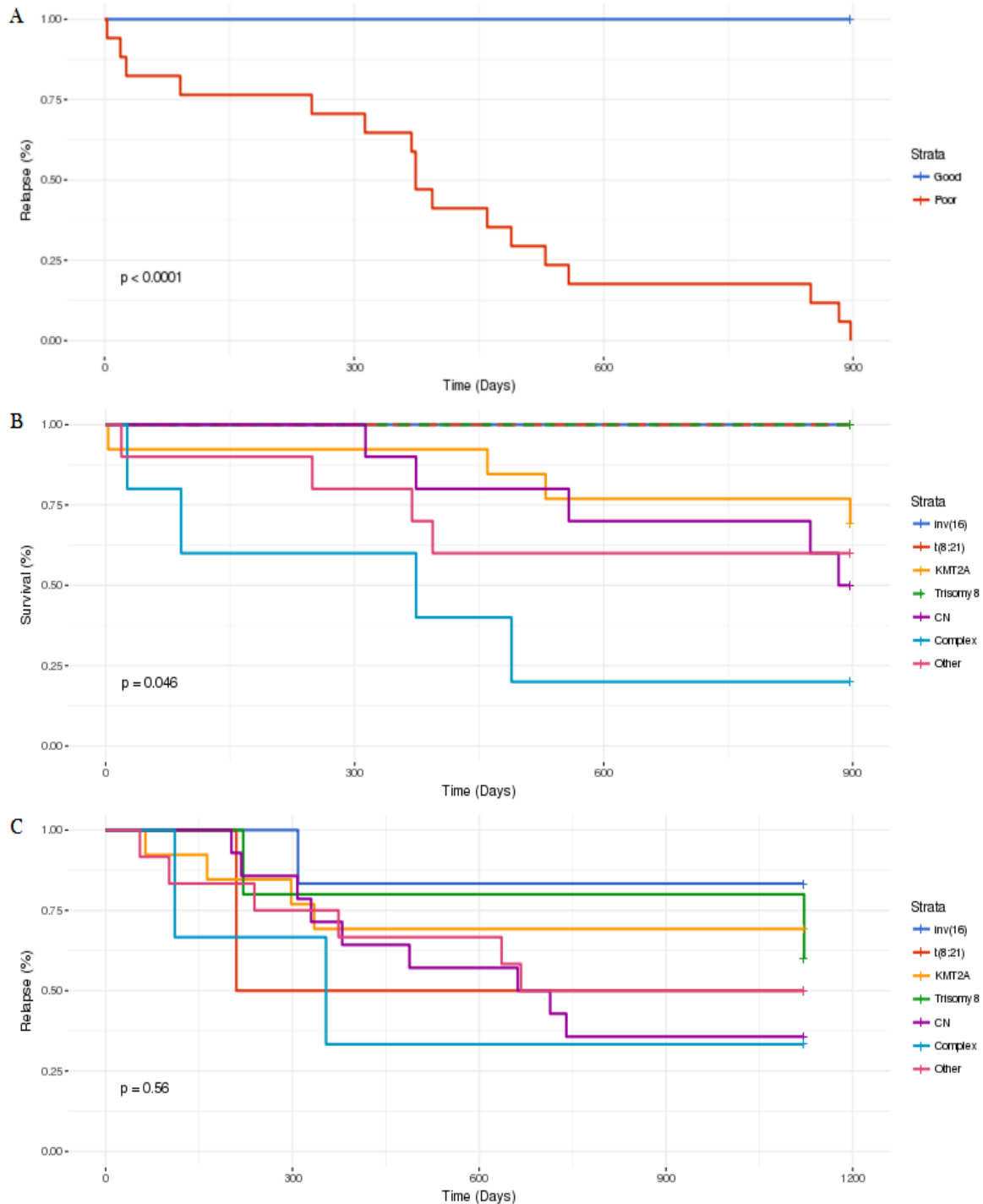


Figure 6.1 Survival and relapse curve analysis based on patient outcome and cytogenetic abnormalities

Time in days from original diagnosis is shown on the x-axis with survival or relapse probability on the y-axis. (A) Time for all patients with poor outcome to relapse (poor outcome, $n=17$, $p<0.0001$, log-rank test). (B) Patients with a complex karyotype ($n=5$) had a significantly shorter survival ($p=0.046$, log-rank test) compared to all other cytogenetic abnormalities ($n = 42$), however (C) cytogenetics did not significantly influence EFS between patients ($n=55$, $p=0.56$, log-rank test).

As a first step to examine the variation in the data, unsupervised principal component analysis (PCA) was undertaken on data from samples sourced at the RCH VIC, RCH QLD and WM NSW (n=73, Good outcome=30, poor outcome=43). Principal component 1 (PC1), which accounts for 12% of the total variation in the dataset, was partially attributed to patient outcome ($p = 0.03$, Student's asymptotic p-value for correlation), however this was confounded by chip effects ($p = 4.0 \times 10^{-4}$) and where the sample originated from ($p = 0.04$). Figure 6.1 illustrates a dimensionality reduction looking at only the PC1 and PC2 sources of variation, highlighting the generally poor separation of good and poor outcome samples. Additional PCs were also unable to distinguish patients by outcome.

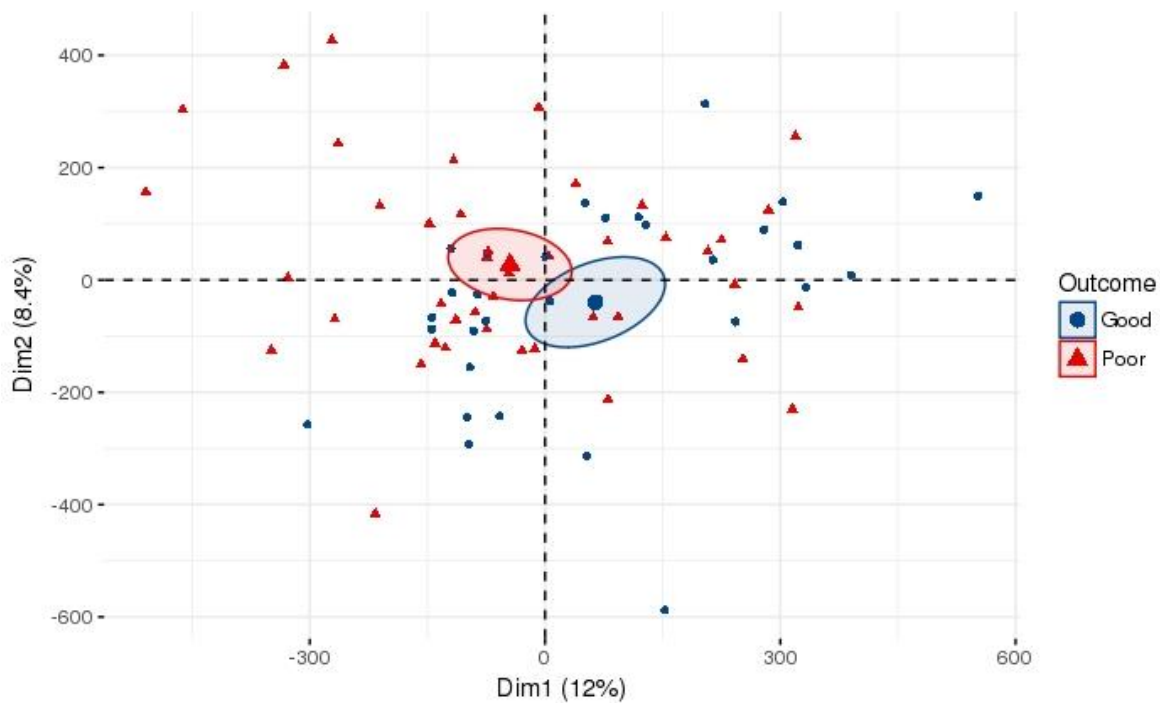


Figure 6.2 Unsupervised principal component analysis plot of poor and good outcome cases

Beta values from all 373,253 probes that passed QC from the HM450k across 73 samples were used in a PCA to determine sources of variation in the dataset. The variation in PC1, although marginally associated with outcome, was not sufficient to clearly separate poor and good outcome cases. This was not improved by including variation as part of PC2. The larger circle and triangle represent their respective group means and ellipses drawn around these averages are defined by 0.95 confidence interval.

While PCA in isolation did not effectively separate patient outcome groups, it did suggest that patient outcome contributed to variation in the data. In order to explore this further, global

methylation analysis of all 373,253 probes was employed to understand if areas of hypo/hypermethylation could be found, which in turn could infer the direction and magnitude of this variation. Overlaying the beta density distribution of diagnostic samples from poor and good outcome cases (Figure 6.3) revealed a slightly hypermethylated profile in poor outcome patients, with fewer hypomethylated probes and more partially methylated and hypermethylated probes. This is reflected in a 2% mean increase in methylation level of all probes ($p = 4.06 \times 10^{-131}$, Student's t-test).

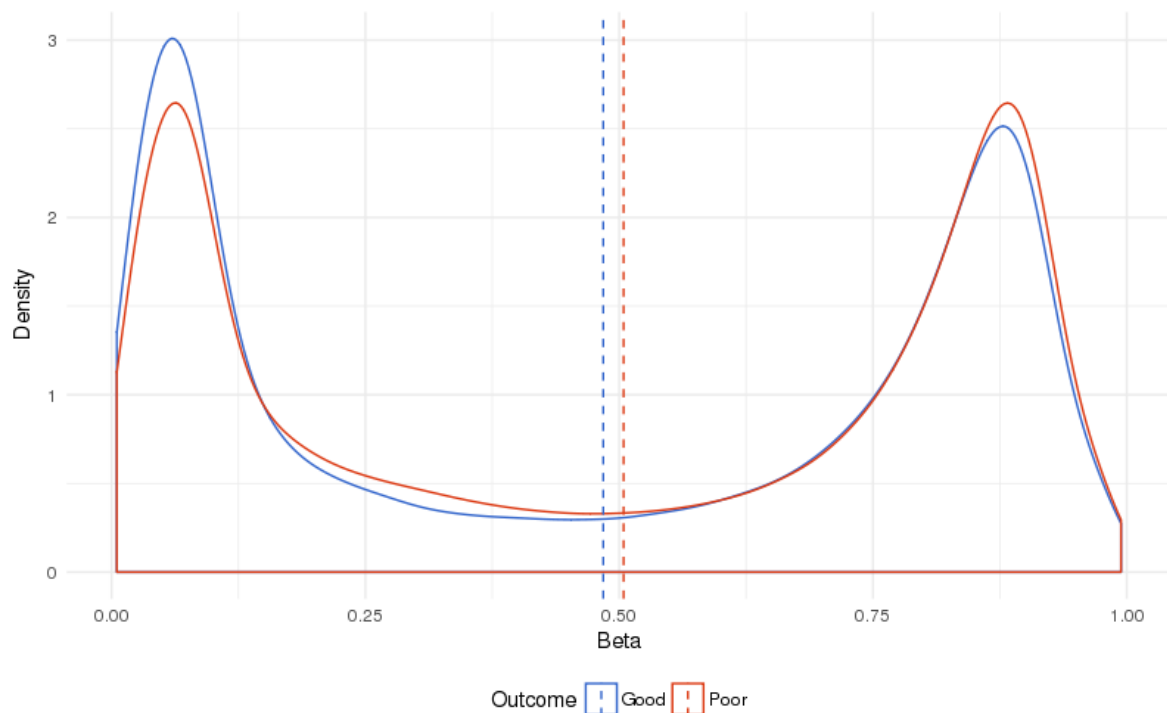


Figure 6.3 Overlaid density plot of methylation distribution between outcome groups

Visualisation of the average beta values from all 373,253 probes between outcome groups ($n=73$) shows a decrease in lowly methylated probes, a slight gain in intermediate methylated probes, and more hypermethylated probes for poor outcome patients. This suggests that some probes are differentially methylated at diagnosis for patients that would later go on to be considered a poor prognostic case. Outcome group means are defined by the dashed lines with a 2% increase in mean methylation in poor outcome patients on average ($p = 4.06 \times 10^{-131}$).

6.3.2 Probe & Gene Specific Differential DNA Methylation at Diagnosis Associated with Poor Patient Outcome

Linear regression analysis, utilising the ‘remove unwanted variance’ (RUV) R package, was used to further identify the sources of variation in the data associated with patient outcome. Regression analysis yielded 2046 statistically significant differentially methylated probes (DMPs, unadjusted $p \leq 0.01$) between good and poor outcome patients in diagnostic samples, however these did not survive correction for multiple-testing.

Of the total 2046 probes that were identified in regression analysis, 517 of these were found in open seas while the remaining 1529 were associated with 1292 genes as defined by the HM450k manifest. Several genes on this list were identified as potential contributors to disease progression and were taken into account for further analysis based on a criterion of number of significant probes, consistency of finding (low p-value), large differences between groups ($\Delta\beta$) or compelling evidence for a role in disease pathogenesis. Overall, 53 probes were selected, spanning 11 genes (listed in Appendix III) and are explored for their prognostic use in Section 6.3.3.

Clustering analysis based on all probes was largely unable to distinguish patients by outcome. By restricting analyses to the top 100 probes ($p \leq 9.11 \times 10^{-4}$) it was possible to partially separate two clusters of patients by outcome using unsupervised hierarchical clustering (Figure 6.4). Two of the three major clusters were associated with poor patient outcome, encompassing 37 of 43 patients, whereas the final cluster associated with good outcome encompassed 24 of 30 patients ($p = 1.602 \times 10^{-8}$, Fisher’s Exact Test). A total of 6 cases each of poor and good outcome did not co-cluster with similar outcome samples.

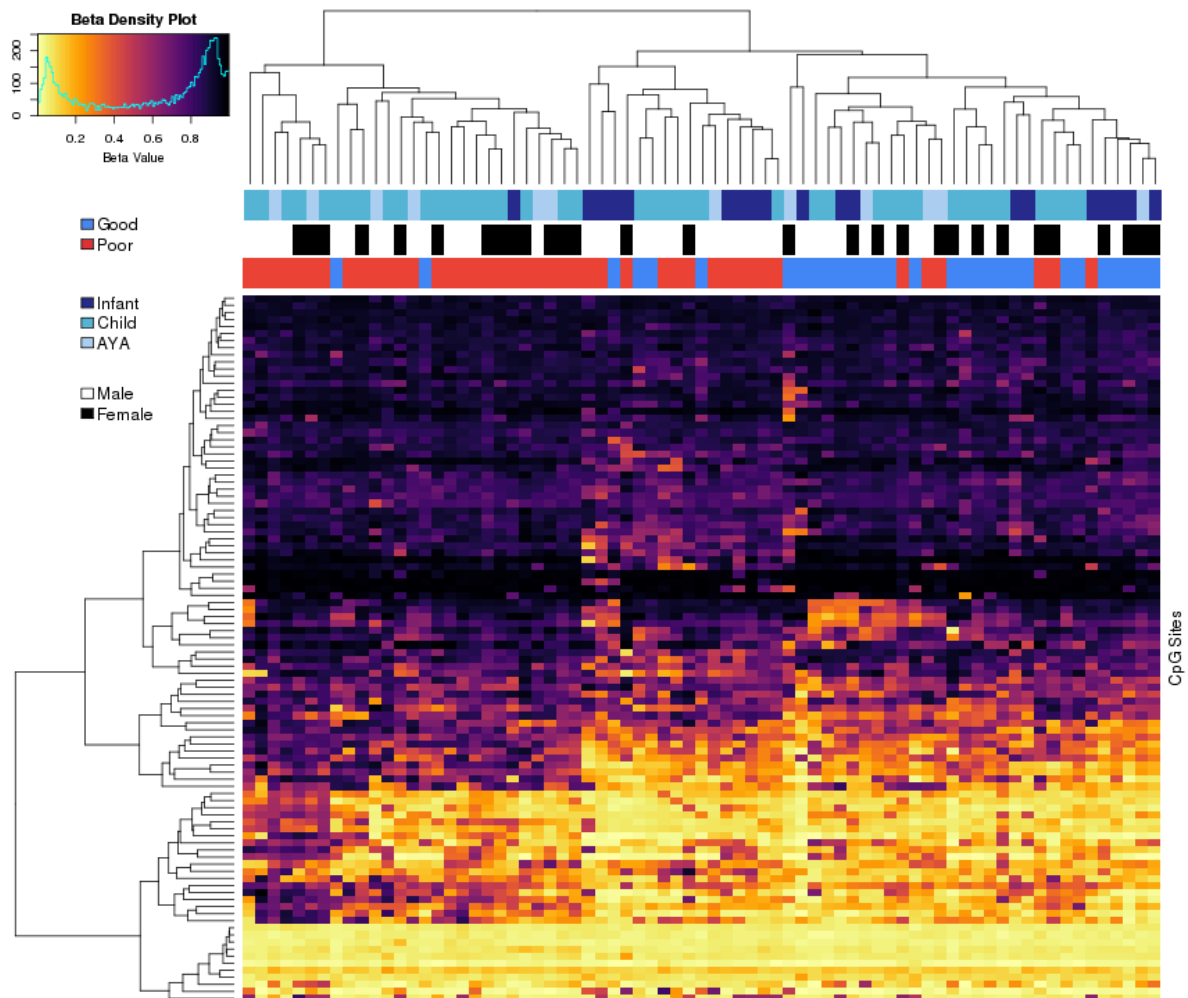


Figure 6.4 Heatmap of beta values for 73 AML patients across the 100 highest ranked probes associated with patient outcome

Heatmap with the y-axis representing the top 100 CpG sites while the x-axis shows 73 patient samples which are both ordered by hierarchical clustering as depicted by the dendrogram. Hypomethylation is depicted in yellow and a gradual increase to hypermethylation is indicated by purple. This heatmap showed that DNAm can fairly accurately define patient outcome by correctly predicting 61 out of 73 patient samples. Patient age group and gender were displayed to ensure that age and sex could not explain the discrepancies in the data.

6.3.3 Using the Top Differentially Methylated Probes to Predict Patient Outcome

The selection of features for predictive modelling is important as the more features that are included in the model the less predictive power that it will have. Furthermore, while many features can contribute individually to prognosis, redundancy can occur if features are not contributing unique information when combined in the same model. To test the prognostic potential of each probe previously identified to potentially contribute to disease, we firstly calculated the Pearson correlation coefficient of all probes previously outlined in Appendix III (Figure 6.5). We discovered that methylation levels at probes found within several genes (*MAEA* & *LAMP1*, *TXNRD1*, *FGFR3* and *PRR34*) were highly correlated (Figure 6.5) and therefore could be used as ‘tagprobes’ (one probe representing a correlated region of probes) for modelling purposes. That is, like a tagSNP, that any site within the correlated region would add the same information to a predictive model, thus one probe is used to encapsulate all probes within that region for predictive purposes. Probes cpg16, cpg7, cpg35 and cpg44 were selected as tagprobes as these had the largest average $\Delta\beta$ between poor and good outcome cases for their respective regions. The remaining probes, accounting for 6 genes, did not cluster together and were therefore unable to be tagged appropriately.

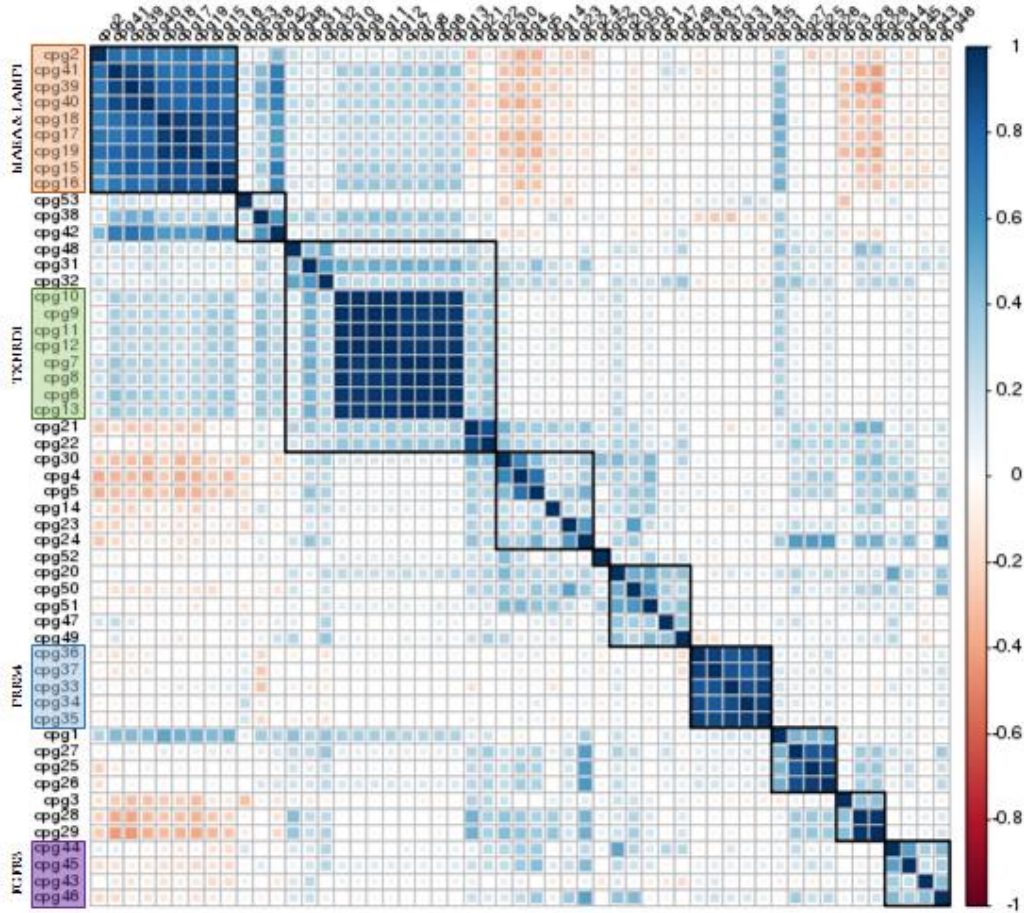


Figure 6.5 Correlation matrix of candidate probes based on disease progression

Correlation of probes ($n=53$) spanning eleven genes selected in section 6.3.2 as potentially effecting disease progression. Probes are ordered by hierarchical clustering and black boxes outline the top ten clusters. High positively correlated probes are in dark blue while dark red suggests high negative correlation. Highlighted areas show probes that are highly correlated and are from the same gene and are ideal candidates as ‘tagprobes’.

Receiver operating characteristic (ROC) analysis was undertaken for all probes identified in section 6.3.2 except for highly correlated probes that were removed and represented by the aforementioned tagprobes (Table 6.3; corresponding HM450K probe names can be found in Appendix III). CpG48 was identified as the optimal probe to distinguish patient outcome ($AUC = 0.770$) and was used as the foundation probe for building a multi-probe signature for ROC analysis. Each probe was sequentially added to the model (in descending order of AUC value) and was kept if it increased the overall AUC. A model including CpG48, CpG38, CpG49, CpG16, CpG3, CpG47 and CpG51 could accurately discriminate patient outcome in 91% of cases (Figure 6.6). Translation of this into a clinical setting would largely revolve around minimising false-negative predictions, therefore calculation of the Youden index (the

maximum effectiveness of a biomarker) will not yield the optimal cut-off value. To truly minimise false-negative rate while obtaining useful prognostic results, an optimal cut-off would be around 4.75 (alpha value) which allows for close to 90% sensitivity with a false positive rate around 20%. Validation of these probes and models in a larger cohort is necessary and will be addressed in section 6.4.3.

Table 6.3 Results of ROC analysis of individual candidate probes

Probe	$\Delta\beta$	AUC (95% CI)	Probe	$\Delta\beta$	AUC (95% CI)
cpg1	-0.205	0.705 (0.583 - 0.828)	cpg28	-0.218	0.6581 (0.528 - 0.788)
cpg2	-0.087	0.653 (0.519 - 0.787)	cpg29	-0.174	0.6543 (0.523 - 0.786)
cpg3	-0.028	0.704 (0.580 - 0.827)	cpg30	-0.074	0.6287 (0.500 - 0.758)
cpg4	0.018	0.585 (0.447 - 0.722)	cpg31	-0.143	0.7078 (0.587 - 0.829)
cpg5	0.016	0.571 (0.427 - 0.714)	cpg32	-0.220	0.7310 (0.616 - 0.846)
cpg7	-0.259	0.698 (0.576 - 0.821)	cpg35	0.151	0.6659 (0.538 - 0.794)
cpg16	-0.206	0.716 (0.593 - 0.839)	cpg38	-0.076	0.7519 (0.638 - 0.866)
cpg20	-0.031	0.723 (0.604 - 0.842)	cpg44	0.069	0.6760 (0.551 - 0.801)
cpg21	-0.182	0.698 (0.573 - 0.822)	cpg47	-0.015	0.6434 (0.512 - 0.775)
cpg22	-0.170	0.679 (0.554 - 0.804)	cpg48	-0.121	0.7698 (0.663 - 0.877)
cpg23	-0.022	0.647 (0.513 - 0.780)	cpg49	-0.030	0.7357 (0.620 - 0.851)
cpg24	-0.057	0.686 (0.560 - 0.812)	cpg50	-0.026	0.6256 (0.488 - 0.763)
cpg25	-0.085	0.631 (0.500 - 0.767)	cpg51	-0.051	0.6264 (0.496 - 0.757)
cpg26	-0.132	0.674 (0.544 - 0.803)	cpg52	-0.028	0.7000 (0.578 - 0.822)
cpg27	-0.115	0.671 (0.542 - 0.799)	cpg53	0.097	0.6636 (0.537 - 0.790)
Best combination of probes			AUC (95% CI)		
cpg48 + cpg38 + cpg49 + cpg16 + cpg3 + cpg47 + cpg51			0.908 (0.841 - 0.975)		

Abbreviations: $\Delta\beta$ - Delta beta (difference in average methylation between two groups); AUC - Area under the curve; CI - Confidence interval.

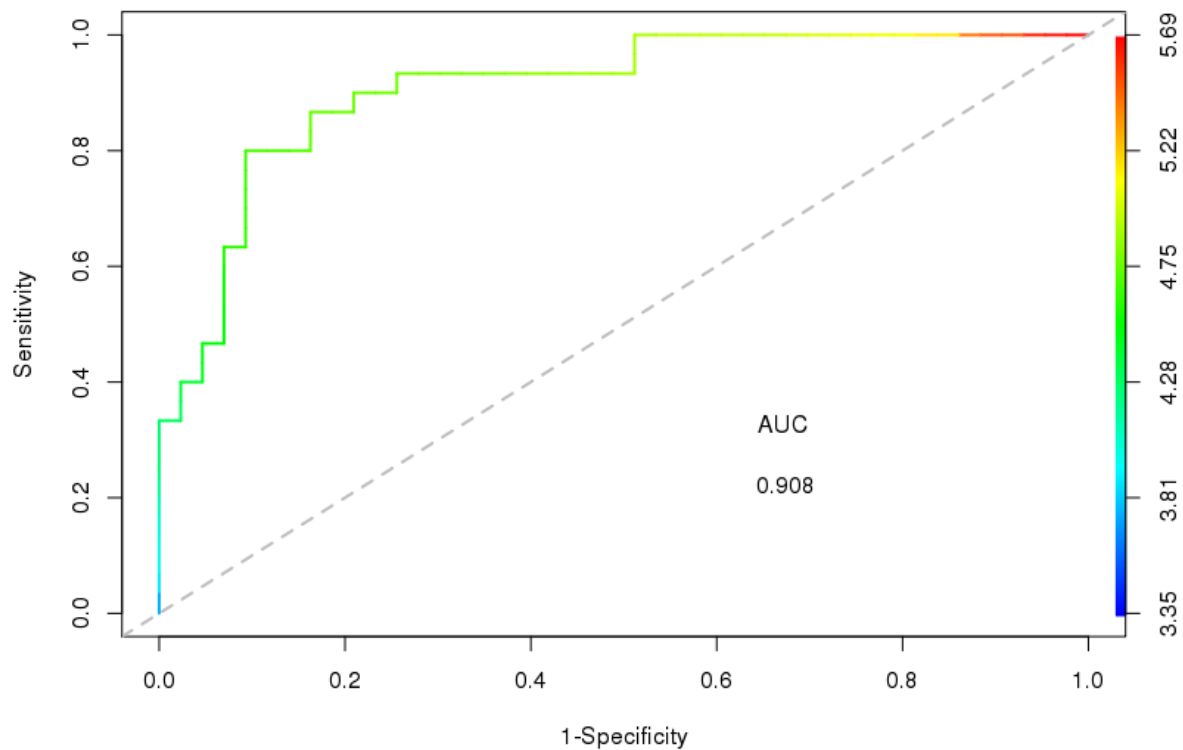


Figure 6.6 Receiver operator characteristic curve of optimum combination of probes to predict patient outcome

With a combination of probe CpG48, CpG38, CpG49, CpG16, CpG3, CpG47 and CpG51 we could accurately predict outcome in 91% of patients ($n=73$, $AUC=0.908$). True positive rate is plotted against false positive rate and the curve is coloured based on cut-off (alpha value).

6.4 Whole-genome DNA Methylation Analysis to Predict Patient Outcome: A Pooled-Analysis

6.4.1 DNA Methylation to Define Patient Outcome in a Combined AML Cohort

With the addition of the TARGET AML cohort epidemiological data such as age, sex and cytogenetics was again analysed for prognostic utility (Table 6.4). Fisher's exact test revealed that patient age group was associated with patient outcome ($p = 0.002$), however this may be due largely to the increase in poor outcome patients, especially AYA age grouped, that the TARGET cohort was built around. Sex and cytogenetics of the patients continued to show no association with patient outcome, confirming the original findings in the discovery cohort.

Receiver operating characteristic analysis showed that there is little predictive value for any epidemiological data, again confirming the original findings despite the newfound age association. Combinations of these data, while an improvement upon their individual prognostic value, still remain unsuitable for prognostic testing. Furthermore, none of the added data altered findings in survival analysis (data not shown), therefore giving more credence to the idea that while relapse is a main cause of death, cytogenetics does not overly dictate who will suffer a recurrence.

Table 6.4 Results of Fishers exact test and ROC analysis for clinical characteristics upon patient outcome in a pooled cohort

Clinical Characteristics	p-value	AUC (95% CI)
Sex (male, female)	0.254	0.5612 (0.4701 - 0.6524)
Age at diagnosis (infant, child, AYA)	0.002	0.6120 (0.5225 - 0.7015)
Patient cytogenetics*	0.141	0.6071 (0.5018 - 0.7124)
All combined	N/A	0.6463 (0.5441 - 0.7485)

Abbreviations: AUC - Area under the curve; AYA - Adolescence/Young adult.

*Patient cytogenetics was split into inv(16)/t(16;16), t(8;21), *KMT2A* rearranged, trisomy 8, CN, complex and all others fell into category 'other'.

Combining the discovery AML cohort with the TARGET AML dataset, which includes a further 79 poor and 6 good outcome patients, increased the sample size to 158.

Unsupervised principal components analysis was undertaken on the total cohort and revealed that patient outcome ($p=2.0 \times 10^{-8}$, Student's asymptotic p-value for correlation) was a component of PC2, contributing 9.5% of the variation in the dataset. Similar to the Australian cohort analysis, this was confounded by chip ($p = 6.0 \times 10^{-16}$) and where the sample originated from ($p = 9.0 \times 10^{-15}$). Figure 6.7 shows dimensionality reduction obtained using methylation values of probes contributing to variation in PC1 and PC2 only. While the pooled-analysis has improved the ability to distinguish patient outcome groups, particularly through PC2, sample overlap is still apparent and cannot solely distinguish patient outcome.

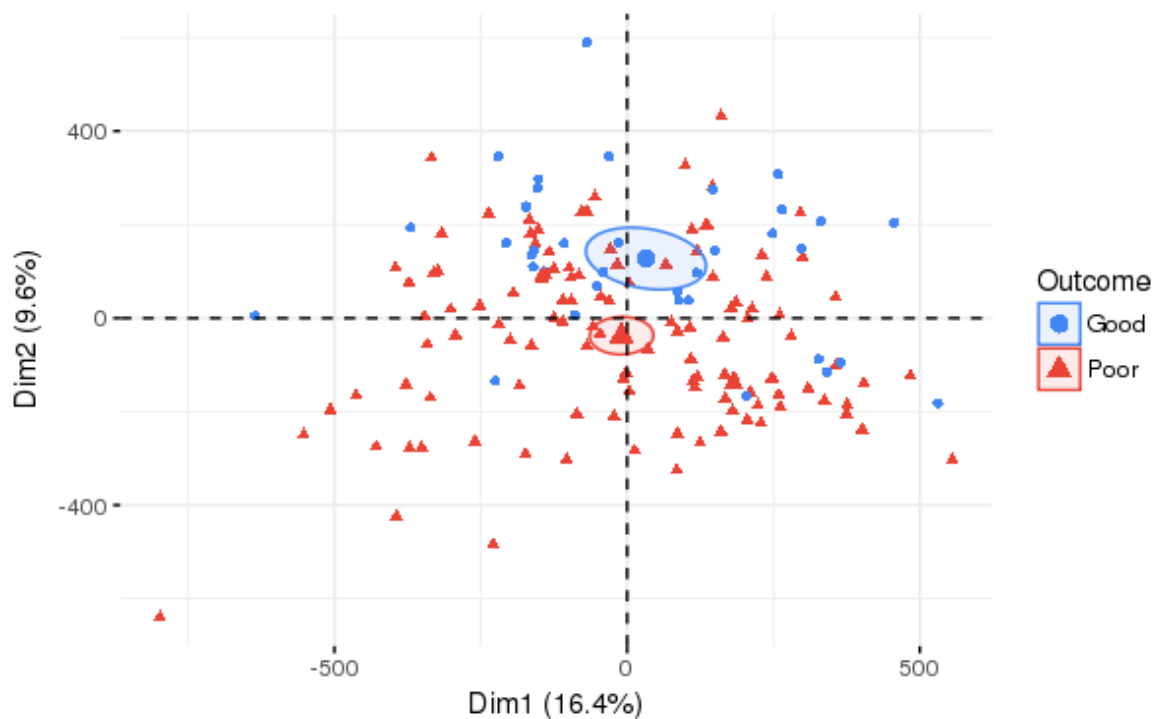


Figure 6.7 Unsupervised principal components analysis plot to define patient outcome in a pooled cohort

Beta values from 373,253 HM450K probes remaining following quality control across 158 samples were used in a PCA to determine sources of variation in the dataset. The variation in PC2 (accounting for 9.5% of the total variation) was partly attributed to patient outcome but this is marginal with little predictive utility (as exemplified by substantial overlap between poor and good outcome cases on the y-axis). The larger blue dot and red triangle represent their respective group means and ellipses drawn around these define the 0.95 confidence intervals.

Global methylation analysis of all 373,253 probes was employed to distinguish global methylation differences between patient outcome groups. Overlaid inspection of beta density (Figure 6.8) supports the previously identified hypermethylated profile in poor outcome

patients with a 2% mean increase ($p = 6.485 \times 10^{-138}$, Student's t-test) in global methylation. In contrast to the Australian only AML cohort, the pooled-analysis revealed a larger density of hypermethylated probes, but also alluded to some more hypomethylated probes in poor outcome cases too. Good outcome cases remained relatively unchanged.

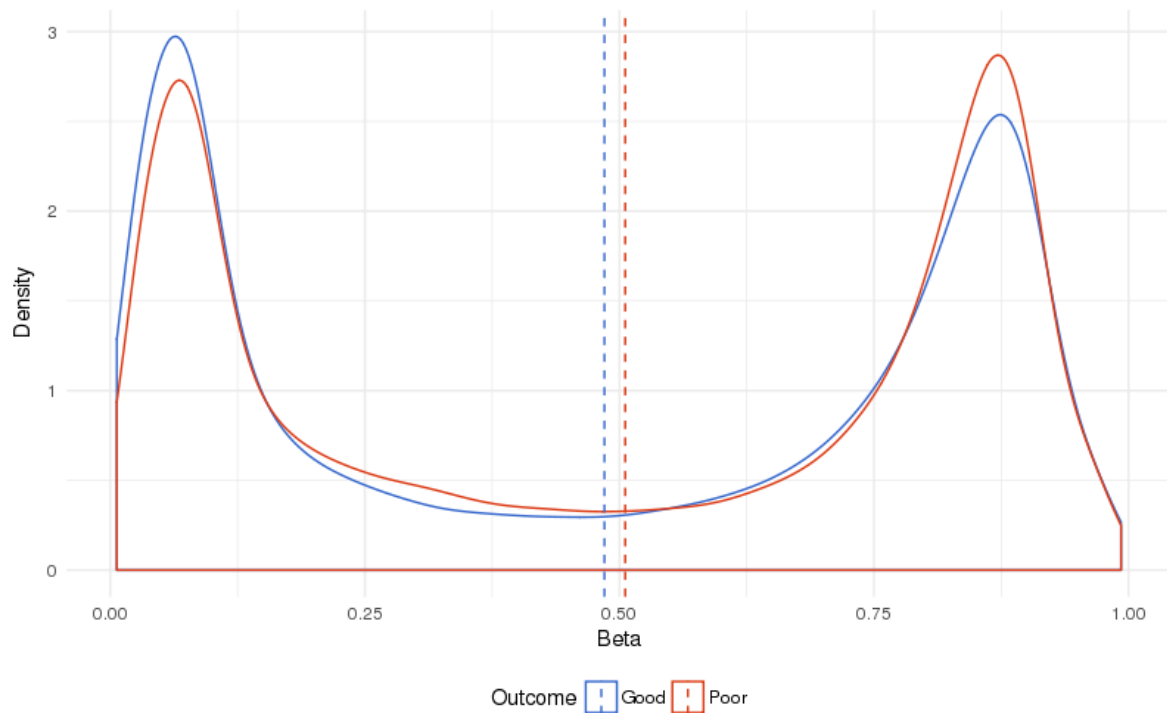


Figure 6.8 Overlaid density plot to interrogate global methylation differences between outcome groups in a pooled cohort

Averaging the beta values from all 373,253 probes between outcome groups ($n = 158$) shows an increase in lowly methylated probes and a slight loss in medium methylated probes for good outcome patients. Pooled-analysis shows that this is a consistent finding as numbers of highly methylated probes in poor outcome patients only continues to rise. Outcome group means are defined by the dashed lines with a 2% increase in average methylation in poor outcome patients ($p = 6.485 \times 10^{-138}$).

6.4.2 Probe & Gene Specific Differential DNA Methylation

Underpinning Poor Patient Outcome: A Pooled-Analysis

Similarly to the previous analysis, linear regression was used, utilising the ‘remove unwanted variance’ (RUV) R package, to identify the most differentially methylated probes between patient outcome groups. Regression analysis yielded 4036 statistically significant differentially methylated probes (DMPs, unadjusted $p \leq 0.01$) between outcome groups, however these did not survive correction for multiple-testing. Of the total 4036 probes that were identified in regression analysis, 866 probes were found in open seas while the remaining 3170 probes were associated to 2485 genes. A total of 3419 probes were specific to the combined analysis, while 1429 were identified only in the Australian samples. A total of 617 probes were found in both analyses (Figure 6.9). This included only 52 of the top 100 from the Australian cohort, highlighting the impact of sample size on this type of differential methylation analysis.

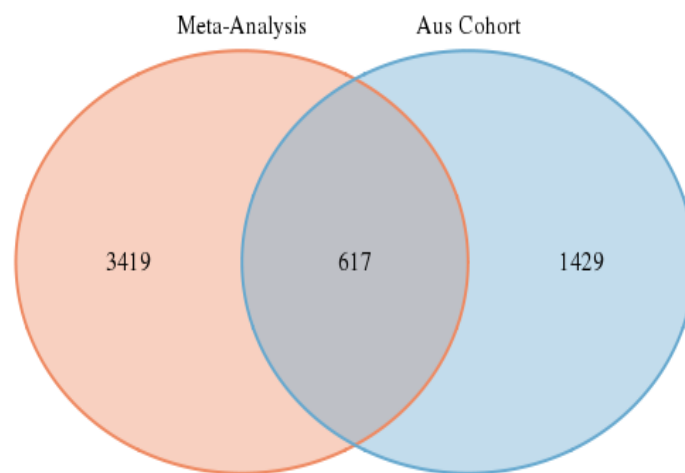


Figure 6.9 Venn diagram of significantly differentially methylated probes between the discovery and combined cohorts

In the pooled-analysis 4036 probes were identified as differentially methylated between patient outcome groups. Of these, 617 probes were common to the previous Australian only analysis.

Clustering analysis was undertaken on all significant probes as well as the 617 common probes between analyses, however this revealed little ability to distinguish patient outcome. By restricting to the top 500 differentially methylated probes ($p \leq 8.08 \times 10^{-4}$) of the current regression analysis we were able to generate a DNAm signature that best distinguished patients by

outcome using unsupervised hierarchical clustering (Figure 6.10). Two primary clusters appeared with cluster 1 featuring 81% of the good outcome cases, but also 43% of the poor outcome cases. Cluster 2 had the remaining 57% of poor outcome cases and 19% of the good outcome cases too ($p=5.704 \times 10^{-5}$, Fisher's Exact Test).

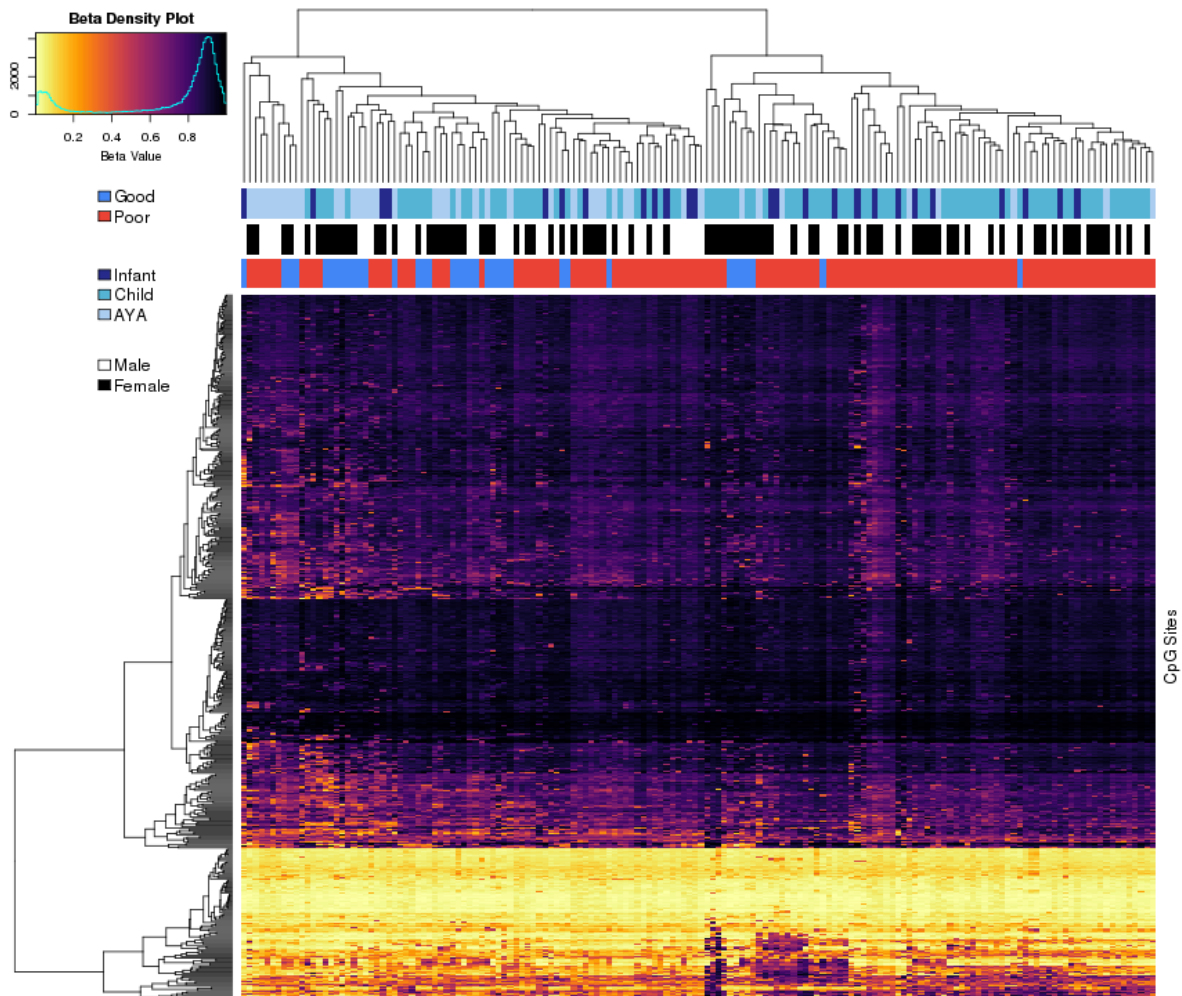


Figure 6.10 Heatmap of beta values for 158 AML patients across the 500 highest ranked probes associated with patient outcome

Heatmap with the y-axis representing 500 CpG sites while the x-axis shows 158 patient samples which are both ordered by hierarchical clustering as depicted by the dendrograms. This heatmap showed that DNAm can fairly accurately define patient outcome by distinguishing 29/36 good outcome patients in cluster 1 and 70/122 poor patient outcome samples in cluster 2. Patient age group and gender were displayed to ensure that age and sex could not explain the discrepancies in the data.

Particular genes that were identified as potential contributors to disease were selected on the same criteria denoted in section 6.3.2. All probes selected can be found in Appendix IV.

Genes of particular interest identified in the discovery analysis and were supported in the pooled-analysis include *MIRLET7A3*, *PTPRN2* and *PAX6*. Correlation analysis was performed on each of these genes/probes to see if the sites could be averaged and used as a singular DMR (Figure 6.11). Probes within *MIRLET7A3* and *PAX6* were all highly correlated. Subsequent DMRs for *MIRLET7A3* ($p=2.54 \times 10^{-4}$, Student's t-test) and *PAX6* ($p=7.38 \times 10^{-9}$, Student's t-test) were significantly differentially methylated, however despite adding more probes from the original analysis, *PTPRN2* remained uncorrelated.

Other genes from Appendix IV were validated in this study, including *BAHCC1* and *PRR34*. *FGFR3* was validated and included several more probes than previously implicated suggesting a key role for this gene in disease progression and patient relapse. *LAMP1* was only slightly validated with 2 out of 4 probes appearing in both analyses, however the difference in methylation was smaller suggesting *LAMP1* could be an artefact of the Australian cohort and is not entirely recapitulated in the total paediatric AML cohort. Unfortunately, many probes from the Australian cohort analysis did not validate including *RUNX3*, *TXNRD1*, *HOXB3* and *MAEA*. With the additional samples from the TARGET AML study we can be more confident with the robustness of these findings implicating aberrant DNAm in different patient outcomes at diagnosis.

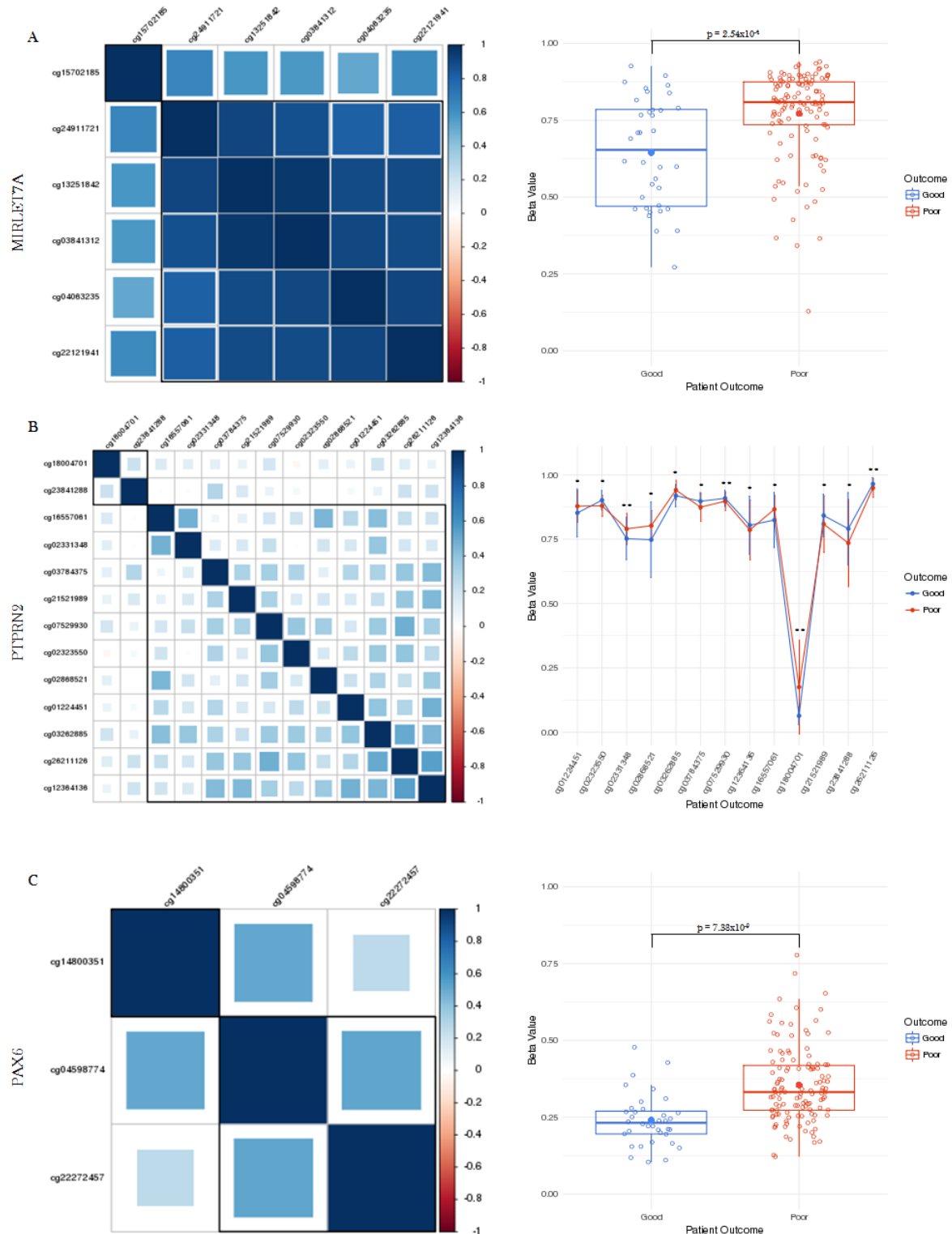


Figure 6.11 Correlation matrix and boxplot/line graph defining differential methylation between pooled patient outcome groups in MIRLET7A, PTPRN2 and PAX6

Caption on next page.

Correlation matrices show the correlation between statistically significant probes within a single gene. Highly correlated probes are averaged and presented as boxplots, while uncorrelated probes are shown as line graphs. Boxplots show the median, 0.25/0.75 IQR, minimum and maximum methylation within patient outcome groups (n=158). The mean is represented by the coloured circle in the middle of the boxplot and outliers lay beyond the whiskers. Line graph show each probe with SD represented by the error bars. A) *MIRLET7A3* had 6 probes, all of which were highly correlated and averaged between patient outcome groups ($p = 2.54 \times 10^{-4}$, Student's t-test). B) *PTPRN2* was validated and included 5 new probes totalling at 13, none of which were correlated. Each probe was averaged between patient outcome groups and visualised (* $p \leq 0.05$, ** $p \leq 0.005$, Student's t-test). C) *PAX6* had 3 probes all of which were correlated and averaged between patient outcome groups ($p = 7.38 \times 10^{-9}$, Student's t-test).

6.4.3 Comparing the Top Differentially Methylated Probes to Predict Patient Outcome: A Pooled-Analysis

Probes that were included in previous ROC analysis of Australian only cases (section 6.3.3) that also survived linear regression analysis in the combined cohort (section 6.4.2) were tested in ROC analysis in the combined cohort to test their predictive value (Table 6.5). Only 14 of 53 probes previously identified probes met these criteria and were included in ROC analysis. Two probes (cpg32 and cpg43) increased their predictive value while the remaining 12 probes all showed less predictive utility in the combined cohort. The average change in AUC (Δ AUC) for these 12 probes was -0.05. Cpg23 was the poorest performer in the combined analysis with a Δ AUC value of -0.112. The $\Delta\beta$ or difference of $\Delta\beta$ between analyses did not influence the AUC, nor did it affect the Δ AUC between analyses. A ROC analysis of the probes that created the optimal prognostic signature in Australian only samples (section 6.4.2) yielded a reduced area under the curve and predictive analysis in the combined cohort (Δ AUC=-0.1184, ROC analysis). As such, a new model encompassing probes solely from the combined analysis was generated.

Table 6.5 ROC analysis in the combined sample of probes from Australian only ROC analysis

Probe	$\Delta\beta$	$\Delta\Delta\beta^*$	AUC (95% CI)	ΔAUC^*
cp15	-0.097	0.048	0.627 (0.520 – 0.743)	-0.090
cp23	-0.015	0.008	0.535 (0.413 – 0.657)	-0.112
cp32	-0.210	0.010	0.763 (0.684 – 0.843)	0.032
cp33	0.096	-0.020	0.607 (0.491 – 0.722)	-0.059
cp34	0.104	-0.025	0.591 (0.473 – 0.591)	-0.075
cp35	0.127	-0.024	0.626 (0.514 – 0.733)	-0.040
cp36	0.091	-0.001	0.636 (0.524 – 0.747)	-0.030
cp37	0.105	0.011	0.635 (0.523 – 0.746)	-0.031
cp43	0.036	-0.018	0.696 (0.485 – 0.700)	0.020
cp44	0.030	-0.040	0.592 (0.485 – 0.592)	-0.084
cp45	0.039	-0.010	0.638 (0.531 – 0.745)	-0.038
cp46	0.022	-0.010	0.596 (0.494 – 0.698)	-0.080
cp49	-0.022	0.007	0.679 (0.583 – 0.775)	-0.057
cp53	0.055	-0.041	0.612 (0.504 – 0.720)	-0.052
Validation of best combination of probes			AUC (95% CI)	ΔAUC^*
cp48 + cp38 + cp49 + cp16 + cp3 + cp47 + cp51			0.789 (0.712 – 0.867)	-0.1184

Abbreviations: $\Delta\beta$ - Delta beta (difference in average methylation between two groups); AUC - Area under the curve; CI - Confidence interval.

* $\Delta\text{AUC}/\Delta\Delta\beta$ is calculated by Table 6.5 AUC/ $\Delta\beta$ minus Table 6.3 AUC/ $\Delta\beta$ for the same probe (or tagprobe).

Feature selection using correlation analysis was again undertaken upon a new set of probes that were identified during regression analysis (Appendix IV). Pearson correlation coefficient was calculated for each pair of probes and ordered by unsupervised hierarchical clustering (Figure 6.12). Probes in *PAX6*, *FLOT1*, *MIRLET7A3* and *PPR34* genes were all highly correlated. *FGF3* was slightly correlated with many probes of other genes. Other genes such as *GRLF1* and *PTPRN2* were not particularly well correlated. Tagprobes for the aforementioned correlated genes were used to represent the region in predictive modelling. Cp179, cp185, cp190 and cp194 were all used to represent probes originating from *PPR34*, *MIRLET7A3*, *PAX6* and *FLOT1* respectively. *FGFR3* was deemed to be insufficiently correlated to have a tagprobe represent the gene.

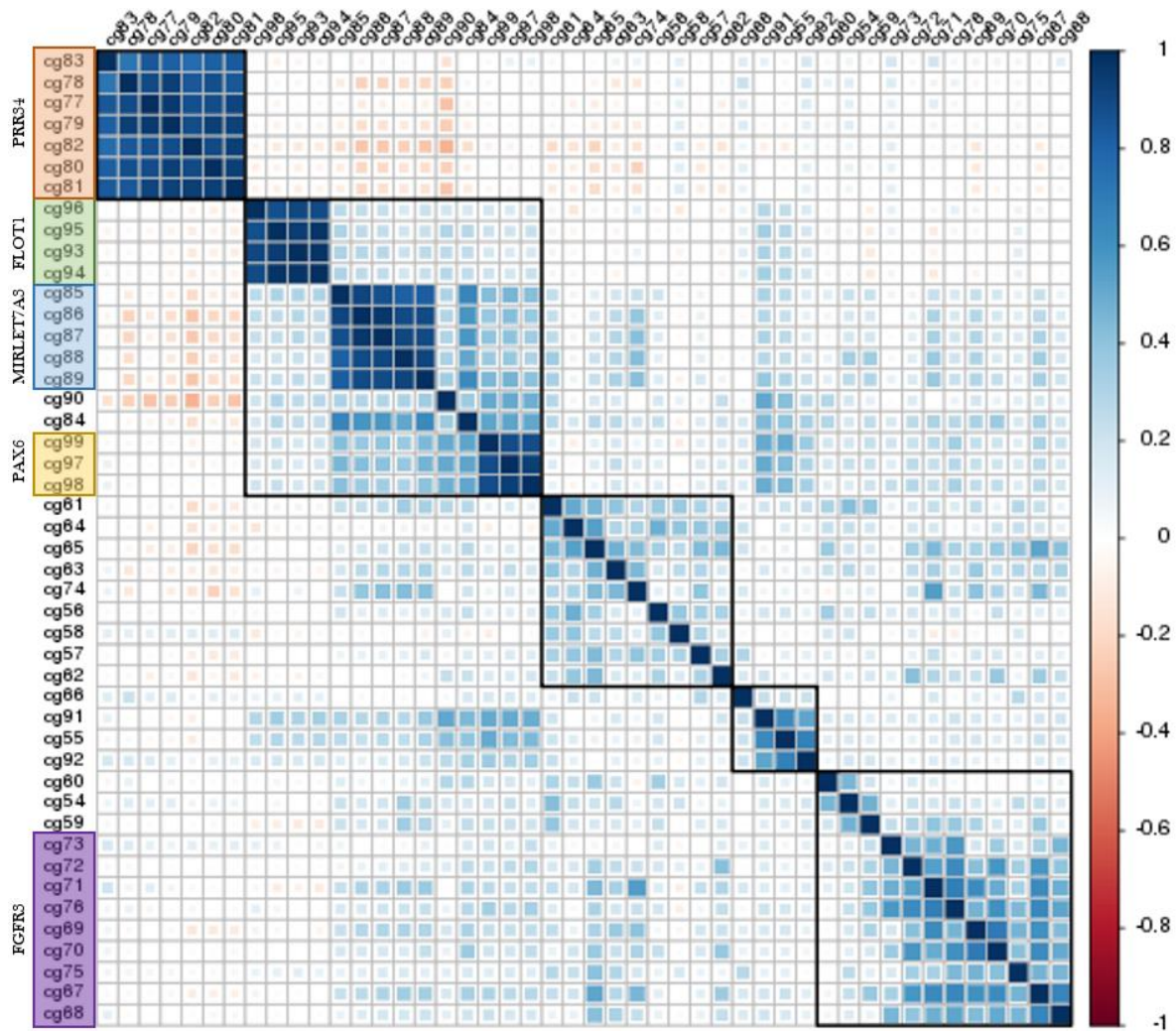


Figure 6.12 Correlation matrix of candidate probes after pooled linear regression based on disease progression

Correlation of probes ($n=158$) spanning seven genes selected in section 6.4.2 as potentially effecting disease progression. Probes are ordered by hierarchical clustering and black boxes outline the top five clusters. High positively correlated probes are in dark blue while dark red suggests high negative correlation. Highlighted areas show probes that are highly correlated and are from the same gene and are ideal candidates as tagprobes.

Receiver operating characteristic analysis was undertaken for probes identified as potentially linked to disease progression and relapse in section 6.4.2. Highly correlated probes were removed from this analysis and tagprobes were used to represent the region as previously outlined. The best probe at distinguishing patient outcome was cpg99 with an AUC of .765 and this was therefore used as the foundation for a multi-probe model. Combining data for cpg99, cpg55, cpg90, cpg94, cpg63, cpg61 and cpg59 resulted in an optimal AUC of 0.848 (Figure 6.13). A cut-off value of 3.45 ± 0.5 (alpha value) would be an optimal cut-off as this has

a sensitivity of around 0.8 while specificity is restricted close to approximately 1 in 5 false positives.

Table 6.6 Results of ROC analysis of candidate probes based on disease progression after pooled-analysis

Probe	$\Delta\beta$	AUC (95% CI)	Probe	$\Delta\beta$	AUC (95% CI)
cpg54	-0.042	0.580 (0.461 - 0.698)	cpg69	0.017	0.479 (0.370 - 0.588)
cpg55	-0.111	0.761 (0.679 - 0.843)	cpg70	0.037	0.687 (0.592 - 0.782)
cpg56	0.011	0.616 (0.501 - 0.731)	cpg71	0.030	0.592 (0.485 - 0.700)
cpg57	0.023	0.651 (0.541 - 0.760)	cpg72	0.039	0.638 (0.531 - 0.745)
cpg58	0.022	0.658 (0.553 - 0.764)	cpg73	0.022	0.596 (0.494 - 0.698)
cpg59	-0.038	0.627 (0.511 - 0.743)	cpg74	0.015	0.717 (0.627 - 0.806)
cpg60	-0.055	0.559 (0.431 - 0.686)	cpg75	0.018	0.607 (0.505 - 0.708)
cpg61	-0.022	0.679 (0.583 - 0.748)	cpg76	0.014	0.505 (0.400 - 0.610)
cpg62	0.033	0.567 (0.460 - 0.675)	cpg79	0.127	0.626 (0.514 - 0.737)
cpg63	-0.026	0.566 (0.451 - 0.681)	cpg85	-0.209	0.695 (0.592 - 0.798)
cpg64	0.015	0.635 (0.529 - 0.741)	cpg90	-0.162	0.743 (0.647 - 0.839)
cpg65	0.018	0.460 (0.341 - 0.578)	cpg94	-0.107	0.681 (0.581 - 0.782)
cpg66	0.055	0.612 (0.504 - 0.720)	cpg97	-0.190	0.720 (0.620 - 0.821)
cpg67	0.011	0.494 (0.391 - 0.597)	cpg98	-0.237	0.724 (0.626 - 0.823)
cpg68	0.036	0.696 (0.598 - 0.794)	cpg99	-0.232	0.765 (0.678 - 0.852)
Best combination of probes			AUC (95% CI)		
cpg99 + cpg55 + cpg90 + cpg94 + cpg63 + cpg61 + cpg59			0.848 (0.773 - 0.924)		

Abbreviations: $\Delta\beta$ - Delta beta (difference in average methylation between two groups); AUC - Area under the curve; CI - Confidence interval.

AUC is calculated from ROC analysis

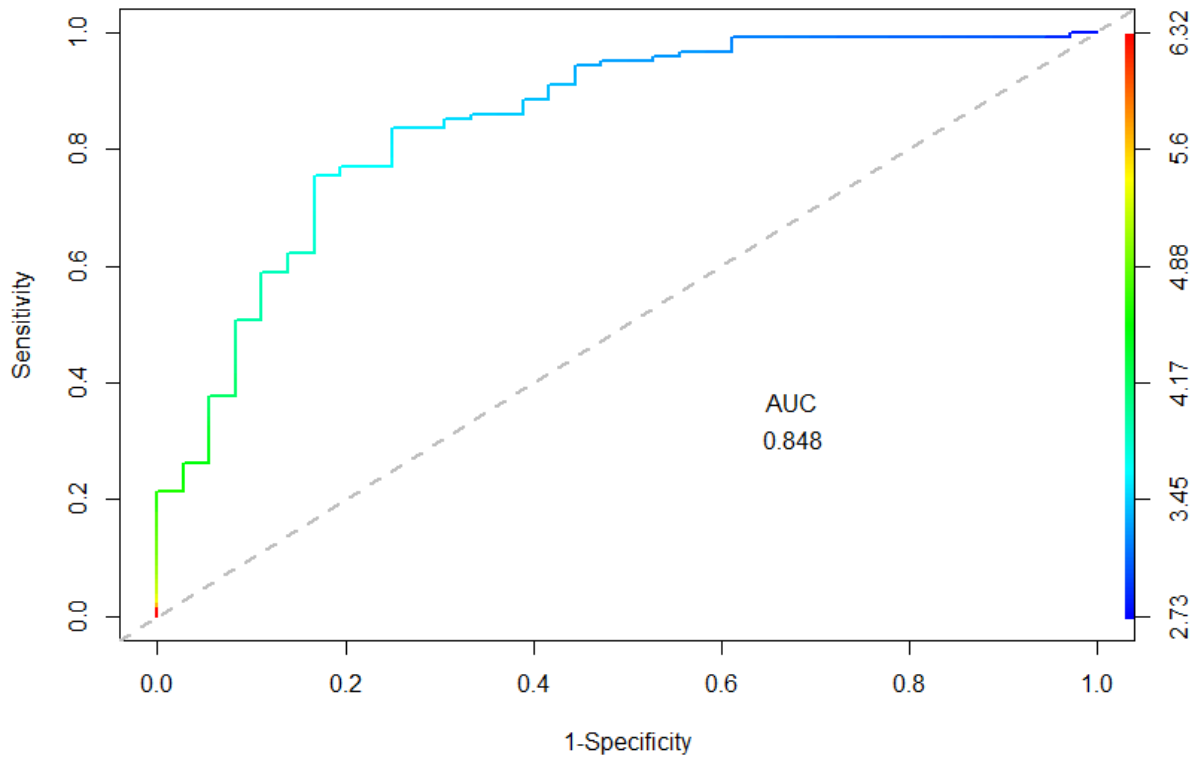


Figure 6.13 Receiver operator characteristic curve of optimum combination of probes to predict patient outcome from pooled linear regression model

With a combination of probe CpG 99, CpG 55, CpG 90, CpG 94, CpG 63, CpG 61 and CpG 59 we were able to predict 84% of patient samples ($n = 158$, $AUC = 0.848$). True positive rate is plotted against false positive rate and the curve is coloured based on cut-off (alpha value).

6.5 Machine-Based Learning Approaches to Patient Outcome Prediction

6.5.1 Finding the Appropriate Machine-Based Learning Algorithm

Machine-based learning (MBL) is an umbrella term for multiple different techniques that can be used to create predictive, self-improving models. We explored several of the prominent methods of MBL contained in the Weka software package. These included J48 decision trees, random forest (RF) analysis, library for single vector machine (LibSVM), decision table (DT) and naïve bayes (NB). Each method was implemented twice, one with and without feature selection. Original models were designed to identify the prognostic utility of findings from the discovery cohort and its application to the greater population. Therefore, models were trained on the 2046 differentially methylated probes identified in the Australian AML cohort regression analysis and were tested on the TARGET AML cohort (n=85). Specifics for each stage in the Weka workflow is detailed in Section 2.5.9.

Precision (positive predictive value), recall (true positive rate), weighted average false positive rate (FPR), F-measure (harmonic average of precision and recall used to determine a tests accuracy) and AUC were calculated for each test result (Table 6.7). A feature selected J48 algorithm performed the best (AUC=0.666) with a moderate F-measure of 0.864, closely followed by both RF analyses. LibSVM and DT underwhelmed as classifiers primarily due to high false positive rates. While J48 and RF algorithms performed best as defined by F-measure, their predictive measures underperformed relative to regression-based ROC analysis, suggesting that the results derived from the discovery AML cohort alone is unsuitable to predict outcome using MBL approaches.

Table 6.7 Weighted average model evaluation of machine-based learning algorithms upon TARGET AML dataset using 2046 differentially methylated probes

Machine-Based Learning Algorithm	Feature Selected	Precision	Recall	False Positive Rate	F-Measure	AUC*
J48 Decision Tree	Y	0.905	0.835	0.475	0.864	0.666
	N	0.879	0.741	0.636	0.798	0.575
Random Forest	Y	0.895	0.553	0.342	0.657	0.577
	N	0.945	0.941	0.775	0.921	0.588
Library for Support Vector Machines	Y	0.882	0.882	0.779	0.882	0.552
	N	N/A	0.929	0.929	N/A	0.500
Decision Table	Y	0.859	0.859	0.935	0.859	0.501
	N	0.862	0.906	0.931	0.884	0.453
Naïve Bayes	Y	0.870	0.365	0.356	0.474	0.569
	N	N/A	0.929	0.929	N/A	0.500

Abbreviations: AUC - Area under the curve; *AUC is calculated from ROC analysis

The same MBL approaches were then tested on the combined TARGET and Australian AML cohorts utilising the 4036 differentially methylated probes identified in the pooled-analysis from section 6.4.2. Weighted average FPR, precision, recall, F-measure and AUC was calculated for each test result (Table 6.8). Again, DT and LibSVM underperformed in both feature and non-feature selected analyses. J48 decision tree analysis did not greatly improve from the previous analysis. Naïve bayes algorithm without feature selection performed well with an AUC of 0.919 and a FPR of 0.129. However, the best performing model was the non-feature selected random forest algorithm (AUC=0.931, Figure 6.14).

Table 6.8 Weighted average model evaluation of machine-based learning algorithms using 5-fold validation of 4036 differentially methylated probes

Machine-Based Learning Algorithm	Feature Selected	Precision	Recall	False Positive Rate	F-Measure	AUC
J48 Decision Tree	Y	0.755	0.759	0.463	0.757	0.626
	N	0.755	0.759	0.463	0.757	0.647
Random Forest	Y	0.690	0.734	0.646	0.704	0.719
	N	0.873	0.848	0.515	0.817	0.931
Library for Support Vector Machines	Y	0.832	0.785	0.729	0.808	0.528
	N	N/A	0.772	0.772	N/A	0.500
Decision Table	Y	0.738	0.755	0.644	0.746	0.563
	N	0.699	0.722	0.591	0.715	0.652
Naïve Bayes	Y	0.716	0.722	0.533	0.719	0.667
	N	0.964	0.962	0.129	0.961	0.919

Abbreviations: AUC - Area under the curve; *AUC is calculated from ROC analysis

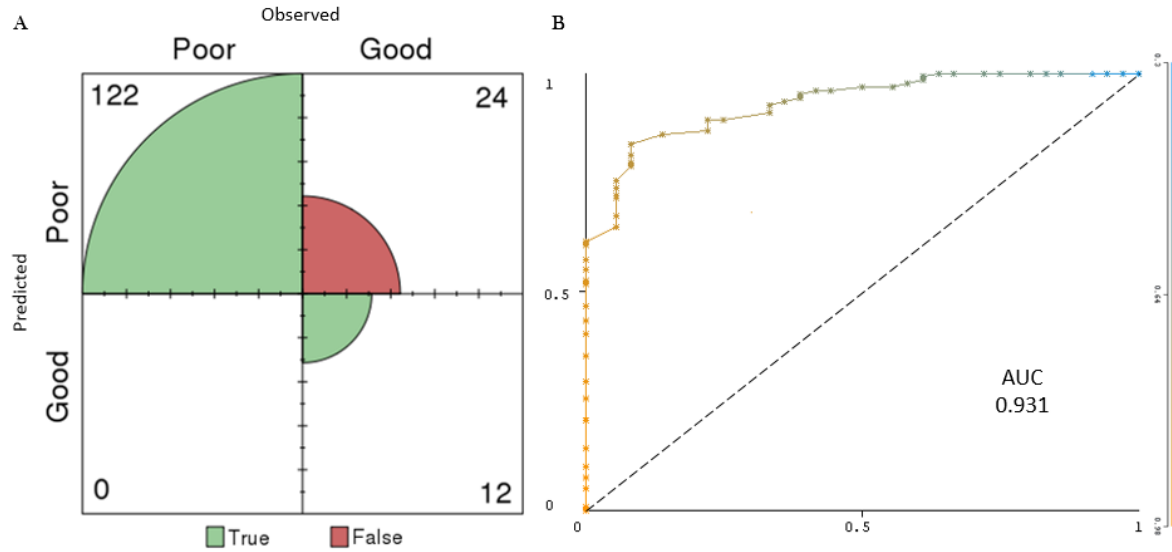


Figure 6.14 Evaluation of random forest model using 4036 differentially methylated probes through confusion matrix and ROC curve

Using the 4036 differentially methylated probes in a random forest analysis to predict patient outcome in a pooled cohort (n=158). A) RF analysis correctly identified all poor outcome and correctly identified 12/36 good outcome cases. B) Overall the predictive value of this model returned an AUC of 0.931

6.5.2 Using Random Forest Analysis to Predict Patient Outcome

Having identified random forest as the optimal machine-based learning algorithm for patient outcome prediction in a combined AML cohort, a more refined RF model was created with greater control over the variables that underpin it. This was done to minimise over-fitting of the model to the specific available dataset. Models were generated from the same probes 4036 that were identified in section 6.5.1.

Hyperparameters such as mtry (the number of probes to be selected from the full dataset at each split of a decision tree) and ntree (the amount of decision trees to be 'grown' in the random forest) are key variables in ensuring the model works in the most efficient and balanced manner for any particular dataset. Figure 6.15 illustrates that the optimal combination of mtry and ntree is 18 and 500 respectively for a RF algorithm to be most effective upon this dataset. The resulting prediction will be using these hyperparameters. Having mtry set to anything between 10 and 15 is mostly unaffected by ntree however this begins to show a larger variable effect as mtry is increased beyond 15.

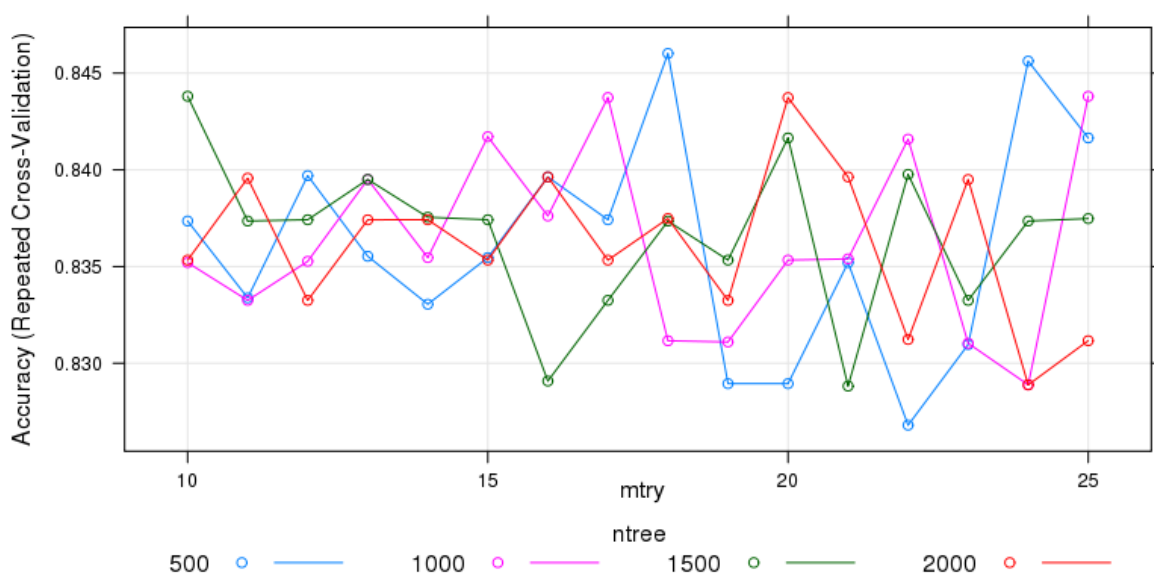


Figure 6.15 Hyperparameter random forest model tuning of ntree and mtry

Line graph illustrating the accuracy of model prediction across a 5-fold cross validated, 3 times repeated dataset (n=158) utilising different hyperparameters for both mtry and ntree. Optimum accuracy is achieved through a hyperparameter combination of mtry=18 and ntree=500.

Once the appropriate model variables were defined, the model was then trained and tested, utilising a more stringent, repeated k-folds cross validation ($k=5$, $\text{repeats}=3$), upon the entire patient cohort ($n=158$). The model created has a comparable AUC and precision suggesting that hyperparameter tuning does not negatively affect the predictive value of the model ($\Delta\text{AUC}=-0.0169$, $\Delta\text{Precision}=-0.0393$), however recall (positive prediction) is improved ($\Delta\text{Recall}=+0.1247$), which ultimately increases the F-measure of the model ($\Delta\text{F-Measure}=+0.0809$). Unfortunately, the false positive rate increased (incorrectly identifying 71/108 good outcome cases) suggesting that while the model succeeds at predicting poor outcome (correctly identifying 356/366 poor outcome cases), it still needs to be further trained on good outcome cases to better identify them.

This model utilises 4036 features which may yield a high amount of redundancy, therefore lowering the ability of the model to accurately predict in a random forest environment. Decreasing the number of features implemented allows for faster running of the model. Thus, by ordering probes used in the RF model by mean decreased gini (the ability of a feature to correctly split data, otherwise known as feature purity) we identify features that are most essential to the model and remove less informative probes. Features that best distinguish patient outcome are clearly defined by the density of diagnostic samples in areas that are not overlapping with distributions of methylation in remission samples, an example of which can be found in Figure 6.16. Table 6.9 demonstrated how the removal of redundant probes from the available pool of features improves the model's performance. The optimal model, comprising the top 100 ranked probes, achieves the highest precision with negligible affects upon recall while greatly improving the false positive rate. This culminates in the highest F-measure of 0.921 of all models and one of the highest AUC (0.933) predictions. Compared to the Weka designed random forest model this model may have a similar AUC, however the false positive rate and F-measure is markedly improved. While the model continues to improve the fewer probes that are added, once we restrict the model to the top 50 probes, it begins to regress, suggesting insufficient information is available to distinguish good from poor outcome cases at that level.

Table 6.9 Final model evaluation of random forest MBL algorithm using 5-fold, 3-times repeated cross validation of varying differentially methylated probes (n = 158)

Machine-Based Learning Algorithm	Probes	Precision	Recall	False Positive Rate	F-Measure	AUC*
Random Forest	4036	0.833	0.973	0.657	0.898	0.914
	1000	0.857	0.973	0.546	0.912	0.930
	500	0.866	0.973	0.509	0.916	0.934
	250	0.871	0.962	0.482	0.914	0.937
	100	0.886	0.959	0.417	0.921	0.933
	50	0.886	0.933	0.407	0.911	0.929

Abbreviations: AUC - Area under the curve; *AUC is calculated by ROC analysis

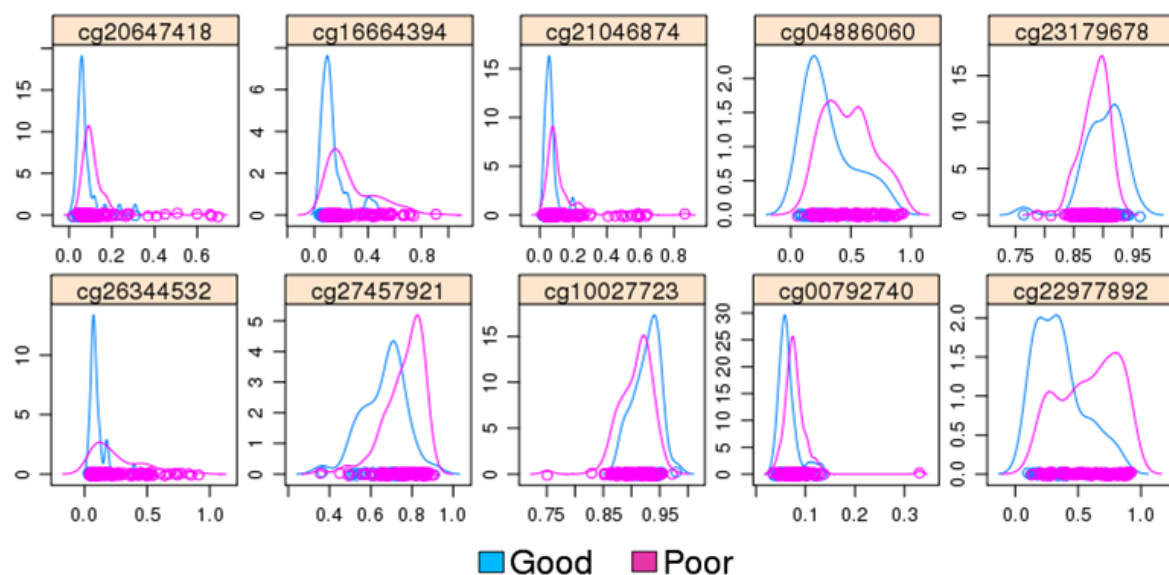


Figure 6.16 Density plots of top ten highest ranked probes by mean decreased Gini in the random forest model

All probes used in the random forest model were ranked by mean decreased Gini (feature purity). The top 10 probes were extracted and beta density was compared for each. X-axis representing the beta value and y-axis represents the relative density.

6.6 Discussion

While prognostic measures, currently based on genetic and chromosomal abnormality, show some utility in predicting severity of disease, there remains an urgent need more precise predictors of relapse at diagnosis, preferably independent of patient subtype. This will allow for patients to obtain a more personalised prognosis that has potential to impact treatment regimens and post-remission care.

The primary aim was to differentiate patient outcome in a retrospective AML cohort using DNAm. We hypothesised that there was differential DNAm at diagnosis between patients that went on to achieve an initial 5-year EFS compared to patients that did not. Furthermore, we hypothesised that these differentially methylated probes would be associated with genes and pathways that drive disease progression, and that in combination, could also be used to distinguish and predict future AML patient outcomes through a machine-based learning approach.

These hypotheses were explored by applying linear regression analysis to determine differential methylation in combination with multiple machine learning algorithms to test their prognostic utility. This was applied to genome-wide DNAm array data of 158 diagnostic AML samples across two cohorts (Australian and TARGET AML cohorts).

6.6.1 Utility of DNA Methylation Profiling to Predict Patient Outcome in Paediatric AML

Several studies have previously attempted to predict 5-year EFS in adult and paediatric AML, however these are largely subtype-specific and rely on various methods such as putative gene expression, mutation or minimal residual disease monitoring (Yagi, Morimoto et al. 2003; Lange, Hubmann et al. 2011; Yin, O'Brien et al. 2012). Unfortunately, many of these have shown to be unreliable in their replication and also suffer from a temporal problem where the assay must be run directly after a sample is taken or require continuous monitoring before coming to an appropriate response. More recent genome-wide DNAm studies have shown prognostic promise such as research conducted by Stieglitz *et al.* This research shows that a

methylation panel of 1527 CpG sites accurately predicted 4-year EFS in juvenile myelomonocytic leukaemia (JMML). This study has shown that not only does this signature exist in bone marrow samples, but is also found in spleen tissue or peripheral blood (Stieglitz, Mazor et al. 2017). While useful, this still retains the problem of being dependent upon specific cytogenetics or mutations.

In the current study we have determined that poor patient outcome is slightly more hypermethylated overall than their good outcome counterparts and there are over 2000 sites of differential DNAm associated with paediatric AML outcome independent of subtype. Upon incorporation of the TARGET cohort, consisting primarily of severely poor outcome patients, we discovered another 3419 probes that were also associated with patient outcome. The strength of this analysis is rooted in multi-cohort approach, which allows for further support of findings through pooled-analysis. Of the total 4036 probes identified, 617 probes were found in both analyses giving further credence to their use as a prognostic biomarker in paediatric AML recurrence. We have demonstrated that, while the full array of significant probes are unable to distinguish patient outcome fully, subsets of probes, namely the top 500 of the pooled-analysis, were able to largely differentiate poor from good patient outcome at diagnosis. This highlights the necessity of having large sample sizes when attempting to determine the more subtle differences in DNAm between cancer samples.

While hierarchical clustering allows for grouping of patients, a more direct approach to prognosis would be preferable to achieve a more personalised outcome. Our study showed that sex and cytogenetics played no role predicting 5-year EFS, however there was a slight association with patient age suggesting that younger patients are more prone to relapse. This is in contrast with a study conducted by Rubnitz *et al.* which suggested that sex was associated with patient outcome (Rubnitz, Razzouk et al. 2007). Pooled-analysis supported the association of 14 probes with patient outcome that were previously identified in the discovery cohort, however individual probes showed little ability to distinguish 5-year EFS. In our attempt to reduce input, the smallest possible combination of CpG sites that could reliably predict patient outcome was reduced to 7 probes, none of which were previously identified in the discovery analysis. This suggests that there is a large heterogeneity of DNAm amongst AML samples in

association with outcome. Future studies would do well to validate this signature to determine its true use in a population setting.

To the best of our knowledge, we have demonstrated the first use of DNAm as a biomarker to predict paediatric AML patient outcome independent of patient cytogenetics. However, considering the full signature found in the pooled-analysis is unable to distinguish patient outcome, this suggests that there are several false discoveries in the data. Furthermore, since no probes reached significance after correction for multiple testing, this also suggests that the differences in DNAm between good and poor outcome groups are likely few in number and subtle in effect size. Future studies into this field would do well to use a greater sample size to truly elucidate probes that can be used to determine patient outcome. Considering we have demonstrated that the best predictive models can be easily influenced by sample size, stronger predictive modelling such as machine-based learning approaches would likely do a better job at not only determining patient outcome, but generating replicable models that can be applied in larger population-based studies.

6.6.2 Machine-Based Learning Approaches to Define Paediatric AML Outcome

Machine learning has been utilised in a myriad of cancers, largely to classify and predict tumour pathogenesis (Furey, Cristianini et al. 2000; Candia, Cherukuri et al. 2015; Bolouri, Farrar et al. 2017). Currently, the largest DNAm-based study to utilise machine learning was conducted by Capper *et al.* where HM450K data were used to classify central nervous system tumours. According to this approach, up to 12% of cases were identified incorrectly by standard histopathological testing, potentially affecting patient treatment regimens (Capper, Jones et al. 2018). In similar fashion, this study was designed to use machine learning to predict patient outcome for early detection of recurrence and potentially personalised intervention.

Our study has applied machine learning in the field of paediatric AML prognosis, and currently stands as the only DNAm-derived approach yet attempted. We discovered the most appropriate model for patient prognosis was the RF algorithm, performing better than four other machine learning methods. This algorithm is commonly used in other cancer-based

classification tools designed by Bolouri *et al.* and Capper *et al.* (Bolouri, Farrar et al. 2017; Capper, Jones et al. 2018). Hyperparameter tuning of the RF model played little role in optimisation of prediction with a $\pm 2\%$ change in accuracy dependant on number of trees and probes available to be randomly selected at any given node. While initial model creation showed similar results to the original Weka model, removal of probes based on mean decreased gini (down to the top 100 probes) thoroughly improved the model's accuracy. This suggests that there is either a large amount of redundancy in the probes that are being selected, or several probes that were identified and used in this analysis are not predictive of patient outcome. Furthermore, while this model has some ability to predict poor patient outcome, the false-positive rate remains too high for realistic clinical translation. To overcome this, the model will need to be trained with larger amounts of patients that achieved a 5-year EFS. This should not only improve the model's inability to predict good outcome patients, but will also better identify pertinent probes that can distinguish patient outcome.

Interestingly, should this approach to AML prognosis be clinically translated, it creates a problem in that improvements in prediction will limit the number of poor outcome cases available for ongoing improvements as the model 'learns' from new data. Therefore, continuation and translation of this work will require increased sample numbers to accurately capture the aforementioned subtle changes in DNAm that is contributing to patient outcome variation and define an end-point to model learning prior to translation.

6.6.3 Epigenetic Disruption in Pathways Associated with Paediatric AML Patient Outcome

Knowing patient outcome *a priori* serves as a strength of this study, allowing for DNAm to serve as a dynamic biomarker and simultaneously discover probes of functional consequence. In this study we discovered several genes, both within the discovery and pooled cohorts, with differential DNAm that may play a role in the recurrence or progression of paediatric AML. Those focussed on include *MIRLET7A1*, *PTPRN2* and *PAX6*.

MIRLET7A3

MicroRNA Let-7a-3 (*MIRLET7A3*) is a member of a microRNA family known as *lethal-7*. Mediated largely by expression of the RNA binding protein LIN28, this coupling of genes have functions involved in timing of development (Ong, Elks et al. 2009) and stem cell pluripotency (Yu, Vodyanik et al. 2007). Furthermore, there has been reports that DNA hypomethylation and overexpression in adult AML and myelodysplastic syndrome to be associated with poor patient outcome (Li, Lin et al. 2013; Wu, Yao et al. 2017). This link of DNA hypomethylation with increased expression has also been identified in lung adenocarcinoma (Brueckner, Stresemann et al. 2007). Conversely, our data challenges these findings, showing that DNA hypermethylation is a biomarker of poor patient outcome. In agreement with Zhu *et al.* we find that *MIRLET7A3* hypomethylation correlates with samples of favourable/intermediate karyotype as defined by the ELN classification (Zhu, Yao et al. 2016), however this does not correlate with patient outcome in our cohort. We hypothesise that this hypermethylation is having a repressive effect upon gene expression, but this requires further study. Ko *et al.* have shown previously that *MIRLET7A3* hypermethylation is negatively correlated with *CEBPA* methylation and may produce a more accurate definition of prognosis when measured in combination (Ko, Fang et al. 2014). From a functional perspective, future studies should elucidate if methylation influences *MIRLET7A3* expression in paediatric AML and if this level of methylation or expression can be correlated with patient outcome. In combination with *LIN28*, it is plausible that overexpression of these genes could contribute to the reprogramming of somatic cell to pluripotent stem cells which are major influences in clonal evolution and AML relapse (Hirsch, Zhang et al. 2016).

PTPRN2

The Protein Tyrosine Phosphate Receptor Type N2 (*PTPRN2*) gene is a protein coding gene largely involved in the regulation of insulin secretion, however, studies have determined that it may play a role in the cellular localisation of cofilin, thus indirectly being involved in cell migration and metastatic relapse (Sengelaub, Navrazhina et al. 2016). This is the first study to implicate *PTPRN2* in AML prognosis and pathogenesis, however, our results show a strange patterning of DNAm with all significant CpG sites identified to not be correlated with one

another and having almost alternating hypo/hypermethylated differences between patient outcome groups. Small regions of DNAm show consistent trends in methylation differences between outcome groups, however all CpGs identified in this analysis are found within introns and are therefore unlikely to influence gene expression. The exception to this is cg23841288, which has been identified to be in an enhancer and an area of DNase hypersensitivity. With reduced DNAm in this area in poor outcome patients, it is plausible that transcription of this gene may be altered through recruitment of other proteins. Future studies would do well to investigate the potential transcription factors associated with *PTPRN2* and if methylation of this site can alter expression in poor outcome paediatric AML patients.

PAX6

Paired Box 6 (*PAX6*) controls gene transcription through the highly conserved paired box and homeobox domains present within the gene. While primarily known for its role in neural tissue development, *PAX6* has been shown to have tumour suppressor and oncogenic effects in lung, prostate, breast and brain cancers (Wu, Rauch et al. 2010; Moelans, Verschuur-Maes et al. 2011; Meng, Zou et al. 2015; Zhang, Yang et al. 2015). Aberrant *PAX6* DNAm has been observed in several cancers including melanoma and hepatocellular carcinoma (Furuta, Nobeyama et al. 2006; Shih, Kuo et al. 2016). Finally, in lung adenocarcinoma, studies have shown that promoter methylation of *PAX6* regulates and drives cancer cells towards a stem-like state (Ooki, Dinalankara et al. 2018). This methylation profile is also found in small cell lung cancer where this is associated with poor patient prognosis (Zhang, Yang et al. 2015). While two probes discovered in our analysis are within the gene body, one CpG site, cg22272457, is found within 200bp of the TSS for the gene. This is the first report of promoter region DNA hypermethylation associated with poor outcome in paediatric AML and warrants further study into the potential pathogenic effects that this aberrant methylation may cause.

6.6.4 Future Directions and Limitations

The approach adopted here assumed that there are universal DNAm differences that occur in association with outcome. Overall this study has identified some DNAm sites as potentially useful biomarkers for paediatric AML outcome. Furthermore, we have shown, in combination

with machine-based learning algorithms, the utility of these biomarkers for prognostic purposes at the time of diagnosis. While these results show potential, there are still many improvements that need to be made before such findings could ever be truly translated. The first major barrier is that only subtle changes in mean DNAm were identified between diagnostic AML samples, independent of subtype. Furthermore, none of these findings survived multiple testing in either the discovery or combined cohorts. This is potentially due to heterogeneity within groups in terms of either methylation level or cell composition of the starting bone marrow aspirate samples. Future studies should investigate sorting for haematopoietic stem cell populations of diagnostic samples and assaying their DNAm in comparison to the remaining cells. This may not only allow for the discovery of DNAm events that drive initiation of paediatric AML, but when used retroactively in combination with outcome data, may also be more telling of a universal DNAm signature associated with patient outcome. Another interesting study would be to investigate the DNAm of HSPCs of patient's remission samples to determine if an 'epigenetic memory' remains that can be associated with those that go on to relapse compared to those that achieve a 5-year EFS.

The limitation of significant probes not surviving multiple testing may have also affected downstream analyses by virtue of false positive inclusion. This is supported by the fact that the model generated from the discovery cohort data could not accurately predict outcome in an independent dataset. The obvious solution to this limitation was to combine the cohorts to create a model based on a larger dataset, however this removes the ability to validate the resulting model beyond internal 5-fold cross validation. By using a larger sample size, something that is a necessity for machine learning algorithms to work effectively (Byrd, Chin et al. 2012), it should be possible to not only refine the model to better reflect good patient outcome prognosis, but also have a separate cohort for validation purposes.

Future improvements to this study can be made regarding feature implementation in the final machine learning model. While our study utilised the most significant probes derived from regression analysis, many other studies have utilised more comprehensive approaches to feature selection. Capper *et al.* and Bolouri *et al.* used an ensemble method which creates multiple models which encapsulates all data, culminating the best performing probes in an averaged single model to determine patient subtype (Bolouri, Farrar et al. 2017; Capper, Jones

et al. 2018). This requires much more computational power than available for this current study but would prove valuable for increasing model accuracy in future studies. Alternatively, we could have approached the study similar to Nordlund *et al.* and Martino *et al.* who used the feature selection function of the nearest shrunken centroid machine learning algorithm to select appropriate features in an unsupervised manner (Martino, Thanh et al. 2015; Nordlund, Bäcklin et al. 2015). While both regression and feature selection are acceptable ways to select for model input, it may be worth exploring alternative approaches in future model generations.

Should these limitations be addressed, it is plausible that a functional prognostic model of patient outcome for paediatric AML based on DNAm can be achieved. This would potentially allow for early detection of recurrence and subsequent personalised treatment. Furthermore, this would create the first prognostic approach to 5-year EFS that is independent of patient subtype, treating all cases equally despite historical-based classification determined by the ELN system (Dohner, Estey et al. 2016).

This page is intentionally left blank.

Chapter 7 General Discussion

In recent decades, there has been tremendous improvement in the rates of complete remission (CR) in those diagnosed with acute myeloid leukaemia, with reports of up to 80% of all patients under 60 years of age entering complete remission (Burnett, Wetzler et al. 2011). This has largely been driven by research into subtype-specific mechanisms and classification, which has subsequently been adapted into the constantly evolving World Health Organisation (WHO) guidelines for patient prognosis (Arber, Orazi et al. 2016). This classification has allowed for the targeted clinical trials within subtypes of AML, but despite this, standard care has remained largely unchanged. While CR rates have improved over time, in recent years this trend has stagnated. Further, it does not consider the poor rates of CR for patients that have relapsed. Studies in adults have reported 5-year OS of patients suffering from AML who relapsed to be anywhere between 10% and 18% (Breems, Van Putten et al. 2005; Rowe, Li et al. 2005). Few studies have looked at the rates of OS of paediatric AML post-relapse, but it has been identified as the major cause of disease related mortality (Kaspers 2012). Rate of relapse has only improved for low-risk patients but remains unchanged for patients at higher risk (Alexander, Wang et al. 2017). More recent studies have identified that the length of time to relapse after bone marrow transplantation (BMT) is strongly associated with OS, yet mortality rates still remain high (Bejanyan, Weisdorf et al. 2015). This suggests that there are still many aspects of AML pathogenesis yet to be discovered, particularly in reference to recurrence of disease. Considering genetic factors in isolation do not appear to fully explain disease aetiology, it is likely that alternative mechanisms, such as epigenetic variation, are also involved.

This study was undertaken to explore the pathogenesis, classification and current prognostic approaches in the understudied field of paediatric AML. In an attempt to discover novel pathogenic pathways, epigenetic analyses (specifically DNA methylation profiling) were applied to a unique collection of bone marrow aspirate samples collected over many decades (Morenos, Saffery et al. 2012; Farrar, Schuback et al. 2016).

There are several reasons to focus on DNA methylation (DNAm) in paediatric AML. Firstly, mutations in several epigenetic modifiers have been shown to contribute to the pathogenesis of adult AML (Shih, Abdel-Wahab et al. 2012). Despite the fact that similar mutations are largely absent in paediatric AML, (Liang, Liu et al. 2013) several studies have nevertheless reported aberrant DNAm as a feature of childhood disease (Bolouri, Farrar et al. 2017).

The overall aim was to characterise DNAm variation associated with active AML, specific subtypes of AML, relapse in AML and OS prognostication. As part of this, differences between adult and paediatric forms of the disease would also be revealed, with potential to improve patient diagnosis, prognosis, treatment and long-term outcomes.

Initial analyses aimed to understand the global disruptions of DNAm in paediatric AML using proxy measures of total 5mC - Genome Wide Average Methylation (GWAM) and repeat element-based (*Alu* and *LINE1*) methylation assessment. Global DNA hypomethylation is a common feature of almost all cancers (Estecio, Gharibyan et al. 2007). In contrast, our results showed that paediatric AML samples are generally hypermethylated when compared to non-diseased samples. The changes in effect size eludes to several genes that were once unmethylated becoming methylated, and ultimately repressed, in paediatric AML. This is in direct contrast with results identified by Deneberg *et al.* and Melki *et al.*; however their studies compared diagnostic AML to healthy donor marrow (Melki and Clark 2002; Deneberg, Grövdal et al. 2010), suggesting that the actual change observed in our study is largely driven by a mass hypomethylation event in remission samples. Should the observed changes be due to blanket hypomethylation, this further evidences that genes that were once repressed need to be hypomethylated in order to appropriately function, and thus, rescue the original phenotype. This may be a consequence of induction therapy and warrants further investigation, either longitudinally or through comparison of donor to remission marrow. When comparing donor marrow to our diagnostic samples, while significantly hypermethylated in diagnostic samples, the difference is fairly negligible and aligns more closely with studies conducted by Pfeifer *et al.* (Pfeifer, Steigerwald et al. 1988) where no differences were observed.

The second focus of this study was aimed at identifying sites and genes of differential methylation that may be playing a role in paediatric AML pathogenesis. This was achieved using array-based DNAm profiling. While aberrant DNAm has been widely studied in adult AML, only limited childhood studies have been undertaken. By combining data from two cohorts independent of AML subtype, it was possible to define a DNAm signature that has capacity to accurately separate diagnostic, remission and healthy donor samples. Findings showed that remission and healthy donor samples were more closely related to one another compared to diagnostic samples.

Individual differentially methylated probes (DMPs) and regions (DMRs) linked to specific genes were also identified. Those showing the most consistent variation in association with AML included *WT1*, *DGKG*, and *ZSCAN1*. Upon pooled-analysis, we also discovered two more genes in *REC8* and *IRX1* that may be involved in leukaemic pathogenesis. *WT1* has been widely studied in AML with mutations in the gene being associated with disease progression yet is not known to be a driver mutation (Metzeler, Herold et al. 2016). It is commonly mutated in both adult (Hou, Huang et al. 2010) and paediatric AML (Hollink, van den Heuvel-Eibrink et al. 2009). Aberrant DNAm of the *WT1* promoter region implicates aberrant *WT1* expression in paediatric AML regardless of mutational status. While previous studies have shown this promoter methylation does not affect *WT1* directly, it does correlate with the expression of alternative transcripts of *WT1* (El Bordiny, Al-Ghandour et al. 2018). This suggests that while *WT1* may not be a driver of disease, it may play an integral role in disease progression through a complex interplay of mutation, aberrant DNAm and alternative transcription, culminating in widespread epigenetic dysregulation in combination with TET proteins (Rampal and Figueroa 2016). This would also help explain the dysregulation of DNAm found in the paediatric setting despite a lack of epigenetic modifier gene mutations. The remaining four genes identified are all plausible candidates to affect AML pathogenesis, having either been previously identified for their role in other forms of leukaemia/cancer (Guo, Liu et al. 2010; Kai, Yamamoto et al. 2017) or function within relevant pathways of the ‘hallmarks of cancer’ (Xu, Beasley et al. 2005; Hanahan and Weinberg 2011). All these genes require additional experiments pertaining to their transcriptional regulation as well as functional studies to define their role in AML initiation and progression.

Although we have shown that universal aberrant DNAm exists in paediatric AML, the various subtypes of AML have unique disease aetiology, suggesting a role for subtype-specific aberrant DNAm as a mechanism of disease. Studies have previously implicated aberrant DNAm in AML as both a mechanism and biomarker of subtype-specific pathogenesis, including our own global methylation study that showed various subtypes have differing global methylation signatures. Through the combination of DNAm and RNA-seq analysis, Bolouri *et al.* was able to distinguish several subtypes including *CBF*-rearranged *KMT2A*-rearranged and *IDH1*-mutated, independent of age, suggesting common pathways are found in both children and adults (Bolouri, Farrar et al. 2017). Similarly, Figueroa *et al.* also showed the potential of

DNAm as a biomarker in adult AML, dividing over 300 samples into 15 subgroups, largely defined by chromosomal rearrangement or mutational status (Figueroa, Lugthart et al. 2010). Akalin *et al.* used reduced representation bisulphite sequencing to determine that *IDH*-mutated AML patients were severely hypermethylated, particularly around CpG islands and gene promoter regions, while *KMT2A*-rearranged AMLs showed extensive loss of methylation in non-coding areas of the genome (Akalin, Garrett-Bakelman et al. 2012). This shows that the dynamic nature of DNAm responds differently in multiple forms of AML, however few studies have looked specifically at paediatric AML subtype-specific DNAm.

Our cohort was designed to encapsulate the large heterogeneity of paediatric AML through inclusion of all subtypes of AML. Therefore, our third study was aimed at defining a signature of DNAm that may be able to distinguish patient subtype while also looking into potential contributors to disease pathogenesis in paediatric AML patients. We were able to identify a signature specific to the *KMT2A* subtype, largely driven by widespread hypomethylation. This is in line with findings by Akalin *et al.* (Akalin, Garrett-Bakelman et al. 2012). We also found clustering of the inv(16) subtype with a mix of hypo/hypermethylation suggesting a more complex mechanism at work than blanket demethylation. Further insight into differential DNAm in inv(16) patients identified *MNI*, *MEIS1* and *RUNX3* as consistently aberrantly methylated compared to other forms of AML, serving as a potent subtype biomarker. Studies have previously shown the link between these genes, cancer and epigenetic regulation, however the mechanism behind this is still being identified. DNA methyltransferases have been implicated as one mechanistic option, but the targeted nature of their function is again, yet to be understood (Ferreira, Heyn et al. 2015; Larmonie, Arentsen-Peters et al. 2017). This suggests that while these genes are being epigenetically regulated, other proteins are playing a role in the targeted aberrant methylation identified. Future studies would do well to identify all *DNMT* functionality in these samples as well as upstream *DNMT*-interacting-proteins such as *UHRF1* (Sharif, Muto et al. 2007) and *LSH* (Dennis, Fan et al. 2001), which may be contributing to *DNMT* regulation and subsequent subtype-specific DNAm signatures.

The final study was designed to tackle the pressing issue of predicting 5-year event-free survival in paediatric AML. As it currently stands, there is no early detection test for AML and attempts to predict EFS are hampered by requiring complete remission to have been achieved thus

diminishing the time available to act, or by complex methodology that is unlikely to translate (Ravandi, Jorgensen et al. 2016). Furthermore, most minimal residual disease monitoring varies between subtypes of AML (Gaidzik, Weber et al. 2017; Hollein, Jeromin et al. 2018), therefore a universal predictor of patient outcome, independent of subtype and can be undertaken upon diagnosis, is desirable. Using our combined discovery and validation cohorts we identified a signature of DNAm that was able to predict patient 5-year EFS with 84% efficiency by regression modelling. Further utilising this set of probes in a machine-based learning environment allowed us to identify random forest as the optimal machine learning algorithm while also improving model accuracy to upwards of 93%. The utility of this method allows for the continued use and improvement of the model as more samples become available. Previous studies conducted by Capper *et al.* and Nordlund *et al.* have shown the application of DNAm as a biomarker of subtype in different cancers (Nordlund, Bäcklin et al. 2015; Capper, Jones et al. 2018) while Bolouri *et al.* has shown this within AML too (Bolouri, Farrar et al. 2017). However, this is the first instance of a machine learning and DNAm combination being used to predict patient outcome. Further studies will be required to further test (stress) the model at a population scale, validate findings and optimise feature selection. Implications of these findings are tailored towards patient monitoring. Considering most children undergo bone marrow transplants (BMT), studies have shown that the reduced intensity conditioning (RIC) is viable post-BMT and in combination with 5-year EFS prognosis, has potential to reduce chemotherapy induced side effects and improve donor engraftment (Niederwieser, Lange et al. 2006; Chiesa and Veys 2014). Alternatively, a poor outcome prediction could lead to alternative or intensified therapy (Sander, Zimmermann et al. 2010) in an attempt to eradicate all disease causing cancer cells likely to cause patient relapse.

Despite the large amount of novel information generated in these studies, each had their own challenges. One of the major limitations of this and many other DNAm-based studies is the inability to examine the direct functional effect that the observed differences are having upon expression of the gene. The majority of patient samples were derived from archival bone marrow, therefore we were unable to examine gene expression as degradation of RNA would have instantly occurred. While one can infer direction of effect from our findings, without expression data the findings of these studies remain associative at best. This highlights the need for additional studies using freshly obtained material as has been used in The Cancer Genome

Atlas or TARGET studies (Weinstein, Collisson et al. 2013; Bolouri, Farrar et al. 2017) to truly elucidate the functional effect that aberrant DNAm is having upon the patient. Furthermore, while we have shown that DNAm plays a pivotal role in paediatric AML pathogenesis, heavily understudied epigenetic mechanisms such as histone modification likely plays a collaborative role in disease pathogenesis too (Krivtsov and Armstrong 2007).

Other limitations were largely cohort related. The Australian discovery cohort, while extensive, did lack in number and fundamental information pertaining to patient mutations and disease burden (blast count). The lack of patient information inhibits the ability to accurately determine associations found between groups, where specific mutations in AML are known to modulate DNAm patterns (Figueroa, Lugthart et al. 2010) while others are known to determine patient subtype and outcome (Dohner, Estey et al. 2016). The low number of patient samples naturally reflects in the ability of findings to be validated. This was evident upon pooled-analysis with the TARGET cohort where several findings were not only diminished, but entirely invalidated upon addition. The problems associated with the lack of a true validation cohort were evident when creating a model to predict paediatric AML 5-year EFS, which has yet to be tested in the broader AML population. This highlights the necessity of larger validation cohorts for future studies.

Despite these limitations, this study is one of the most comprehensive insights into aberrant DNAm in the paediatric AML setting yet undertaken. We have shown the utility of DNAm arrays in particular regards to biomarking disease, copy number variation, predicting 5-year EFS and patient subtyping while also identifying/validating aberrant methylation believed to be involved in pathogenesis of disease, including genes such as *WT1*, *MEIS1* and *MNI*. We believe that this provides a foundation for future paediatric AML studies to further elucidate functionally relevant pathways caused by discordant DNAm. Ultimately, we have demonstrated that DNAm contributes greatly to AML pathogenesis and that these findings may begin a journey towards a more personalised experience for all involved in the treatment of paediatric AML, from patient subtyping and prognosis to therapeutic intervention.

This page is intentionally left blank.

References

- Abdel-Wahab, O. and Levine, R. L. (2013). Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. *Blood* 121(18): 3563.
- Abrahamsson, J., Clausen, N., Gustafsson, G., Hovi, L., Jonmundsson, G., Zeller, B., Forestier, E., Heldrup, J. and Hasle, H. (2007). Improved outcome after relapse in children with acute myeloid leukaemia. *Br J Haematol* 136(2): 229-236.
- Abrahamsson, J., Forestier, E., Heldrup, J., Jahnukainen, K., Jónsson, O. G., Lausen, B., Palle, J., Zeller, B. and Hasle, H. (2010). Response-guided induction therapy in pediatric acute myeloid leukemia with excellent remission rate. *Journal of Clinical Oncology* 29(3): 310-315.
- Agrawal, S., Unterberg, M., Koschmieder, S., zur Stadt, U., Brunnberg, U., Verbeek, W., Buchner, T., Berdel, W. E., Serve, H. and Muller-Tidow, C. (2007). DNA methylation of tumor suppressor genes in clinical remission predicts the relapse risk in acute myeloid leukemia. *Cancer Res* 67(3): 1370-1377.
- Akalin, A., Garrett-Bakelman, F. E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T. A., Huang, Y., Biswas, D. and Hess, J. L. (2012). Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS genetics* 8(6): e1002781.
- Alcalay, M., Orleth, A., Sebastiani, C., Meani, N., Chiaradonna, F., Casciari, C., Sciarpi, M. T., Gelmetti, V., Riganelli, D., Minucci, S., Fagioli, M. and Pelicci, P. G. (2001). Common themes in the pathogenesis of acute myeloid leukemia. *Oncogene* 20(40): 5680.
- Alessandra, I., Angelo, C., Mario, F., Barbara, G., Luisa, A., Simona, R., Laura, F., Andrea, V. and Pier Giuseppe, P. (2013). DNA damage in stem cells activates p21, inhibits p53, and induces symmetric self-renewing divisions. *Proceedings of the National Academy of Sciences of the United States of America* 110(10): 3931.
- Alexander, T. B., Wang, L., Inaba, H., Triplett, B. M., Pounds, S., Ribeiro, R. C., Pui, C. H. and Rubnitz, J. E. (2017). Decreased relapsed rate and treatment-related mortality contribute to improved outcomes for pediatric acute myeloid leukemia in successive clinical trials. *Cancer*.
- Amin, H. M., Yang, Y., Shen, Y., Estey, E. H., Giles, F. J., Pierce, S. A., Kantarjian, H. M., O'Brien, S. M., Jilani, I. and Albitar, M. (2005). Having a higher blast percentage in circulation than bone marrow: clinical implications in myelodysplastic syndrome and acute lymphoid and myeloid leukemias. *Leukemia* 19(9): 1567-1572.

Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., Bloomfield, C. D., Cazzola, M. and Vardiman, J. W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. United States, AMERICAN SOCIETY OF HEMATOLOGY: 2391.

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D. and Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30(10): 1363-1369.

Bachas, C., Schuurhuis, G. J., Zwaan, C. M., van den Heuvel-Eibrink, M. M., den Boer, M. L., de Bont, E. S. J. M., Kwidama, Z. J., Reinhardt, D., Creutzig, U., de Haas, V., Kaspers, G. J. L. and Cloos, J. (2015). Gene Expression Profiles Associated with Pediatric Relapsed AML. *PLoS One* 10(4): e0121730.

Balgobind, B. V., Raimondi, S. C., Harbott, J., Zimmermann, M., Alonzo, T. A., Auvrignon, A., Beverloo, H. B., Chang, M., Creutzig, U., Dworzak, M. N., Forestier, E., Gibson, B., Hasle, H., Harrison, C. J., Heerema, N. A., Kaspers, G. J. L., Leszl, A., Litvinko, N., Nigro, L. L., Morimoto, A., Perot, C., Pieters, R., Reinhardt, D., Rubnitz, J. E., Smith, F. O., Stary, J., Stasevich, I., Strehl, S., Taga, T., Tomizawa, D., Webb, D., Zemanova, Z., Zwaan, C. M. and van den Heuvel-Eibrink, M. M. (2009). Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: results of an international retrospective study. *Blood* 114(12): 2489-2496.

Bannister, A. J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research* 21(3): 381-395.

Barchitta, M., Quattrocchi, A., Maugeri, A., Canto, C., La Rosa, N., Cantarella, M. A., Spampinato, G., Scalisi, A. and Agodi, A. (2017). LINE-1 hypermethylation in white blood cell DNA is associated with high-grade cervical intraepithelial neoplasia. *BMC Cancer* 17: 601.

Barth, T. K. and Imhof, A. (2010). Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci* 35(11): 618-626.

Bejanyan, N., Weisdorf, D. J., Logan, B. R., Wang, H.-L., Devine, S. M., de Lima, M., Bunjes, D. W. and Zhang, M.-J. (2015). Survival of patients with acute myeloid leukemia relapsing after allogeneic hematopoietic cell transplantation: a center for international blood and marrow transplant research study. *Biol Blood Marrow Transplant* 21(3): 454-459.

Belinsky, S. A., Nikula, K. J., Palmisano, W. A., Michels, R., Saccomanno, G., Gabrielson, E., Baylin, S. B. and Herman, J. G. (1998). Aberrant Methylation of p16 INK4a Is an Early Event in Lung Cancer and a Potential Biomarker for Early Diagnosis, National Academy of Sciences of the United States of America: 11891.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R. and Sultan, C. (1976). Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *British Journal Of Haematology* 33(4): 451-458.

Bennett, K. L., Karpenko, M., Lin, M. T., Claus, R., Arab, K., Dyckhoff, G., Plinkert, P., Herpel, E., Smiraglia, D. and Plass, C. (2008). Frequently methylated tumor suppressor genes in head and neck squamous cell carcinoma. *Cancer Res* 68(12): 4494-4499.

Bernt, K. M., Zhu, N., Sinha, A. U., Vempati, S., Faber, J., Krivtsov, A. V., Feng, Z., Punt, N., Daigle, A. and Bullinger, L. (2011). MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell* 20(1): 66-78.

Bird, A. (2007). Perceptions of epigenetics. *Nature* 447(7143): 396-398.

Bjornsson, H. T., Fallin, M. D. and Feinberg, A. P. (2004). An integrated epigenetic and genetic approach to common human disease. *Trends Genet* 20(8): 350-358.

Boers, A., Wang, R., van Leeuwen, R. W., Klip, H. G., de Bock, G. H., Hollema, H., van Criekinge, W., de Meyer, T., Denil, S., van der Zee, A. G. J., Schuurin, E. and Wisman, G. B. A. (2016). Discovery of new methylation markers to improve screening for cervical intraepithelial neoplasia grade 2/3. *Clinical Epigenetics* 8: 29-29.

Bolouri, H., Farrar, J. E., Triche, T., Jr., Ries, R. E., Lim, E. L., Alonzo, T. A., Ma, Y., Moore, R., Mungall, A. J., Marra, M. A., Zhang, J., Ma, X., Liu, Y., Liu, Y., Auvil, J. M. G., Davidsen, T. M., Gesuwan, P., Hermida, L. C., Salhia, B., Capone, S., Ramsingh, G., Zwaan, C. M., Noort, S., Piccolo, S. R., Kolb, E. A., Gamis, A. S., Smith, M. A., Gerhard, D. S. and Meshinchi, S. (2017). The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med*.

Bolouri, H., Farrar, J. E., Triche, T., Jr., Ries, R. E., Lim, E. L., Alonzo, T. A., Ma, Y., Moore, R., Mungall, A. J., Marra, M. A., Zhang, J., Ma, X., Liu, Y., Liu, Y., Auvil, J. M. G., Davidsen, T. M., Gesuwan, P., Hermida, L. C., Salhia, B., Capone, S., Ramsingh, G., Zwaan, C. M., Noort, S., Piccolo, S. R., Kolb, E. A., Gamis, A. S., Smith, M. A., Gerhard, D. S. and Meshinchi, S. (2018). Erratum: The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* 24(4): 526.

Boulton, J. and Wainscoat, J. S. (2007). Gene silencing by DNA methylation in haematological malignancies. *Br J Haematol* 138(1): 3-11.

Bozic, T., Lin, Q., Frobel, J., Wilop, S., Hoffmann, M., Muller-Tidow, C., Brummendorf, T. H., Jost, E. and Wagner, W. (2015). DNA-methylation in C1R is a prognostic biomarker for acute myeloid leukemia. *Clin Epigenetics* 7: 116.

Breems, D. A., Van Putten, W. L., Huijgens, P. C., Ossenkoppele, G. J., Verhoef, G. E., Verdonck, L. F., Vellenga, E., De Greef, G. E., Jacky, E. and Van der Lelie, J. (2005). Prognostic index for adult patients with acute myeloid leukemia in first relapse. *Journal of Clinical Oncology* 23(9): 1969-1978.

Brown, P., McIntyre, E., Rau, R., Meshinchi, S., Lacayo, N., Dahl, G., Alonzo, T. A., Chang, M., Arceci, R. J. and Small, D. (2007). The incidence and clinical significance of nucleophosmin mutations in childhood AML. *Blood* 110(3): 979-985.

Brueckner, B., Stresemann, C., Kuner, R., Mund, C., Musch, T., Meister, M., Sultmann, H. and Lyko, F. (2007). The Human *let-7a-3* Locus Contains an Epigenetically Regulated MicroRNA Gene with Oncogenic Function. *Cancer research* 67(4): 1419.

Burnett, A., Wetzler, M. and Lowenberg, B. (2011). Therapeutic advances in acute myeloid leukemia. *J Clin Oncol* 29(5): 487-494.

Byrd, R. H., Chin, G. M., Nocedal, J. and Wu, Y. (2012). Sample size selection in optimization methods for machine learning. *Mathematical programming* 134(1): 127-155.

Candia, J., Cherukuri, S., Guo, Y., Doshi, K. A., Banavar, J. R., Civin, C. I. and Losert, W. (2015). Uncovering low-dimensional, miR-based signatures of acute myeloid and lymphoblastic leukemias with a machine-learning-driven network approach. *Converg Sci Phys Oncol* 1(2).

Candoni, A., De Marchi, F., Zannier, M. E., Lazzarotto, D., Fili, C., Dubbini, M. V., Rabassi, N., Toffoletti, E., Lau, B. W. and Fanin, R. (2017). High prognostic value of pre-allogeneic stem cell transplantation minimal residual disease detection by WT1 gene expression in AML transplanted in cytologic complete remission. *Leuk Res* 63: 22-27.

Capper, D., Jones, D. T. W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D. E., Kratz, A., Wefers, A. K., Huang, K., Pajtler, K. W., Schweizer, L., Stichel, D., Olar, A., Engel, N. W., Lindenberg, K., Harter, P. N., Braczynski, A. K., Plate, K. H., Dohmen, H., Garvalov, B. K., Coras, R., Holsken, A., Hewer, E., Bewerunge-Hudler, M., Schick, M., Fischer, R., Beschoner, R., Schittenhelm, J., Staszewski, O., Wani, K., Varlet, P., Pages, M., Temming, P., Lohmann, D., Selt, F., Witt, H., Milde, T., Witt, O., Aronica, E., Giangaspero, F., Rushing, E., Scheurlen, W., Geisenberger, C., Rodriguez, F. J., Becker, A., Preusser, M., Haberler, C., Bjerkvig, R., Cryan, J., Farrell, M., Deckert, M., Hench, J., Frank, S., Serrano, J., Kannan, K., Tsirigos, A., Bruck, W., Hofer, S., Brehmer, S., Seiz-Rosenhagen, M., Hanggi, D., Hans, V., Rozsnoki, S., Hansford, J. R., Kohlhof, P., Kristensen, B. W., Lechner, M., Lopes, B., Mawrin, C., Ketter, R., Kulozik, A., Khatib, Z., Heppner, F., Koch, A., Jouvett, A., Keohane, C., Muhleisen, H., Mueller, W., Pohl, U., Prinz, M., Benner, A., Zapatka, M., Gottardo, N. G., Driever, P. H., Kramm, C. M., Muller, H. L., Rutkowski, S., von Hoff, K., Fruhwald, M. C., Gnekow, A., Fleischhack, G., Tippelt, S., Calaminus, G., Monoranu, C. M., Perry, A., Jones, C., Jacques, T. S.,

Radlwimmer, B., Gessi, M., Pietsch, T., Schramm, J., Schackert, G., Westphal, M., Reifemberger, G., Wesseling, P., Weller, M., Collins, V. P., Blumcke, I., Bendszus, M., Debus, J., Huang, A., Jabado, N., Northcott, P. A., Paulus, W., Gajjar, A., Robinson, G. W., Taylor, M. D., Jaunmuktane, Z., Ryzhova, M., Platten, M., Unterberg, A., Wick, W., Karajannis, M. A., Mittelbronn, M., Acker, T., Hartmann, C., Aldape, K., Schuller, U., Buslei, R., Lichter, P., Kool, M., Herold-Mende, C., Ellison, D. W., Hasselblatt, M., Snuderl, M., Brandner, S., Korshunov, A., von Deimling, A. and Pfister, S. M. (2018). DNA methylation-based classification of central nervous system tumours. *Nature* 555(7697): 469-474.

Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., Schrock, A., Campbell, B., Shlien, A., Chmielecki, J., Huang, F., Yuting, H., Sun, J., Tabori, U., Kennedy, M., Lieber, D. S., Roels, S., White, J., Otto, G. A. and Ross, J. S. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine* 9: 1-14.

Chatterton, Z., Morenos, L., Mechinaud, F., Ashley, D. M., Craig, J. M., Sexton-Oates, A., Halemba, M. S., Parkinson-Bates, M., Ng, J., Morrison, D., Carroll, W. L., Saffery, R. and Wong, N. C. (2014). Epigenetic deregulation in pediatric acute lymphoblastic leukemia. *Epigenetics* 9(3): 459-467.

Cheng, C. K., Li, L., Cheng, S. H., Lau, K. M., Chan, N. P., Wong, R. S., Shing, M. M., Li, C. K. and Ng, M. H. (2008). Transcriptional repression of the RUNX3/AML2 gene by the t (8; 21) and inv (16) fusion proteins in acute myeloid leukemia. *Blood* 112(8): 3391-3402.

Chiesa, R. and Veys, P. (2014). Reduced intensity conditioning in paediatric haematopoietic cell transplantation. Controversies in Pediatric and Adolescent Hematology, Karger Publishers. 17: 116-134.

Cole, C. B., Verdoni, A. M., Ketkar, S., Leight, E. R., Russler-Germain, D. A., Lamprecht, T. L., Demeter, R. T., Magrini, V. and Ley, T. J. (2015). PML-RARA requires DNA methyltransferase 3A to initiate acute promyelocytic leukemia. *J Clin Invest*.

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414): 57-74.

Consortium, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Coolen, M. W., Statham, A. L., Gardiner-Garden, M. and Clark, S. J. (2007). Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. *Nucleic Acids Res* 35(18): e119.

Cooper, T. M., Franklin, J., Gerbing, R. B., Alonzo, T. A., Hurwitz, C., Raimondi, S. C., Hirsch, B., Smith, F. O., Mathew, P. and Arceci, R. J. (2012). AAML03P1, a pilot study of the

safety of gemtuzumab ozogamicin in combination with chemotherapy for newly diagnosed childhood acute myeloid leukemia: a report from the Children's Oncology Group. *Cancer* 118(3): 761-769.

Cordaux, R. and Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* 10: 691.

Costello, J. F., Frühwald, M. C., Smiraglia, D. J., Rush, L. J., Robertson, G. P., Gao, X., Wright, F. A., Feramisco, J. D., Peltomäki, P., Lang, J. C., Schuller, D. E., Yu, L., Bloomfield, C. D., Caligiuri, M. A., Yates, A., Nishikawa, R., Su Huang, H. J., Petrelli, N. J., Zhang, X., O'Dorisio, M. S., Held, W. A., Cavenee, W. K. and Plass, C. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nature Genetics* 24: 132.

Creutzig, U., Zimmermann, M., Bourquin, J.-P., Dworzak, M. N., Fleischhack, G., Graf, N., Klingebiel, T., Kremens, B., Lehrnbecher, T. and von Neuhoff, C. (2013). Randomized trial comparing liposomal daunorubicin with idarubicin as induction for pediatric acute myeloid leukemia: results from Study AML-BFM 2004. *Blood* 122(1): 37-43.

Creutzig, U., Zimmermann, M., Reinhardt, D., Rasche, M., von Neuhoff, C., Alpermann, T., Dworzak, M., Perglerova, K., Zemanova, Z., Tchinda, J., Bradtke, J., Thiede, C. and Haferlach, C. (2016). Changes in cytogenetics and molecular genetics in acute myeloid leukemia from childhood to adult age groups. *Cancer*.

Creutzig, U., Zimmermann, M., Ritter, J., Reinhardt, D., Hermann, J., Henze, G., Jurgens, H., Kabisch, H., Reiter, A. and Riehm, H. (2005). Treatment strategies and long-term results in paediatric patients treated in four consecutive AML-BFM trials. Great Britain, Nature Publishing Group: 2030.

Crispino, J. D. (2005). GATA1 mutations in Down syndrome: implications for biology and diagnosis of children with transient myeloproliferative disorder and acute megakaryoblastic leukemia. *Pediatr Blood Cancer* 44(1): 40-44.

Dallosso, A. R., Hancock, A. L., Brown, K. W., Williams, A. C., Jackson, S. and Malik, K. (2004). Genomic imprinting at the WT1 gene involves a novel coding transcript (AWT1) that shows deregulation in Wilms' tumours. *Hum Mol Genet* 13(4): 405-415.

Dallosso, A. R., Hancock, A. L., Malik, S., Salpekar, A., King-Underwood, L., Pritchard-Jones, K., Peters, J., Moorwood, K., Ward, A., Malik, K. T. A. and Brown, K. W. (2007). Alternately spliced WT1 antisense transcripts interact with WT1 sense RNA and show epigenetic and splicing defects in cancer. *RNA (New York, N.Y.)* 13(12): 2287-2299.

Danielsson, A., Nemes, S., Tisell, M., Lannering, B., Nordborg, C., Sabel, M. and Caren, H. (2015). MethPed: a DNA methylation classifier tool for the identification of pediatric brain tumor subtypes. *Clin Epigenetics* 7(1): 62.

Daura-Oller, E., Cabre, M., Montero, M. A., Paternain, J. L. and Romeu, A. (2009). Specific gene hypomethylation and cancer: New insights into coding region feature trends. *Bioinformatics* 3(8): 340-343.

de Figueiredo, A. F., Mkrtchyan, H., Liehr, T., Ventura, E. M. S., de Jesus Marques-Salles, T., Santos, N., Ribeiro, R. C., Abdelhay, E. and Silva, M. L. M. (2009). A case of childhood acute myeloid leukemia AML (M5) with a neocentric chromosome neo (1)(qter→ q23~ 24:: q23~ 24→ q43→ neo→ q43→ qter) and tetrasomy of chromosomes 8 and 21. *Cancer Genetics and Cytogenetics* 193(2): 123-126.

Demokan, S., Chuang, A. Y., Pattani, K. M., Sidransky, D., Koch, W. and Califano, J. A. (2014). Validation of nucleolar protein 4 as a novel methylated tumor suppressor gene in head and neck cancer. *Oncology Reports* 31(2): 1014-1020.

Deneberg, S., Grövdal, M., Karimi, M., Jansson, M., Nahi, H., Corbacioglu, A., Gaidzik, V., Döhner, K., Paul, C., Ekström, T. J., Hellström-Lindberg, E. and Lehmann, S. (2010). Gene-specific and global methylation patterns predict outcome in patients with acute myeloid leukemia. *Leukemia* 24: 932.

Deneberg, S., Guardiola, P., Lennartsson, A., Qu, Y., Gaidzik, V., Blanchet, O., Karimi, M., Bengtzen, S., Nahi, H. and Uggla, B. (2011). Prognostic DNA methylation patterns in cytogenetically normal acute myeloid leukemia are predefined by stem cell chromatin marks. *Blood* 118(20): 5573-5582.

Dennis, K., Fan, T., Geiman, T., Yan, Q. and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev* 15(22): 2940-2944.

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D., McMichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., Schmidt, H., Kalicki-Veizer, J., Magrini, V. J., Cook, L., McGrath, S. D., Vickery, T. L., Wendl, M. C., Heath, S., Watson, M. A., Link, D. C., Tomasson, M. H., Shannon, W. D., Payton, J. E., Kulkarni, S., Westervelt, P., Walter, M. J., Graubert, T. A., Mardis, E. R., Wilson, R. K. and DiPersio, J. F. (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481(7382): 506-510.

Dohner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F. R., Buchner, T., Dombret, H., Ebert, B. L., Fenaux, P., Larson, R. A., Levine, R. L., Lo-Coco, F., Naoe, T., Niederwieser, D., Ossenkoppele, G. J., Sanz, M., Sierra, J., Tallman, M. S., Tien, H. F., Wei, A. H., Lowenberg, B. and Bloomfield, C. D. (2016). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*.

Dohner, H., Estey, E. H., Amadori, S., Appelbaum, F. R., Buchner, T., Burnett, A. K., Dombret, H., Fenaux, P., Grimwade, D., Larson, R. A., Lo-Coco, F., Naoe, T., Niederwieser,

D., Ossenkoppele, G. J., Sanz, M. A., Sierra, J., Tallman, M. S., Lowenberg, B. and Bloomfield, C. D. (2010). Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 115(3): 453-474.

Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., Miller, J., Schlaeger, T., Daley, G. Q. and Feinberg, A. P. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 41(12): 1350-1353.

Downing, J. R., Wilson, R. K., Zhang, J., Mardis, E. R., Pui, C. H., Ding, L., Ley, T. J. and Evans, W. E. (2012). The Pediatric Cancer Genome Project. *Nat Genet* 44(6): 619-622.

Du, T., Zhang, B., Zhang, S., Jiang, X., Zheng, P., Li, J., Yan, M., Zhu, Z. and Liu, B. (2016). Decreased expression of long non-coding RNA WT1-AS promotes cell proliferation and invasion in gastric cancer. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1862(1): 12-19.

Eccleston, A., DeWitt, N., Gunter, C., Marte, B. and Nath, D. (2007). Epigenetics. *Nature*(7143): 395.

Ehrich, M., Nelson, M. R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C. R., Field, J. K. and van den Boom, D. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* 102(44): 15785-15790.

Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. *Oncogene* 21(35): 5400.

Ehrlich, M., Gama-Sosa, M. A., Huang, L., Midgett, R. M., Kuo, K. C., McCune, R. A. and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research* 10(B): 2709 - 2721.

El Bordiny, M., Al-Ghandour, A., Abo Elwafa, R. A. and Fayed, O. (2018). The clinical significance of the alternative Wilms tumor gene overexpression-hypermethylation signature in acute myeloid leukemia. *Clin Transl Oncol*.

Elias, M. H., Azlan, H., Sulong, S., Baba, A. A. and Ankathil, R. (2018). Aberrant DNA methylation at HOXA4 and HOXA5 genes are associated with resistance to imatinib mesylate among chronic myeloid leukemia patients. *Cancer Reports* 1(2): e1111.

Estecio, M. R., Gharibyan, V., Shen, L., Ibrahim, A. E., Doshi, K., He, R., Jelinek, J., Yang, A. S., Yan, P. S., Huang, T. H., Tajara, E. H. and Issa, J. P. (2007). LINE-1 hypomethylation

in cancer is highly variable and inversely correlated with microsatellite instability. *PLoS One* 2(5): e399.

Estécio, M. R., Maddipoti, S., Bueso-Ramos, C., DiNardo, C. D., Yang, H., Wei, Y., Kondo, K., Fang, Z., Stevenson, W. and Chang, K. S. (2015). RUNX 3 promoter hypermethylation is frequent in leukaemia cell lines and associated with acute myeloid leukaemia inv (16) subtype. *British Journal Of Haematology* 169(3): 344-351.

Estecio, M. R., Maddipoti, S., DiNardo, C. D., Yang, H., Stevenson, W. S., Bueso-Ramos, C. E., Pierce, S. R., Wei, Y. and Garcia-Manero, G. (2014). Association Between RUNX3 Hypermethylation and Acute Myeloid Leukemia Inv(16) Subtype. *Blood* 124(21): 3548.

Esteller, M. (2005). Aberrant DNA methylation as a cancer-inducing mechanism. *Annu Rev Pharmacol Toxicol* 45: 629-656.

Esteller, M. (2008). Epigenetics in cancer. *New England Journal of Medicine* 358(11): 1148-1159.

Estey, E. (2016). Acute myeloid leukemia: 2016 Update on risk-stratification and management. *Am J Hematol* 91(8): 824-846.

Estey, E. and Dohner, H. (2006). Acute myeloid leukaemia. *The Lancet*(9550): 1894.

Estey, E. H. (2014). Acute myeloid leukemia: 2014 update on risk-stratification and management. *Am J Hematol* 89(11): 1063-1081.

Faderl, S., Talpaz, M., Estrov, Z., O'Brien, S., Kurzrock, R. and Kantarjian, H. M. (1999). The biology of chronic myeloid leukemia. *The New England Journal Of Medicine*(3): 164.

Farrar, J. E., Schuback, H. L., Ries, R. E., Wai, D., Hampton, O. A., Trevino, L. R., Alonzo, T. A., Guidry Auvil, J. M., Davidsen, T. M., Gesuwan, P., Hermida, L., Muzny, D. M., Dewal, N., Rustagi, N., Lewis, L. R., Gamis, A. S., Wheeler, D. A., Smith, M. A., Gerhard, D. S. and Meshinchi, S. (2016). Genomic profiling of pediatric acute myeloid leukemia reveals a changing mutational landscape from disease diagnosis to relapse. *Cancer Res*.

Feinberg, A. P., Ohlsson, R. and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. *Nat Rev Genet* 7(1): 21-33.

Feinberg, A. P. and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301: 89.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 136(5): E359-386.

Ferrara, J. L. M., Levine, J. E., Reddy, P. and Holler, E. (2009). Graft-versus-host disease. *The Lancet* 373(9674): 1550-1561.

Ferreira, H. J., Heyn, H., Vizoso, M., Moutinho, C., Vidal, E., Gomez, A., Martinez-Cardus, A., Simo-Riudalbas, L., Moran, S., Jost, E. and Esteller, M. (2015). DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia. *Oncogene*.

Figueroa, M. E., Chen, S.-C., Andersson, A. K., Phillips, L. A., Li, Y., Sotzen, J., Kundu, M., Downing, J. R., Melnick, A. and Mullighan, C. G. (2013). Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *J Clin Invest* 123(7): 3099-3111.

Figueroa, M. E., Lugthart, S., Li, Y., Erpelinck-Verschueren, C., Deng, X., Christos, P. J., Schifano, E., Booth, J., van Putten, W., Skrabanek, L., Campagne, F., Mazumdar, M., Grealley, J. M., Valk, P. J., Lowenberg, B., Delwel, R. and Melnick, A. (2010). DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* 17(1): 13-27.

Figueroa, M. E., Lugthart, S., Li, Y., Erpelinck-Verschueren, C., Deng, X., Christos, P. J., Schifano, E., Booth, J., van Putten, W., Skrabanek, L., Campagne, F., Mazumdar, M., Grealley, J. M., Valk, P. J. M., Löwenberg, B., Delwel, R. and Melnick, A. (2010). DNA Methylation Signatures Identify Biologically Distinct Subtypes in Acute Myeloid Leukemia. *Cancer Cell* 17(1): 13-27.

Figueroa, M. E., Skrabanek, L., Li, Y., Jiemjit, A., Fandy, T. E., Paietta, E., Fernandez, H., Tallman, M. S., Grealley, J. M., Carraway, H., Licht, J. D., Gore, S. D. and Melnick, A. (2009). MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. *Blood* 114(16): 3448.

Flores, K. B., Wolschin, F. and Amdam, G. V. (2013). The role of methylation of DNA in environmental adaptation. *Integr Comp Biol* 53(2): 359-372.

Franklin, J., Alonzo, T., Hurwitz, C. A., Gerbing, R., Smith, F., Arceci, R. J., Feusner, J., Mathew, P., Iannone, R., Meshinchi, S., Lavey, R., Wallace, J. D., Anderson, K., Winter, L., Glick, K., Hirsch, B., Raimondi, S. C. and Gamis, A. (2008). COG AAML03P1: Efficacy and Safety in a Pilot Study of Intensive Chemotherapy Including Gemtuzumab in Children Newly Diagnosed with Acute Myeloid Leukemia (AML). *Blood* 112(11): 136-136.

Friedman, A. D. (1999). Leukemogenesis by CBF oncoproteins. *Leukemia (08876924)* 13(12): 1932.

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M. and Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 16(10): 906-914.

Furuhata, A., Murakami, M., Ito, H., Gao, S., Yoshida, K., Sobue, S., Kikuchi, R., Iwasaki, T., Takagi, A., Kojima, T., Suzuki, M., Abe, A., Naoe, T. and Murate, T. (2009). GATA-1 and GATA-2 binding to 3' enhancer of WT1 gene is essential for its transcription in acute leukemia and solid tumor cell lines. *Leukemia* 23: 1270.

Furuta, J., Nobeyama, Y., Umebayashi, Y., Otsuka, F., Kikuchi, K. and Ushijima, T. (2006). Silencing of Peroxiredoxin 2 and aberrant methylation of 33 CpG islands in putative promoter regions in human malignant melanomas. *Cancer Res* 66(12): 6080-6086.

Gaidzik, V. I., Schlenk, R. F., Moschny, S., Becker, A., Bullinger, L., Corbacioglu, A., Krauter, J., Schlegelberger, B., Ganser, A. and Döhner, H. (2009). Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group. *Blood* 113(19): 4505-4511.

Gaidzik, V. I., Weber, D., Paschka, P., Kaumanns, A., Krieger, S., Corbacioglu, A., Kronke, J., Kapp-Schwoerer, S., Kramer, D., Horst, H. A., Schmidt-Wolf, I., Held, G., Kundgen, A., Ringhoffer, M., Gotze, K., Kindler, T., Fiedler, W., Wattad, M., Schlenk, R. F., Bullinger, L., Teleanu, V., Schlegelberger, B., Thol, F., Heuser, M., Ganser, A., Dohner, H. and Dohner, K. (2017). DNMT3A mutant transcript levels persist in remission and do not predict outcome in patients with acute myeloid leukemia. *Leukemia*.

Galm, O., Wilop, S., Lüders, C., Jost, E., Gehbauer, G., Herman, J. G. and Osieka, R. (2005). Clinical implications of aberrant DNA methylation patterns in acute myelogenous leukemia. *Annals of Hematology* 84(1): 39-46.

Gama-Sosa, M. A., Slagel, V. A., Trewyn, R. W., Oxenhandler, R., Kuo, K. C., Gehrke, C. W. and Ehrlich, M. (1983). The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Research* 11(19): 6883-6894.

Garcia-Cruz, R., Brieno, M. A., Roig, I., Grossmann, M., Velilla, E., Pujol, A., Cabero, L., Pessarrodona, A., Barbero, J. L. and Garcia Caldes, M. (2010). Dynamics of cohesin proteins REC8, STAG3, SMC1 beta and SMC3 are consistent with a role in sister chromatid cohesion during meiosis in human oocytes. *Hum Reprod* 25(9): 2316-2327.

Garcia-Manero, G., Bueso-Ramos, C., Daniel, J., Williamson, J., Kantarjian, H. M. and Issa, J.-P. J. (2002). DNA methylation patterns at relapse in adult acute lymphocytic leukemia. *Clinical Cancer Research* 8(6): 1897-1903.

Garcia-Manero, G., Gore, S. D., Cogle, C., Ward, R., Shi, T., MacBeth, K. J., Laille, E., Giordano, H., Sakoian, S., Jabbour, E., Kantarjian, H. and Skikne, B. (2011). Phase I Study

of Oral Azacitidine in Myelodysplastic Syndromes, Chronic Myelomonocytic Leukemia, and Acute Myeloid Leukemia. *Journal of Clinical Oncology* 29(18): 2521-2527.

Garg, M., Nagata, Y., Kanojia, D., M, T. A., Yoshida, K., Keloth, S. H., Jiang, Z. Z., Okuno, Y., Shiraishi, Y., Chiba, K., Tanaka, H., Miyano, S., Ding, L. W., Alpermann, T., Sun, Q. Y., Lin, D. C., Chien, W., Madan, V., Liu, L. Z., Tan, K. T., Sampath, A., Venkatesan, S., Inokuchi, K., Wakita, S., Yamaguchi, H., Chng, W. J., Kham, S. Y., Yeoh, A. E., Sanada, M., Schiller, J., Kreuzer, K. A., Kornblau, S. M., Kantarjian, H. M., Haferlach, T., Lill, M., Kuo, M. C., Shih, L. Y., Blau, I. W., Blau, O., Yang, H., Ogawa, S. and Koeffler, H. P. (2015). Profiling of somatic mutations of acute myeloid leukemia, FLT3-ITD subgroup at diagnosis and relapse. *Blood*.

Geyh, S., Rodriguez-Paredes, M., Jager, P., Koch, A., Bormann, F., Gutekunst, J., Zilkens, C., Germing, U., Kobbe, G., Lyko, F., Haas, R. and Schroeder, T. (2018). Transforming growth factor beta1-mediated functional inhibition of mesenchymal stromal cells in myelodysplastic syndromes and acute myeloid leukemia. *Haematologica* 103(9): 1462-1471.

Gibson, B. E., Wheatley, K., Hann, I. M., Stevens, R. F., Webb, D., Hills, R. K., De Graaf, S. S. and Harrison, C. J. (2005). Treatment strategy and long-term results in paediatric patients treated in consecutive UK AML trials. Great Britain, Nature Publishing Group: 2130.

Goemans, B. F., Tamminga, R. Y., Corbijn, C. M., Hahlen, K. and Kaspers, G. J. (2008). Outcome for children with relapsed acute myeloid leukemia in the Netherlands following initial treatment between 1980 and 1998: survival after chemotherapy only? *Haematologica* 93(9): 1418-1420.

Goldberg, A. D., Allis, C. D. and Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* 128(4): 635-638.

Goldman, S. L., Hassan, C., Khunte, M., Soldatenko, A., Jong, Y., Afshinnkoo, E. and Mason, C. E. (2019). Epigenetic Modifications in Acute Myeloid Leukemia: Prognosis, Treatment, and Heterogeneity. *Frontiers in genetics* 10: 133-133.

Goltz, D., Gevensleben, H., Grunen, S., Dietrich, J., Kristiansen, G., Landsberg, J. and Dietrich, D. (2016). PD-L1 (CD274) promoter methylation predicts survival in patients with acute myeloid leukemia. *Leukemia*.

Gorman, M. F., Ji, L., Ko, R. H., Barnette, P., Bostrom, B., Hutchinson, R., Raetz, E., Seibel, N. L., Twist, C. J., Eckroth, E., Sposto, R., Gaynon, P. S. and Loh, M. L. (2010). Outcome for children treated for relapsed or refractory acute myelogenous leukemia (rAML): a Therapeutic Advances in Childhood Leukemia (TACL) Consortium study. *Pediatr Blood Cancer* 55(3): 421-429.

Green, A., Whiteman, D., Frost, C. and Battistutta, D. (1999). Sun Exposure, Skin Cancers and Related Skin Conditions. *J Epidemiol* 9(sup): 7-13.

Greer, J. M. and McCombe, P. A. (2012). The role of epigenetic mechanisms and processes in autoimmune disorders. *Biologics* 6: 307-327.

Griffiths, E. A., Gore, S. D., Hooker, C. M., Mohammad, H. P., McDevitt, M. A., Smith, B. D., Karp, J. E., Herman, J. G. and Carraway, H. E. (2010). Epigenetic differences in cytogenetically normal versus abnormal acute myeloid leukemia. *Epigenetics* 5(7): 590-600.

Grisolano, J. L., Wesselschmidt, R. L., Pelicci, P. G. and Ley, T. J. (1997). Altered Myeloid Development and Acute Leukemia in Transgenic Mice Expressing PML-RAR a Under Control of Cathepsin G Regulatory Sequences. *Blood* 89(2): 276-387.

Grosveld, G. C. (2007). MN1, a novel player in human AML. *Blood Cells, Molecules, and Diseases* 39(3): 336-339.

Guo, H., Zhang, T. J., Wen, X. M., Zhou, J. D., Ma, J. C., An, C., Zhang, W., Xu, Z. J., Lin, J. and Qian, J. (2017). Hypermethylation of secreted frizzled-related proteins predicts poor prognosis in non-M3 acute myeloid leukemia. *Onco Targets Ther* 10: 3635-3644.

Guo, Q., Luan, J., Li, N., Zhang, Z., Zhu, X., Zhao, L., Wei, R., Sun, L., Shi, Y., Yin, X., Ding, N., Jiang, G. and Li, X. (2017). MicroRNA-181 as a prognostic biomarker for survival in acute myeloid leukemia: a meta-analysis. *Oncotarget* 8(51): 89130-89141.

Guo, R. J., Atenafu, E. G., Schimmer, A. D., Minden, M. D. and Chang, H. (2017). Expression of CD4 is correlated with an unfavorable prognosis in wild-type NPM1, FLT3-ITD-negative cytogenetically normal adult acute myeloid leukemia. *Int J Lab Hematol*.

Guo, X., Liu, W., Pan, Y., Ni, P., Ji, J., Guo, L., Zhang, J., Wu, J., Jiang, J., Chen, X., Cai, Q., Li, J., Zhang, J., Gu, Q., Liu, B., Zhu, Z. and Yu, Y. (2010). Homeobox gene IRX1 is a tumor suppressor gene in gastric carcinoma. *Oncogene* 29: 3908.

Hanahan, D. and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* 103(1): 57.

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell* 144(5): 646-674.

Hashim, D. and Boffetta, P. (2014). Occupational and environmental exposures and cancers in developing countries. *Annals Of Global Health* 80(5): 393-411.

Hasle, H., Abrahamsson, J., Forestier, E., Ha, S.-Y., Heldrup, J., Jahnukainen, K., Jónsson, Ó. G., Lausen, B., Palle, J. and Zeller, B. (2012). Gemtuzumab ozogamicin as

postconsolidation therapy does not prevent relapse in children with AML: results from NOPHO-AML 2004. *Blood* 120(5): 978-984.

Hefazi, M., Siddiqui, M., Patnaik, M., Wolanskyj, A., Alkhateeb, H., Zblewski, D., Elliott, M., Hogan, W., Litzow, M. and Al-Kali, A. (2015). Prognostic impact of combined NPM1+/FLT3-genotype in patients with acute myeloid leukemia with intermediate risk cytogenetics stratified by age and treatment modalities. *Leuk Res*.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M. and Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459: 108.

Herman, J. G. (1999). Hypermethylation of tumor suppressor genes in cancer. *Semin Cancer Biol* 9(5): 359-367.

Heuser, M., Beutel, G., Krauter, J., Döhner, K., von Neuhoff, N., Schlegelberger, B. and Ganser, A. (2006). High meningeoma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. *Blood* 108(12): 3898-3905.

Hirsch, P., Zhang, Y., Tang, R., Joulin, V., Boutroux, H., Pronier, E., Moatti, H., Flandrin, P., Marzac, C., Bories, D., Fava, F., Mokrani, H., Betems, A., Lorre, F., Favier, R., Feger, F., Mohty, M., Douay, L., Legrand, O., Billhou-Nabera, C., Louache, F. and Delhommeau, F. (2016). Genetic hierarchy and temporal variegation in the clonal history of acute myeloid leukaemia. *Nat Commun* 7: 12475.

Ho, P. A., Alonzo, T. A., Gerbing, R. B., Pollard, J., Stirewalt, D. L., Hurwitz, C., Heerema, N. A., Hirsch, B., Raimondi, S. C., Lange, B., Franklin, J. L., Radich, J. P. and Meshinchi, S. (2009). Prevalence and prognostic implications of CEBPA mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group. *Blood* 113(26): 6558-6566.

Ho, P. A., Alonzo, T. A., Gerbing, R. B., Pollard, J., Stirewalt, D. L., Hurwitz, C., Heerema, N. A., Hirsch, B., Raimondi, S. C., Lange, B., Franklin, J. L., Radich, J. P. and Meshinchi, S. (2009). Prevalence and prognostic implications of CEBPA mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group. *Blood* 113(26): 6558-6566.

Ho, P. A., Kutny, M. A., Alonzo, T. A., Gerbing, R. B., Joaquin, J., Raimondi, S. C., Gamis, A. S. and Meshinchi, S. (2011). Leukemic mutations in the methylation-associated genes DNMT3A and IDH2 are rare events in pediatric AML: A report from the Children's Oncology Group. *Pediatric blood & cancer* 57(2): 204-209.

Ho, P. A., Zeng, R., Alonzo, T. A., Gerbing, R. B., Miller, K. L., Pollard, J. A., Stirewalt, D. L., Heerema, N. A., Raimondi, S. C., Hirsch, B., Franklin, J. L., Lange, B. and Meshinchi, S. (2010). Prevalence and prognostic implications of WT1 mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group. *Blood* 116(5): 702-710.

Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., Leiserson, M. D., Niu, B., McLellan, M. D., Uzunangelov, V., Zhang, J., Kandoth, C., Akbani, R., Shen, H., Omberg, L., Chu, A., Margolin, A. A., Van't Veer, L. J., Lopez-Bigas, N., Laird, P. W., Raphael, B. J., Ding, L., Robertson, A. G., Byers, L. A., Mills, G. B., Weinstein, J. N., Van Waes, C., Chen, Z., Collisson, E. A., Benz, C. C., Perou, C. M. and Stuart, J. M. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158(4): 929-944.

Hollein, A., Jeromin, S., Meggendorfer, M., Fasan, A., Nadarajah, N., Kern, W., Haferlach, C. and Haferlach, T. (2018). Minimal residual disease (MRD) monitoring and mutational landscape in AML with RUNX1-RUNX1T1: a study on 134 patients. *Leukemia*.

Hollink, I. H. I. M., van den Heuvel-Eibrink, M. M., Zimmermann, M., Balgobind, B. V., Arentsen-Peters, S. T. C. J. M., Alders, M., Willasch, A., Kaspers, G. J. L., Trka, J. and Baruchel, A. (2009). Clinical relevance of Wilms tumor 1 gene mutations in childhood acute myeloid leukemia. United States, AMERICAN SOCIETY OF HEMATOLOGY: 5951.

Hopp, L., Lübbert, M., Loeffler, M. and Galle, J. (2017). Targeting DNA hypermethylation: Computational modeling of DNA demethylation treatment of acute myeloid leukemia AU - Przybilla, Jens. *Epigenetics* 12(10): 886-896.

Hou, H.-A., Huang, T.-C., Lin, L.-I., Liu, C.-Y., Chen, C.-Y., Chou, W.-C., Tang, J.-L., Tseng, M.-H., Huang, C.-F., Chiang, Y.-C., Lee, F.-Y., Liu, M.-C., Yao, M., Huang, S.-Y., Ko, B.-S., Hsu, S.-C., Wu, S.-J., Tsay, W., Chen, Y.-C. and Tien, H.-F. (2010). WT1 mutation in 470 adult patients with acute myeloid leukemia: stability during disease evolution and implication of its incorporation into a survival scoring system. *Blood* 115(25): 5222.

Hou, H. A., Liu, C. Y., Kuo, Y. Y., Chou, W. C., Tsai, C. H., Lin, C. C., Lin, L. I., Tseng, M. H., Chiang, Y. C., Liu, M. C., Liu, C. W., Tang, J. L., Yao, M., Li, C. C., Huang, S. Y., Ko, B. S., Hsu, S. C., Chen, C. Y., Lin, C. T., Wu, S. J., Tsay, W. and Tien, H. F. (2016). Splicing factor mutations predict poor prognosis in patients with de novo acute myeloid leukemia. *Oncotarget*.

Houseman, E. A., Kile, M. L., Christiani, D. C., Ince, T. A., Kelsey, K. T. and Marsit, C. J. (2016). Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC Bioinformatics* 17(1): 259.

Hu, X., Chen, Y. and Zhao, Z. J. (2015). Chapter 17 - Structure, regulation, and function of TET family proteins. Epigenetic Gene Expression and Regulation. S. Huang, M. D. Litt and C. A. Blakey. Oxford, Academic Press: 379-395.

Hulegardh, E., Nilsson, C., Lazarevic, V., Garelius, H., Antunovic, P., Rangert Derolf, A., Mollgard, L., Uggla, B., Wennstrom, L., Wahlin, A., Hoglund, M., Juliusson, G., Stockelberg, D. and Lehmann, S. (2015). Characterization and prognostic features of secondary acute myeloid leukemia in a population-based setting: a report from the Swedish Acute Leukemia Registry. *Am J Hematol* 90(3): 208-214.

Hutter, J. J. (2010). Childhood leukemia. *Pediatrics In Review / American Academy Of Pediatrics* 31(6): 234-241.

Im, A. P., Sehgal, A. R., Carroll, M. P., Smith, B. D., Tefferi, A., Johnson, D. E. and Boyiadzis, M. (2014). DNMT3A and IDH mutations in acute myeloid leukemia and other myeloid malignancies: associations with prognosis and potential treatment strategies. *Leukemia* 28: 1774.

Ismail, M. M. and Abdulateef, N. A. (2016). Bone marrow T-cell percentage: A novel prognostic indicator in acute myeloid leukemia. *Int J Hematol*.

Jaenisch, R. and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33 Suppl: 245-254.

Jiang, Y., Dunbar, A., Gondek, L. P., Mohan, S., Rataul, M., O'Keefe, C., Sekeres, M., Sauntharajah, Y. and Maciejewski, J. P. (2009). Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. *Blood* 113(6): 1315-1325.

Jin, Z. and Liu, Y. (2018). DNA methylation in human diseases. *Genes & Diseases* 5(1): 1-8.

Jo, J., Nam, C. M., Sull, J. W., Yun, J. E., Kim, S. Y., Lee, S. J., Kim, Y. N., Park, E. J., Kimm, H. and Jee, S. H. (2012). Prediction of Colorectal Cancer Risk Using a Genetic Risk Score: The Korean Cancer Prevention Study-II (KCPS-II). *Genomics & informatics* 10(3): 175-183.

Jones, C., Perryman, L. and Hargrave, D. (2012). Paediatric and adult malignant glioma: close relatives or distant cousins? *Nat Rev Clin Oncol* 9(7): 400-413.

Jones, P. A. and Baylin, S. B. (2007). The epigenomics of cancer. *Cell* 128(4): 683-692.

Jung, N., Dai, B., Gentles, A. J., Majeti, R. and Feinberg, A. P. (2015). An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nat Commun* 6: 8489.

Kai, M., Yamamoto, E., Sato, A., Yamano, H. O., Niinuma, T., Kitajima, H., Harada, T., Aoki, H., Maruyama, R., Toyota, M., Hatahira, T., Nakase, H., Sugai, T., Yamashita, T., Toyota, M. and Suzuki, H. (2017). Epigenetic silencing of diacylglycerol kinase gamma in colorectal cancer. *Mol Carcinog* 56(7): 1743-1752.

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., Leiserson, M. D. M., Miller, C. A., Welch, J. S., Walter, M. J., Wendl, M. C., Ley, T. J., Wilson, R. K., Raphael, B. J. and Ding, L. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*(7471): 333.

Kaneda, A., Tsukamoto, T., Takamura-Enya, T., Watanabe, N., Kaminishi, M., Sugimura, T., Tatematsu, M. and Ushijima, T. (2004). Frequent hypomethylation in multiple promoter CpG islands is associated with global hypomethylation, but not with frequent promoter hypermethylation. *Cancer science* 95(1): 58-64.

Karsy, M., Guan, J. and Huang, L. E. (2018). Prognostic role of mitochondrial pyruvate carrier in isocitrate dehydrogenase-mutant glioma. *Journal of neurosurgery*: 1-11.

Kaspers, G. (2014). How I treat paediatric relapsed acute myeloid leukaemia. *Br J Haematol* 166(5): 636-645.

Kaspers, G. J. L. (2012). Pediatric acute myeloid leukemia. *Expert Rev Anticancer Ther* 12(3): 405-413.

Kelly, A. D., Kroeger, H., Yamazaki, J., Taby, R., Neumann, F., Yu, S., Lee, J. T., Patel, B., Li, Y., He, R., Liang, S., Lu, Y., Cesaroni, M., Pierce, S. A., Kornblau, S. M., Bueso-Ramos, C. E., Ravandi, F., Kantarjian, H. M., Jelinek, J. and Issa, J. P. (2017). A CpG island methylator phenotype in acute myeloid leukemia independent of IDH mutations and associated with a favorable outcome. *Leukemia*.

Kelly, A. D., Madzo, J., Madiredi, P., Kropf, P., Good, C. R., Jelinek, J. and Issa, J. J. (2018). Demethylator phenotypes in acute myeloid leukemia. *Leukemia*.

Kim, H., Seol, Y. M., Song, M. K., Choi, Y. J., Shin, H. J., Park, S. H., Lee, E. Y. and Chung, J. S. (2016). Evaluation of prognostic factors in patients with relapsed AML: Clonal evolution versus residual disease. *Blood Res* 51(3): 175-180.

Ko, Y. C., Fang, W. H., Lin, T. C., Hou, H. A., Chen, C. Y., Tien, H. F. and Lin, L. I. (2014). MicroRNA let-7a-3 gene methylation is associated with karyotyping, CEBPA promoter methylation, and survival in acute myeloid leukemia. *Leuk Res* 38(5): 625-631.

Kohli, R. M. and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502: 472.

Konopleva, M., Pollyea, D. A., Potluri, J., Chyla, B., Hogdal, L., Busman, T., McKeegan, E., Salem, A. H., Zhu, M., Ricker, J. L., Blum, W., DiNardo, C. D., Kadia, T., Dunbar, M., Kirby, R., Falotico, N., Levenson, J., Humerickhouse, R., Mabry, M., Stone, R., Kantarjian, H. and Letai, A. (2016). Efficacy and Biological Correlates of Response in a Phase II Study of Venetoclax Monotherapy in Patients with Acute Myelogenous Leukemia. *Cancer Discov* 6(10): 1106.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128(4): 693-705.

Krivtsov, A. V. and Armstrong, S. A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. *Nature Reviews Cancer* 7(11): 823.

Kroeger, H., Jelinek, J., Estécio, M. R., He, R., Kondo, K., Chung, W., Zhang, L., Shen, L., Kantarjian, H. M. and Bueso-Ramos, C. E. (2008). Aberrant CpG island methylation in acute myeloid leukemia is accentuated at relapse. *Blood* 112(4): 1366-1373.

Kühn, A., Löscher, D. and Marschalek, R. (2016). The IRX1/HOXA connection: insights into a novel t(4;11)- specific cancer mechanism. *Oncotarget* 7(23): 35341-35352.

Kühnl, A., Kaiser, M., Neumann, M., Fransecky, L., Heesch, S., Radmacher, M., Marcucci, G., Bloomfield, C. D., Hofmann, W.-K. and Thiel, E. (2011). High expression of IGFBP2 is associated with chemoresistance in adult acute myeloid leukemia. *Leukemia research* 35(12): 1585-1590.

Kumar, C. C. (2011). Genetic abnormalities and challenges in the treatment of acute myeloid leukemia. *Genes & Cancer* 2(2): 95-107.

Lamba, J. K., Cao, X., Raimondi, S. C., Rafiee, R., Downing, J. R., Lei, S., Gruber, T., Ribeiro, R. C., Rubnitz, J. E. and Pounds, S. B. (2018). Integrated epigenetic and genetic analysis identifies markers of prognostic significance in pediatric acute myeloid leukemia. *Oncotarget* 9(42): 26711-26723.

Lancet, J. E., Cortes, J. E., Hogge, D. E., Tallman, M. S., Kovacsosics, T. J., Damon, L. E., Komrokji, R., Solomon, S. R., Kolitz, J. E., Cooper, M., Yeager, A. M., Louie, A. C. and Feldman, E. J. (2014). Phase 2 trial of CPX-351, a fixed 5:1 molar ratio of cytarabine/daunorubicin, vs cytarabine/daunorubicin in older adults with untreated AML. *Blood* 123(21): 3239-3246.

Lange, B. J., Smith, F. O., Feusner, J., Barnard, D. R., Dinndorf, P., Feig, S., Heerema, N. A., Arndt, C., Arceci, R. J. and Seibel, N. (2008). Outcomes in CCG-2961, a Children's Oncology Group Phase 3 Trial for untreated pediatric acute myeloid leukemia: a report from the Children's Oncology Group. United States, AMERICAN SOCIETY OF HEMATOLOGY: 1044.

Lange, T., Hubmann, M., Burkhardt, R., Franke, G., Cross, M., Scholz, M., Leiblein, S., Al-Ali, H., Edelmann, J. and Thiery, J. (2011). Monitoring of WT1 expression in PB and CD34+ donor chimerism of BM predicts early relapse in AML and MDS patients after hematopoietic cell transplantation with reduced-intensity conditioning. *Leukemia* 25(3): 498.

Larmonie, N. S. D., Arentsen-Peters, T., Obulkasim, A., Valerio, D., Sonneveld, E., Danen-van Oorschot, A. A., de Haas, V., Reinhardt, D., Zimmermann, M., Trka, J., Baruchel, A., Pieters, R., van den Heuvel-Eibrink, M. M., Zwaan, C. M. and Fornerod, M. (2017). MN1 overexpression is driven by loss of DNMT3B methylation activity in inv(16) pediatric AML. *Oncogene*.

Larmonie, N. S. D., van den Heuvel-Eibrink, M. M., Obulkasim, A., de Haas, V., Reinhardt, D., Zimmermann, M., Trka, J., Baruchel, A., Stam, R. W., Pieters, R., Fornerod, M. W. J. and Zwaan, C. M. (2014). DNA Methylation Profiling of Pediatric AML Reveals That Hypomethylation of MN1 Is Characteristic of Inv(16) AML and a Driver of MN1 Overexpression. *Blood* 124(21): 867.

Lasa, A., Carnicer, M., Aventin, A., Estivill, C., Brunet, S., Sierra, J. and Nomdedeu, J. (2004). MEIS 1 expression is downregulated through promoter hypermethylation in AML1-ETO acute myeloid leukemias. *Leukemia* 18(7): 1231.

Laura, D. W., Parsons, D. W., Siân, J., Jimmy, L., Tobias, S., Rebecca, J. L., Dong, S., Simina, M. B., Thomas, B., Janine, P., Natalie, S., Steve, S., Zoltan, D., Vadim, U., Tatiana, N., Yuri, N., Rachel, K., Paul, A. W., Joshua, S. K., Zemin, Z., Randal, C., Joseph, W., Dawn, D., Michail, S., James, K. V. W., Saraswati, S., Kornelia, P., Ben Ho, P., Charit, L. P., Pant, P. V. K., Dennis, G. B., Andrew, B. S., James, H., Douglas, R. S., Erick, S., Nickolas, P., Phillip, B., Sanford, D. M., Giovanni, P., Kenneth, W. K., Victor, E. V. and Bert, V. (2007). The Genomic Landscapes of Human Breast and Colorectal Cancers. *Science*(5853): 1108.

Law, J. A. and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature reviews. Genetics* 11(3): 204-220.

Lawrence, H. J. and Largman, C. (2005). The HOX-MEIS connection in AML. *Blood* 106(1): 6.

LeBaron, S., Zeltzer, L. K., LeBaron, C., Scott, S. E. and Zeltzer, P. M. (1988). Chemotherapy side effects in pediatric oncology patients: drugs, age, and sex as risk factors. *Medical And Pediatric Oncology* 16(4): 263-268.

Lee, S. H., Chiu, Y. C., Li, Y. H., Lin, C. C., Hou, H. A., Chou, W. C. and Tien, H. F. (2017). High expression of dedicator of cytokinesis 1 (DOCK1) confers poor prognosis in acute myeloid leukemia. *Oncotarget* 8(42): 72250-72259.

Lee, S. I., Celik, S., Logsdon, B. A., Lundberg, S. M., Martins, T. J., Oehler, V. G., Estey, E. H., Miller, C. P., Chien, S., Dai, J., Saxena, A., Blau, C. A. and Becker, P. S. (2018). A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat Commun* 9(1): 42.

Lekanne, R. D., Riegman, P., Groen, N. A., Warringa, U. L., Molijn, A., Bootsma, D., Menon, A. and Kley, N. (1995). Cloning and characterization of MN1, a gene from chromosome 22q11, which is disrupted by a balanced translocation in a meningioma. *Oncogene* 10(8): 1521-1528.

Li, M., Gao, L., Li, Z., Sun, J., Zhang, H., Duan, H., Ma, Y. and Wang, C. (2016). CTNNA1 hypermethylation, a frequent event in acute myeloid leukemia, is independently associated with an adverse outcome. *Oncotarget*.

Li, S., Chen, L., Jin, W., Ma, X., Ma, Y., Dong, F., Zhu, H., Li, J. and Wang, K. (2017). Influence of body mass index on incidence and prognosis of acute myeloid leukemia and acute promyelocytic leukemia: A meta-analysis. *Sci Rep* 7(1): 17998.

Li, S., Garrett-Bakelman, F. E., Chung, S. S., Sanders, M. A., Hricik, T., Rapaport, F., Patel, J., Dillon, R., Vijay, P., Brown, A. L., Perl, A. E., Cannon, J., Bullinger, L., Luger, S., Becker, M., Lewis, I. D., To, L. B., Delwel, R., Lowenberg, B., Dohner, H., Dohner, K., Guzman, M. L., Hassane, D. C., Roboz, G. J., Grimwade, D., Valk, P. J., D'Andrea, R. J., Carroll, M., Park, C. Y., Neuberg, D., Levine, R., Melnick, A. M. and Mason, C. E. (2016). Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat Med*.

Li, S., Mason, C. and Melnick, A. (2016). Genetic and epigenetic heterogeneity in acute myeloid leukemia. *Curr Opin Genet Dev* 36: 100-106.

Li, X. X., Zhou, J. D., Zhang, T. J., Yang, L., Wen, X. M., Ma, J. C., Yang, J., Zhang, Z. H., Lin, J. and Qian, J. (2017). Epigenetic dysregulation of NKD2 is a valuable predictor assessing treatment outcome in acute myeloid leukemia. *J Cancer* 8(3): 460-468.

Li, Y., Lin, J., Yang, J., Qian, J., Qian, W., Yao, D. M., Deng, Z. Q., Liu, Q., Chen, X. X., Xie, D., An, C. and Tang, C. Y. (2013). Overexpressed let-7a-3 is associated with poor outcome in acute myeloid leukemia. *Leuk Res* 37(12): 1642-1647.

Li, Z., Herold, T., He, C., Valk, P. J., Chen, P., Jurinovic, V., Mansmann, U., Radmacher, M. D., Maharry, K. S., Sun, M., Yang, X., Huang, H., Jiang, X., Sauerland, M. C., Buchner, T., Hiddemann, W., Elkahlon, A., Neilly, M. B., Zhang, Y., Larson, R. A., Le Beau, M. M., Caligiuri, M. A., Dohner, K., Bullinger, L., Liu, P. P., Delwel, R., Marcucci, G., Lowenberg, B., Bloomfield, C. D., Rowley, J. D., Bohlander, S. K. and Chen, J. (2013). Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. *J Clin Oncol* 31(9): 1172-1181.

Liang, D.-C., Liu, H.-C., Yang, C.-P., Jaing, T.-H., Hung, I.-J., Yeh, T.-C., Chen, S.-H., Hou, J.-Y., Huang, Y.-J. and Shih, Y.-S. (2013). Cooperating gene mutations in childhood acute myeloid leukemia with special reference on mutations of ASXL1, TET2, IDH1, IDH2 and DNMT3A. *Blood*: blood-2012-2006-436782.

Lin, X., Wang, Z., Wang, Y. and Feng, W. (2015). Serum MicroRNA-370 as a potential diagnostic and prognostic biomarker for pediatric acute myeloid leukemia. *Int J Clin Exp Pathol* 8(11): 14658-14666.

Liu, J., Qin, Y. Z., Yang, S., Wang, Y., Chang, Y. J., Zhao, T., Jiang, Q. and Huang, X. J. (2017). Meis1 is critical to the maintenance of human acute myeloid leukemia cells independent of MLL rearrangements. *Ann Hematol* 96(4): 567-574.

Lobo, N. A., Shimono, Y., Qian, D. and Clarke, M. F. (2007). The biology of cancer stem cells. *Annu Rev Cell Dev Biol* 23: 675-699.

Locatelli, F., Masetti, R., Rondelli, R., Zecca, M., Fagioli, F., Rovelli, A., Messina, C., Lanino, E., Bertaina, A., Favre, C., Giorgiani, G., Ripaldi, M., Ziino, O., Palumbo, G., Pillon, M., Pession, A., Rutella, S. and Prete, A. (2015). Outcome of children with high-risk acute myeloid leukemia given autologous or allogeneic hematopoietic cell transplantation in the aieop AML-2002/01 study. *Bone Marrow Transplant* 50(2): 181-188.

Loeb, D. M., Evron, E., Patel, C. B., Sharma, P. M., Niranjana, B., Buluwela, L., Weitzman, S. A., Korz, D. and Sukumar, S. (2001). Wilms' tumor suppressor gene (WT1) is expressed in primary breast tumors despite tumor-specific promoter methylation. *Cancer Res* 61(3): 921-925.

Loeb, L. A., Loeb, K. R. and Anderson, J. P. (2003). Multiple mutations and cancer. *Proc Natl Acad Sci U S A* 100(3): 776-781.

Loke, Y. J., Galati, J. C., Saffery, R. and Craig, J. M. (2015). Association of in vitro fertilization with global and IGF2/H19 methylation variation in newborn twins. *Journal of Developmental Origins of Health and Disease* 6(2): 115-124.

Loke, Y. J., Muggli, E., Nguyen, L., Ryan, J., Saffery, R., Elliott, E. J., Halliday, J. and Craig, J. M. (2018). Time- and sex-dependent associations between prenatal alcohol exposure and placental global DNA methylation. *Epigenomics*.

Löwenberg, B., Beck, J., Graux, C., van Putten, W., Schouten, H. C., Verdonck, L. F., Ferrant, A., Sonneveld, P., Jongen-Lavrencic, M., von Lilienfeld-Toal, M., Biemond, B. J., Vellenga, E., Breems, D., de Muijnck, H., Schaafsma, R., Verhoef, G., Döhner, H., Gratwohl, A., Pabst, T., Ossenkoppele, G. J. and Maertens, J. (2010). Gemtuzumab ozogamicin as postremission treatment in AML at 60 years of age or more: results of a multicenter phase 3 study. *Blood* 115(13): 2586.

Luskin, M. R., Gimotty, P. A., Smith, C., Loren, A. W., Figueroa, M. E., Harrison, J., Sun, Z., Tallman, M. S., Paietta, E. M., Litzow, M. R., Melnick, A. M., Levine, R. L., Fernandez, H. F., Luger, S. M., Carroll, M., Master, S. R. and Wertheim, G. B. (2016). A clinical measure of DNA methylation predicts outcome in de novo acute myeloid leukemia. *JCI Insight* 1(9).

Maegawa, S., Gough, S. M., Watanabe-Okochi, N., Lu, Y., Zhang, N., Castoro, R. J., Estecio, M. R. H., Jelinek, J., Liang, S., Kitamura, T., Aplan, P. D. and Issa, J.-P. J. (2014). Age-related epigenetic drift in the pathogenesis of MDS and AML. *Genome Research* 24(4): 580-591.

Maksimovic, J., Gagnon-Bartsch, J. A., Speed, T. P. and Oshlack, A. (2015). Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. *Nucleic Acids Res* 43(16): e106.

Maksimovic, J., Gordon, L. and Oshlack, A. (2012). SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol* 13(6): R44.

Marlton, P., Keating, M., Kantarjian, H., Pierce, S., O'Brien, S., Freireich, E. J. and Estey, E. (1995). Cytogenetic and clinical correlates in AML patients with abnormalities of chromosome 16. *Leukemia* 9(6): 965-971.

Marschalek, R. (2015). MLL Leukemia and Future Treatment Strategies. *Archiv der Pharmazie* 348(4): 1-8.

Martino, D., Thanh, D., Sexton-Oates, A., Prescott, S., Tang, M. L. K., Shyamali, D., Gurrin, L., Koplin, J., Ponsonby, A. L., Allen, K. J. and Saffery, R. (2015). Blood DNA methylation biomarkers predict clinical reactivity in food-sensitized infants. *Journal of Allergy and Clinical Immunology* 135(5): 1319.

Marx, V. (2012). Reading the second genomic code. *Nature* 491: 143.

Matsuo, H., Nakamura, N., Tomizawa, D., Saito, A. M., Kiyokawa, N., Horibe, K., Nishinaka-Arai, Y., Tokumasu, M., Itoh, H., Kamikubo, Y., Nakayama, H., Kinoshita, A., Taga, T., Tawa, A., Taki, T., Tanaka, S. and Adachi, S. (2016). CXCR4 Overexpression is a Poor Prognostic Factor in Pediatric Acute Myeloid Leukemia With Low Risk: A Report From the Japanese Pediatric Leukemia/Lymphoma Study Group. *Pediatr Blood Cancer*.

Matsuzaki, K., Deng, G., Tanaka, H., Kakar, S., Miura, S. and Kim, Y. S. (2005). The relationship between global methylation level, loss of heterozygosity, and microsatellite instability in sporadic colorectal cancer. *Clin Cancer Res* 11(24 Pt 1): 8564-8569.

Melki, J. R. and Clark, S. J. (2002). DNA methylation changes in leukaemia. *Seminars in Cancer Biology* 12(5): 347-357.

Melki, J. R., Vincent, P. C. and Clark, S. J. (1999). Concurrent DNA Hypermethylation of Multiple Genes in Acute Myeloid Leukemia. *Cancer research* 59(15): 3730.

Meng, Y., Zou, Q., Liu, T., Cai, X., Huang, Y. and Pan, J. (2015). microRNA-335 inhibits proliferation, cell-cycle progression, colony formation, and invasion via targeting PAX6 in breast cancer cells. *Mol Med Rep* 11(1): 379-385.

Meshinchi, S., Alonzo, T. A., Stirewalt, D. L., Zwaan, M., Zimmerman, M., Reinhardt, D., Kaspers, G. J. L., Heerema, N. A., Gerbing, R., Lange, B. J. and Radich, J. P. (2006). Clinical implications of FLT3 mutations in pediatric AML. *Blood* 108(12): 3654-3661.

Metzeler, K. H., Herold, T., Rothenberg-Thurley, M., Amler, S., Sauerland, M. C., Goerlich, D., Schneider, S., Konstandin, N. P., Dufour, A., Braundl, K., Ksienzyk, B., Zellmeier, E., Hartmann, L., Greif, P. A., Fiegl, M., Subklewe, M., Bohlander, S. K., Krug, U., Faldum, A., Berdel, W. E., Wormann, B., Buchner, T., Hiddemann, W., Braess, J. and Spiekermann, K. (2016). Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. *Blood*.

Metzeler, K. H., Hummel, M., Bloomfield, C. D., Spiekermann, K., Braess, J., Sauerland, M. C., Heinecke, A., Radmacher, M., Marcucci, G., Whitman, S. P., Maharry, K., Paschka, P., Larson, R. A., Berdel, W. E., Buchner, T., Wormann, B., Mansmann, U., Hiddemann, W., Bohlander, S. K. and Buske, C. (2008). An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood* 112(10): 4193-4201.

Meyer, C., Burmeister, T., Groger, D., Tsaar, G., Fechina, L., Renneville, A., Sutton, R., Venn, N. C., Emerenciano, M., Pombo-de-Oliveira, M. S., Barbieri Blunck, C., Almeida Lopes, B., Zuna, J., Trka, J., Ballerini, P., Lapillonne, H., De Braekeleer, M., Cazzaniga, G., Corral Abascal, L., van der Velden, V. H. J., Delabesse, E., Park, T. S., Oh, S. H., Silva, M. L. M., Lund-Aho, T., Juvonen, V., Moore, A. S., Heidenreich, O., Vormoor, J., Zerkalenkova, E., Olshanskaya, Y., Bueno, C., Menendez, P., Teigler-Schlegel, A., Zur Stadt, U., Lentjes, J., Gohring, G., Kustanovich, A., Aleinikova, O., Schafer, B. W., Kubetzko, S., Madsen, H. O., Gruhn, B., Duarte, X., Gameiro, P., Lippert, E., Bidet, A., Cayuela, J. M., Clappier, E., Alonso, C. N., Zwaan, C. M., van den Heuvel-Eibrink, M. M., Izraeli, S., Trakhtenbrot, L., Archer, P., Hancock, J., Moricke, A., Alten, J., Schrappe, M., Stanulla, M., Strehl, S., Attarbaschi, A., Dworzak, M., Haas, O. A., Panzer-Grumayer, R., Sedek, L., Szczepanski, T., Caye, A., Suarez, L., Cave, H. and Marschalek, R. (2018). The MLL recombinome of acute leukemias in 2017. *Leukemia* 32(2): 273-284.

Milani, L., Lundmark, A., Kiialainen, A., Nordlund, J., Flaegstad, T., Forestier, E., Heyman, M., Jonmundsson, G., Kanerva, J., Schmiegelow, K., Söderhäll, S., Gustafsson, M. G., Lönnerholm, G. and Syvänen, A.-C. (2010). DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood* 115(6): 1214.

Mill, J. and Heijmans, B. T. (2013). From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet* 14(8): 585-594.

Moelans, C. B., Verschuur-Maes, A. H. and van Diest, P. J. (2011). Frequent promoter hypermethylation of BRCA2, CDH13, MSH6, PAX5, PAX6 and WT1 in ductal carcinoma in situ and invasive breast cancer. *J Pathol* 225(2): 222-231.

Mohr, S., Doebele, C., Comoglio, F., Berg, T., Beck, J., Bohnenberger, H., Alexe, G., Corso, J., Strobel, P., Wachter, A., Beissbarth, T., Schnutgen, F., Cremer, A., Haetscher, N., Gollner, S., Rouhi, A., Palmqvist, L., Rieger, M. A., Schroeder, T., Bonig, H., Muller-Tidow, C., Kuchenbauer, F., Schutz, E., Green, A. R., Urlaub, H., Stegmaier, K., Humphries, R. K., Serve, H. and Oellerich, T. (2017). Hoxa9 and Meis1 Cooperatively Induce Addiction to Syk Signaling by Suppressing miR-146a in Acute Myeloid Leukemia. *Cancer Cell* 31(4): 549-562.e511.

Morenos, L., Saffery, R., Mechinaud, F., Ashley, D., Elwood, N., Craig, J. M. and Wong, N. C. (2012). Evaluation of microRNA expression in patient bone marrow aspirate slides. *PLoS One* 7(8): e42951.

Morenos, L. C., Z.; Ng, J. L.; Halemba, M. S.; Parkinson-Bates, M.; Mechinaud, F.; Elwood, N.; Saffery, R.; Wong, N. C. (2014). Hypermethylation and down-regulation of DLEU2 in paediatric acute myeloid leukaemia independent of embedded tumour suppressor miR-15a/16-1. *Mol Cancer* 13(1): 1-24.

Mrozek, K., Marcucci, G., Nicolet, D., Maharry, K. S., Becker, H., Whitman, S. P., Metzeler, K. H., Schwind, S., Wu, Y. Z., Kohlschmidt, J., Pettenati, M. J., Heerema, N. A., Block, A. W., Patil, S. R., Baer, M. R., Kolitz, J. E., Moore, J. O., Carroll, A. J., Stone, R. M., Larson, R. A. and Bloomfield, C. D. (2012). Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. *J Clin Oncol* 30(36): 4515-4523.

Muntean, A. G. and Hess, J. L. (2012). The pathogenesis of mixed-lineage leukemia. *Annu Rev Pathol* 7: 283-301.

Musialik, E., Bujko, M., Kober, P., Grygorowicz, M. A., Libura, M., Przestrzelska, M., Juszczynski, P., Borg, K., Florek, I. and Jakóbczyk, M. (2015). Promoter DNA methylation and expression levels of HOXA4, HOXA5 and MEIS1 in acute myeloid leukemia. *Mol Med Rep* 11(5): 3948-3954.

Nakano, H., Fujiwara, S. I., Ito, S., Mashima, K., Umino, K., Minakata, D., Yamasaki, R., Kawasaki, Y., Sugimoto, M., Ashizawa, M., Yamamoto, C., Hatano, K., Okazuka, K., Sato, K., Oh, I., Ohmine, K., Suzuki, T., Muroi, K. and Kanda, Y. (2015). The prognostic significance of rapid peripheral blood blast clearance during the initial course of induction chemotherapy in young patients with de novo acute myeloid leukemia. *Hematol Oncol*.

Namasu, C. Y., Katzerke, C., Brauer-Hartmann, D., Wurm, A. A., Gerloff, D., Hartmann, J. U., Schwind, S., Muller-Tidow, C., Hilger, N., Fricke, S., Christopeit, M., Niederwieser, D. and Behre, G. (2017). ABR, a novel inducer of transcription factor C/EBPalpha, contributes to myeloid differentiation and is a favorable prognostic factor in acute myeloid leukemia. *Oncotarget* 8(61): 103626-103639.

Neale, R. E., Clark, P. J., Fawcett, J., Fritschi, L., Nagler, B. N., Risch, H. A., Walters, R. J., Crawford, W. J., Webb, P. M., Whiteman, D. C. and Buchanan, D. D. (2014). Association between hypermethylation of DNA repetitive elements in white blood cell DNA and pancreatic cancer. *Cancer Epidemiol* 38(5): 576-582.

Newcombe, A. A., Gibson, B. E. S. and Keeshan, K. (2018). Harnessing the potential of epigenetic therapies for childhood acute myeloid leukemia. *Exp Hematol* 63: 1-11.

Niavarani, A., Herold, T., Rey, Y., Sauerland, M. C., Buchner, T., Hiddemann, W., Bohlander, S. K., Valk, P. J. and Bonnet, D. (2015). A 4-gene expression score associated with high levels of Wilms Tumor-1 (WT1) expression is an adverse prognostic factor in acute myeloid leukaemia. *Br J Haematol*.

Niavarani, A., Horswell, S., Sadri, R. and Bonnet, D. (2015). The Wilms Tumor-1 (WT1) rs2234593 variant is a prognostic factor in normal karyotype acute myeloid leukemia. *Ann Hematol*.

Nibourel, O., Guihard, S., Roumier, C., Pottier, N., Terre, C., Paquet, A., Peyrouze, P., Geffroy, S., Quentin, S., Alberdi, A., Abdelali, R. B., Renneville, A., Demay, C., Celli-Lebras, K., Barbry, P., Quesnel, B., Castaigne, S., Dombret, H., Soulier, J., Preudhomme, C. and Cheok, M. H. (2016). Copy-number analysis identified new prognostic marker in acute myeloid leukemia. *Leukemia*.

Niederwieser, D., Lange, T., Cross, M., Basara, N. and Al-Ali, H. (2006). Reduced intensity conditioning (RIC) haematopoietic cell transplants in elderly patients with AML. *Best Pract Res Clin Haematol* 19(4): 825-838.

Nordlund, J., Bäcklin, C. L., Wahlberg, P., Busche, S., Berglund, E. C., Eloranta, M.-L., Flaegstad, T., Forestier, E., Frost, B.-M., Harila-Saari, A., Heyman, M., Jónsson, Ó. G., Larsson, R., Palle, J., Rönnblom, L., Schmiegelow, K., Sinnett, D., Söderhäll, S., Pastinen, T., Gustafsson, M. G., Lönnholm, G. and Syvänen, A.-C. (2013). Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome biology* 9(9).

Nordlund, J., Bäcklin, C. L., Zachariadis, V., Cavelier, L., Dahlberg, J., Öfverholm, I., Barbany, G., Nordgren, A., Övernäs, E. and Abrahamsson, J. (2015). DNA methylation-based subtype prediction for pediatric acute lymphoblastic leukemia. *Clinical Epigenetics* 7(1): 11.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* 99(3): 247-257.

Okuda, T., Cai, Z., Yang, S., Lenny, N., Lyu, C. J., van Deursen, J. M., Harada, H. and Downing, J. R. (1998). Expression of a knocked-in AML1-ETO leukemia gene inhibits the establishment of normal definitive hematopoiesis and directly generates dysplastic hematopoietic progenitors. *Blood* 91(9): 3134-3143.

Ong, K. K., Elks, C. E., Li, S., Zhao, J. H., Luan, J. a., Andersen, L. B., Bingham, S. A., Brage, S., Smith, G. D. and Ekelund, U. (2009). Genetic variation in LIN28B is associated with the timing of puberty. *Nature Genetics* 41(6): 729.

Ooki, A., Dinalankara, W., Marchionni, L., Tsay, J. J., Goparaju, C., Maleki, Z., Rom, W. N., Pass, H. I. and Hoque, M. O. (2018). Epigenetically regulated PAX6 drives cancer cells toward a stem-like state via GLI-SOX2 signaling axis in lung adenocarcinoma. *Oncogene* 37(45): 5967-5981.

OpenStax (2016). Anatomy & Physiology. *OpenStax CNX*.

Orlovsky, K., Kalinkovich, A., Rozovskaia, T., Shezen, E., Itkin, T., Alder, H., Ozer, H. G., Carramusa, L., Avigdor, A., Volinia, S., Buchberg, A., Mazo, A., Kollet, O., Largman, C., Croce, C. M., Nakamura, T., Lapidot, T. and Canaani, E. (2011). Down-regulation of homeobox genes MEIS1 and HOXA in MLL-rearranged acute leukemia impairs engraftment and reduces proliferation. *Proc Natl Acad Sci U S A* 108(19): 7956-7961.

Pallares, V., Hoyos, M., Chillon, M. C., Barragan, E., Prieto Conde, M. I., Llop, M., Cespedes, M. V., Nomdedeu, J. F., Brunet, S., Sanz, M. A., Gonzalez-Diaz, M., Sierra, J., Casanova, I. and Manges, R. (2017). NEDD9, an independent good prognostic factor in intermediate-risk acute myeloid leukemia patients. *Oncotarget*.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., Potter, N. E., Heuser, M., Thol, F., Bolli, N., Gundem, G., Van Loo, P., Martincorena, I., Ganly, P., Mudie, L., McLaren, S., O'Meara, S., Raine, K., Jones, D. R., Teague, J. W., Butler, A. P., Greaves, M. F., Ganser, A., Dohner, K., Schlenk, R. F., Dohner, H. and Campbell, P. J. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 374(23): 2209-2221.

Paprocka, M., Bielawska-Pohl, A., Rossowska, J., Krawczyński, A., Dus, D., Kielbinski, M., Haus, O., Podolak-Dawidziak, M. and Kuliczowski, K. (2017). MRP1 protein expression in leukemic stem cells as a negative prognostic marker in acute myeloid leukemia patients. *Eur J Haematol*.

Paschka, P., Du, J., Schlenk, R. F., Gaidzik, V. I., Bullinger, L., Corbacioglu, A., Späth, D., Kayser, S., Schlegelberger, B., Krauter, J., Ganser, A., Köhne, C.-H., Held, G., von Lilienfeld-Toal, M., Kirchen, H., Rummel, M., Götze, K., Horst, H.-A., Ringhoffer, M., Lübbert, M., Wattad, M., Salih, H. R., Kündgen, A., Döhner, H. and Döhner, K. (2013). Secondary genetic lesions in acute myeloid leukemia with inv(16) or t(16;16): a study of the German-Austrian AML Study Group (AMLSG). *Blood* 121(1): 170.

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. and Bejerano, G. (2013). Enhancers: five essential questions. *Nature Reviews Genetics* 14: 288.

Perea, G., Domingo, A., Villamor, N., Palacios, C., Junca, J., Torres, P., Llorente, A., Fernandez, C., Tormo, M., Queipo de Llano, M. P., Bargay, J., Gallart, M., Florensa, L., Vivancos, P., Marti, J. M., Font, L., Berlanga, J., Esteve, J., Bueno, J., Ribera, J. M., Brunet, S., Sierra, J. and Nomdedeu, J. F. (2005). Adverse prognostic impact of CD36 and CD2 expression in adult de novo acute myeloid leukemia patients. *Leuk Res* 29(10): 1109-1116.

Pession, A., Masetti, R., Rizzari, C., Putti, M. C., Casale, F., Fagioli, F., Luciani, M., Nigro, L. L., Menna, G. and Micalizzi, C. (2013). Results of the AIEOP AML 2002/01 multicenter prospective trial for the treatment of children with acute myeloid leukemia. *Blood* 122(2): 170-178.

Pfeifer, G. P., Steigerwald, S., Boehm, T. L. and Drahovsky, D. (1988). DNA methylation levels in acute human leukemia. *Cancer Lett* 39(2): 185-192.

Pinto-Medel, M. J., Oliver-Martos, B., Urbaneja-Romero, P., Hurtado-Guerrero, I., Ortega-Pinazo, J., Serrano-Castro, P., Fernández, Ó. and Leyva, L. (2017). Global methylation correlates with clinical status in multiple sclerosis patients in the first year of IFNbeta treatment. *Sci Rep* 7: 8727.

Plon, S. E. and Nathanson, K. (2005). Inherited Susceptibility for Pediatric Cancer. *Cancer Journal* 11(4): 255.

Portela, A. and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat Biotechnol* 28(10): 1057-1068.

Pui, C. H., Carroll, W. L., Meshinchi, S. and Arceci, R. J. (2011). Biology, risk stratification, and therapy of pediatric acute leukemias: an update. *J Clin Oncol* 29(5): 551-565.

Puumala, S. E., Ross, J. A., Aplenc, R. and Spector, L. G. (2013). Epidemiology of Childhood Acute Myeloid Leukemia. *Pediatric blood & cancer* 60(5): 728-733.

Qu, Y., Lennartsson, A., Gaidzik, V. I., Deneberg, S., Karimi, M., Bengtzen, S., Höglund, M., Bullinger, L., Döhner, K. and Lehmann, S. (2014). Differential methylation in CN-AML

preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. *Epigenetics* 9(8): 1108-1119.

Qu, Y., Siggins, L., Cordeddu, L., Gaidzik, V. I., Karlsson, K., Bullinger, L., Dohner, K., Ekwall, K., Lehmann, S. and Lennartsson, A. (2016). Cancer specific changes in DNA methylation reveal aberrant silencing and activation of enhancers in leukemia. *Blood*.

Queensland, C. C. (2014). A summary of childhood cancer in Australia 1983-2014. *Cancer Council Queensland*.

Radtke, I., Mullighan, C. G., Ishii, M., Su, X., Cheng, J., Ma, J., Ganti, R., Cai, Z., Goorha, S., Pounds, S. B., Cao, X., Obert, C., Armstrong, J., Zhang, J., Song, G., Ribeiro, R. C., Rubnitz, J. E., Raimondi, S. C., Shurtleff, S. A. and Downing, J. R. (2009). Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. *Proceedings of the National Academy of Sciences* 106(31): 12944.

Raffel, S., Falcone, M., Kneisel, N., Hansson, J., Wang, W., Lutz, C., Bullinger, L., Poschet, G., Nonnenmacher, Y., Barnert, A., Bahr, C., Zeisberger, P., Przybylla, A., Sohn, M., Tonjes, M., Erez, A., Adler, L., Jensen, P., Scholl, C., Frohling, S., Cocciardi, S., Wuchter, P., Thiede, C., Florcken, A., Westermann, J., Ehninger, G., Lichter, P., Hiller, K., Hell, R., Herrmann, C., Ho, A. D., Krijgsveld, J., Radlwimmer, B. and Trumpp, A. (2017). BCAT1 restricts alphaKG levels in AML stem cells leading to IDHmut-like DNA hypermethylation. *Nature* 551(7680): 384-388.

Ramos, K. N., Ramos, I. N., Zeng, Y. and Ramos, K. S. (2018). Genetics and epigenetics of pediatric leukemia in the era of precision medicine. *F1000Research* 7: F1000 Faculty Rev-1104.

Rampal, R. and Figueroa, M. E. (2016). Wilms tumor 1 mutations in the pathogenesis of acute myeloid leukemia. *Haematologica* 101(6): 672-679.

Randhawa, J. K., Kantarjian, H. M., Borthakur, G., Thompson, P. A., Konopleva, M., Daver, N., Pemmaraju, N., Jabbour, E., Kadia, T. M., Estrov, Z., Ramachandran, A., Paradela, J., Andreeff, M., Levis, M., Ravandi, F. and Cortes, J. E. (2014). Results of a Phase II Study of Crenolanib in Relapsed/Refractory Acute Myeloid Leukemia Patients (Pts) with Activating FLT3 Mutations. *Blood* 124(21): 389-389.

Rasche, M., Zimmermann, M., Borschel, L., Bourquin, J.-P., Dworzak, M., Klingebiel, T., Lehnbecher, T., Creutzig, U., Klusmann, J.-H. and Reinhardt, D. (2018). Successes and challenges in the treatment of pediatric acute myeloid leukemia: a retrospective analysis of the AML-BFM trials from 1987 to 2012. *Leukemia*.

Ravandi, F., Jorgensen, J., Borthakur, G., Jabbour, E., Kadia, T., Pierce, S., Brandt, M., Wang, S., Konoplev, S., Wang, X., Huang, X., Daver, N., DiNardo, C., Andreeff, M., Konopleva,

M., Estrov, Z., Garcia-Manero, G., Cortes, J. and Kantarjian, H. (2016). Persistence of minimal residual disease assessed by multiparameter flow cytometry is highly prognostic in younger patients with acute myeloid leukemia. *Cancer*.

Ravindranath, Y., Chang, M., Steuber, C. P., Becton, D., Dahl, G., Civin, C., Camitta, B., Carroll, A., Raimondi, S. C. and Weinstein, H. J. (2005). Pediatric Oncology Group (POG) studies of acute myeloid leukemia (AML): a review of four consecutive childhood AML trials conducted between 1981 and 2000. *Leukemia* 19(12): 2101-2116.

Reyngold, M., Turcan, S., Giri, D., Kannan, K., Walsh, L. A., Viale, A., Drobnjak, M., Vahdat, L. T., Lee, W. and Chan, T. A. (2014). Remodeling of the methylation landscape in breast cancer metastasis. *PLoS One* 9(8): e103896.

Rhoades, K. L., Hetherington, C. J., Harakawa, N., Yergeau, D. A., Zhou, L., Liu, L. Q., Little, M. T., Tenen, D. G. and Zhang, D. E. (2000). Analysis of the role of AML1-ETO in leukemogenesis, using an inducible transgenic mouse model. *Blood* 96(6): 2108-2115.

Richards, K. L., Zhang, B., Baggerly, K. A., Colella, S., Lang, J. C., Schuller, D. E. and Krahe, R. (2009). Genome-wide hypomethylation in head and neck cancer is more pronounced in HPV-negative tumors and is associated with genomic instability. *PLoS One* 4(3): e4941.

Robertson, K. D. (2005). DNA methylation and human disease. *Nature Reviews Genetics* 6: 597.

Rodriguez, J., Frigola, J., Vendrell, E., Risques, R.-A., Fraga, M. F., Morales, C., Moreno, V., Esteller, M., Capella, G. and Ribas, M. (2006). Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. *Cancer research* 66(17): 8462-9468.

Rowe, J. M., Li, X., Cassileth, P. A., Appelbaum, F. R., Schiffer, C. A., Wiernik, P. H., Litzow, M. R., Cripe, L. D., Lazarus, H. M. and Paietta, E. (2005). Very Poor Survival of Patients with AML Who Relapse after Achieving a First Complete Remission: The Eastern Cooperative Oncology Group Experience, Am Soc Hematology.

Rubnitz, J. E., Inaba, H., Ribeiro, R. C., Pounds, S., Rooney, B., Bell, T., Pui, C.-H. and Leung, W. (2010). NKAML: a pilot study to determine the safety and feasibility of haploidentical natural killer cell transplantation in childhood acute myeloid leukemia. *Journal of Clinical Oncology* 28(6): 955.

Rubnitz, J. E., Razzouk, B. I., Lensing, S., Pounds, S., Pui, C. H. and Ribeiro, R. C. (2007). Prognostic factors and outcome of recurrence in childhood acute myeloid leukemia. *Cancer* 109(1): 157-163.

Sandanger, T. M., Nøst, T. H., Guida, F., Rylander, C., Campanella, G., Muller, D. C., van Dongen, J., Boomsma, D. I., Johansson, M., Vineis, P., Vermeulen, R., Lund, E. and Chadeau-Hyam, M. (2018). DNA methylation and associated gene expression in blood prior to lung cancer diagnosis in the Norwegian Women and Cancer cohort. *Sci Rep* 8(1): 16714.

Sander, A., Zimmermann, M., Dworzak, M., Fleischhack, G., von Neuhoff, C., Reinhardt, D., Kaspers, G. J. and Creutzig, U. (2010). Consequent and intensified relapse therapy improved survival in pediatric AML: results of relapse treatment in 379 patients of three consecutive AML-BFM trials. *Leukemia* 24(8): 1422-1428.

Sandoval, J. and Esteller, M. (2012). Cancer epigenomics: beyond genomics. *Curr Opin Genet Dev* 22(1): 50-55.

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M. A., Bibikova, M. and Esteller, M. (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6(6): 692-702.

Schnittger, S., Bacher, U., Haferlach, C., Kern, W. and Haferlach, T. (2007). Rare CBFβ-MYH11 fusion transcripts in AML with inv(16)/t(16;16) are associated with therapy-related AML M4eo, atypical cytomorphology, atypical immunophenotype, atypical additional chromosomal rearrangements and low white blood cell count: a study on 162 patients. *Leukemia* 21(4): 725-731.

Schoofs, T., Berdel, W. E. and Muller-Tidow, C. (2014). Origins of aberrant DNA methylation in acute myeloid leukemia. *Leukemia* 28(1): 1-14.

Schwalbe, E. C., Williamson, D., Lindsey, J. C., Hamilton, D., Ryan, S. L., Megahed, H., Garami, M., Hauser, P., Dembowska-Baginska, B., Perek, D., Northcott, P. A., Taylor, M. D., Taylor, R. E., Ellison, D. W., Bailey, S. and Clifford, S. C. (2013). DNA methylation profiling of medulloblastoma allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. *Acta Neuropathol* 125(3): 359-371.

Sengelaub, C. A., Navrazhina, K., Ross, J. B., Halberg, N. and Tavazoie, S. F. (2016). PTPRN2 and PLCβ1 promote metastatic breast cancer cell migration through PI(4,5)P2-dependent actin remodeling. *The EMBO journal* 35(1): 62-76.

Shao, X., Liu, Y., Li, Y., Xian, M., Zhou, Q., Yang, B., Ying, M. and He, Q. (2016). The HER2 inhibitor TAK165 Sensitizes Human Acute Myeloid Leukemia Cells to Retinoic Acid-Induced Myeloid Differentiation by activating MEK/ERK mediated RARα/STAT1 axis. *Sci Rep* 6: 24589.

Sharif, J., Muto, M., Takebayashi, S.-i., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T. and Okamura, K. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450(7171): 908.

Sharma, S., Kelly, T. K. and Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis* 31(1): 27-36.

Shih, A. H., Abdel-Wahab, O., Patel, J. P. and Levine, R. L. (2012). The role of mutations in epigenetic regulators in myeloid malignancies. *Nat Rev Cancer* 12(9): 599-612.

Shih, Y. L., Kuo, C. C., Yan, M. D., Lin, Y. W., Hsieh, C. B. and Hsieh, T. Y. (2016). Quantitative methylation analysis reveals distinct association between PAX6 methylation and clinical characteristics with different viral infections in hepatocellular carcinoma. *Clin Epigenetics* 8: 41.

Shlush, L. I., Mitchell, A., Heisler, L., Abelson, S., Ng, S. W. K., Trotman-Grant, A., Medeiros, J. J. F., Rao-Bhatia, A., Jaciw-Zurakowsky, I., Marke, R., McLeod, J. L., Doedens, M., Bader, G., Voisin, V., Xu, C., McPherson, J. D., Hudson, T. J., Wang, J. C. Y., Minden, M. D. and Dick, J. E. (2017). Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature*.

Sinclair, S. H. G., Yegnasubramanian, S. and Dumler, J. S. (2015). Global DNA methylation changes and differential gene expression in *Anaplasma phagocytophilum*-infected human neutrophils. *Clinical Epigenetics* 7(1): 77.

Sinha, S. T., D.; Yu, L.; Gentles, A. J.; Jung, N.; Corces-Zimmerman, M. R.; Chan, S. M.; Reinisch, A.; Feinberg, A. P.; Dill, D. L.; Majeti, R. (2014). Mutant WT1 is associated with DNA hypermethylation of PRC2 targets in AML and responds to EZH2 inhibition. *The American Society of Hematology* 125(2): 316-326.

Smith, F. O., Alonzo, T. A., Gerbing, R. B., Woods, W. G. and Arceci, R. J. (2005). Long-term results of children with acute myeloid leukemia: a report of three consecutive Phase III trials by the Children's Cancer Group: CCG 251, CCG 213 and CCG 2891. *Leukemia* (08876924) 19(12): 2054-2062.

Sobieszkoda, D., Czech, J., Gablo, N., Kopanska, M., Tabarkiewicz, J., Kolacinska, A., Robak, T. and Zawlik, I. (2017). MGMT promoter methylation as a potential prognostic marker for acute leukemia. *Arch Med Sci* 13(6): 1433-1441.

Stahnke, K., Boos, J., Bender-Gotze, C., Ritter, J., Zimmermann, M. and Creutzig, U. (1998). Duration of first remission predicts remission rates and long-term survival in children with relapsed acute myelogenous leukemia. *Leukemia* 12(10): 1534-1538.

Stein, E. M. and Tallman, M. S. (2015). Emerging therapeutic drugs for AML. *Blood*.

Stengel, A., Kern, W., Meggendorfer, M., Nadarajah, N., Perglerova, K., Haferlach, T. and Haferlach, C. (2017). Number of RUNX1 mutations, wild-type allele loss and additional mutations impact on prognosis in adult RUNX1 mutated AML. *Leukemia*.

Stieglitz, E., Mazor, T., Olshen, A. B., Geng, H., Gelston, L. C., Akutagawa, J., Lipka, D. B., Plass, C., Flotho, C., Chehab, F. F., Braun, B. S., Costello, J. F. and Loh, M. L. (2017). Genome-wide DNA methylation is predictive of outcome in juvenile myelomonocytic leukemia. *Nat Commun* 8(1): 2127.

Stone, R. M., Mandrekar, S., Sanford, B. L., Geyer, S., Bloomfield, C. D., Dohner, K., Thiede, C., Marcucci, G., Lo-Coco, F., Klisovic, R. B., Wei, A., Sierra, J., Sanz, M. A., Brandwein, J. M., de Witte, T., Niederwieser, D., Appelbaum, F. R., Medeiros, B. C., Tallman, M. S., Krauter, J., Schlenk, R. F., Ganser, A., Serve, H., Ehninger, G., Amadori, S., Larson, R. A. and Dohner, H. (2015). The Multi-Kinase Inhibitor Midostaurin (M) Prolongs Survival Compared with Placebo (P) in Combination with Daunorubicin (D)/Cytarabine (C) Induction (ind), High-Dose C Consolidation (consol), and As Maintenance (maint) Therapy in Newly Diagnosed Acute Myeloid Leukemia (AML) Patients (pts) Age 18-60 with *FLT3* Mutations (muts): An International Prospective Randomized (rand) P-Controlled Double-Blind Trial (CALGB 10603/RATIFY [Alliance]). *Blood* 126(23): 6-6.

Stumpel, D. J. P. M., Schneider, P., van Roon, E. H. J., Pieters, R. and Stam, R. W. (2013). Absence of global hypomethylation in promoter hypermethylated Mixed Lineage Leukaemia-rearranged infant acute lymphoblastic leukaemia. *European Journal of Cancer* 49(1): 175-184.

Sturm, D., Orr, B. A., Toprak, U. H., Hovestadt, V., Jones, D. T. W., Capper, D., Sill, M., Buchhalter, I., Northcott, P. A., Leis, I., Ryzhova, M., Koelsche, C., Pfaff, E., Allen, S. J., Balasubramanian, G., Worst, B. C., Pajtler, K. W., Brabetz, S., Johann, P. D., Sahm, F., Reimand, J., Mackay, A., Carvalho, D. M., Remke, M., Phillips, J. J., Perry, A., Cowdrey, C., Drissi, R., Fouladi, M., Giangaspero, F., Łastowska, M., Grajkowska, W., Scheurlen, W., Pietsch, T., Hagel, C., Gojo, J., Lötsch, D., Berger, W., Slavc, I., Haberler, C., Jouvett, A., Holm, S., Hofer, S., Prinz, M., Keohane, C., Fried, I., Mawrin, C., Scheie, D., Mobley, B. C., Schniederjan, M. J., Santi, M., Buccoliero, A. M., Dahiya, S., Kramm, C. M., von Bueren, A. O., von Hoff, K., Rutkowski, S., Herold-Mende, C., Frühwald, M. C., Milde, T., Hasselblatt, M., Wesseling, P., Rößler, J., Schüller, U., Ebinger, M., Schittenhelm, J., Frank, S., Grobholz, R., Vajtai, I., Hans, V., Schneppenheim, R., Zitterbart, K., Collins, V. P., Aronica, E., Varlet, P., Puget, S., Dufour, C., Grill, J., Figarella-Branger, D., Wolter, M., Schuhmann, M. U., Shalaby, T., Grotzer, M., van Meter, T., Monoranu, C.-M., Felsberg, J., Reifenberger, G., Snuderl, M., Forrester, L. A., Koster, J., Versteeg, R., Volckmann, R., van Sluis, P., Wolf, S., Mikkelsen, T., Gajjar, A., Aldape, K., Moore, A. S., Taylor, M. D., Jones, C., Jabado, N., Karajannis, M. A., Eils, R., Schlesner, M., Lichter, P., von Deimling, A., Pfister, S. M., Ellison, D. W., Korshunov, A. and Kool, M. (2016). New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. *Cell* 164(5): 1060-1072.

Summers, K., Stevens, J., Kakkas, I., Smith, M., Smith, L. L., MacDougall, F., Cavenagh, J., Bonnet, D., Young, B. D., Lister, T. A. and Fitzgibbon, J. (2007). Wilms' tumour 1 mutations are associated with *FLT3*-ITD and failure of standard induction chemotherapy in patients with normal karyotype AML. *Leukemia* 21(3): 550-551.

Tammen, S. A., Friso, S. and Choi, S. W. (2013). Epigenetics: the link between nature and nurture. *Mol Aspects Med* 34(4): 753-764.

Tan, Y., Liu, H. and Chen, S. (2015). Mutant DNA methylation regulators endow hematopoietic stem cells with the preleukemic stem cell property, a requisite of leukemia initiation and relapse. *Front Med*.

Tang, X., Chen, L., Yan, X., Li, Y., Xiong, Y. and Zhou, X. (2015). Overexpression of miR-210 is Associated with Poor Prognosis of Acute Myeloid Leukemia. *Med Sci Monit* 21: 3427-3433.

Tao, Y. F., Fang, F., Hu, S. Y., Lu, J., Cao, L., Zhao, W. L., Xiao, P. F., Li, Z. H., Wang, N. N., Xu, L. X., Du, X. J., Sun, L. C., Li, Y. H., Li, Y. P., Xu, Y. Y., Ni, J., Wang, J., Feng, X. and Pan, J. (2015). Hypermethylation of the GATA binding protein 4 (GATA4) promoter in Chinese pediatric acute myeloid leukemia. *BMC Cancer* 15(1): 756.

Tao, Y. F., Hu, S. Y., Lu, J., Cao, L., Zhao, W. L., Xiao, P. F., Xu, L. X., Li, Z. H., Wang, N. N., Du, X. J., Sun, L. C., Zhao, H., Fang, F., Su, G. H., Li, Y. H., Li, Y. P., Xu, Y. Y., Ni, J., Wang, J., Feng, X. and Pan, J. (2014). Zinc finger protein 382 is downregulated by promoter hypermethylation in pediatric acute myeloid leukemia patients. *Int J Mol Med* 34(6): 1505-1515.

Tao, Y. F., Xu, L. X., Lu, J., Hu, S. Y., Fang, F., Cao, L., Xiao, P. F., Du, X. J., Sun, L. C., Li, Z. H., Wang, N. N., Su, G. H., Li, Y. H., Li, G., Zhao, H., Li, Y. P., Xu, Y. Y., Zhou, H. T., Wu, Y., Jin, M. F., Liu, L., Zhu, X. M., Ni, J., Wang, J., Xing, F., Zhao, W. L. and Pan, J. (2015). Early B-cell factor 3 (EBF3) is a novel tumor suppressor gene with promoter hypermethylation in pediatric acute myeloid leukemia. *J Exp Clin Cancer Res* 34: 4.

Tarlock, K., He, Y., Sun, J., Ries, R. E., Severson, E. A., Bailey, M., Morley, S., Balasubramanian, S., Erlich, R., Lipson, D., Otto, G., Vergillo, J.-A., Ross, J. S., Mughal, T., Stephens, P., Miller, V., Meshinchi, S. and He, J. (2017). Recurrent Copy Number Variants Are Highly Prevalent in Acute Myeloid Leukemia. *Blood* 130(Suppl 1): 3800.

Tarlock, K. and Meshinchi, S. (2015). Pediatric acute myeloid leukemia: biology and therapeutic implications of genomic variants. *Pediatr Clin North Am* 62(1): 75-93.

Taskesen, E., Babaei, S., Reinders, M. M. J. and de Ridder, J. (2015). Integration of gene expression and DNA-methylation profiles improves molecular subtype classification in acute myeloid leukemia. *BMC Bioinformatics* 16(Suppl 4): S5-S5.

Teng, E. C., Racioppi, L. and Means, A. R. (2011). A cell-intrinsic role for CaMKK2 in granulocyte lineage commitment and differentiation. *Journal of leukocyte biology* 90(5): 897-909.

Teschendorff, A. E., Zhuang, J. and Widschwendter, M. (2011). Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics* 27(11): 1496-1505.

The Cancer Genome Atlas Research, N. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368(22): 2059-2074.

Thomas, D. and Majeti, R. (2017). Biology and relevance of human acute myeloid leukemia stem cells. *Blood*.

Thorsteinsdottir, U., Kroon, E., Jerome, L., Blasi, F. and Sauvageau, G. (2001). Defining Roles for HOX and MEIS1 Genes in Induction of Acute Myeloid Leukemia. *Mol Cell Biol* 21(1): 224.

Thursfield, V., Staines, C., Giles, G. and Farrugia, H. (2013). Cancer in Victoria: Statistics and Trends 2012. *Cancer Council Victoria*.

Tomlinson, I., Sasieni, P. and Bodmer, W. (2002). How many mutations in a cancer? *The American Journal Of Pathology* 160(3): 755-758.

Toyota, M., Kopecky, K.J., Toyota, M.-O., Jair, K.-W., Willman, C. L. and Issa, J.-P.J. (2001). Methylation profiling in acute myeloid leukemia. *Blood* 97(9): 2823-2829.

Trka, J., Kalinova, M., Hrušák, O., Zuna, J., Krejčí, O., Madžo, J., Sedláček, P., Vavra, V., Michalova, K. and Jarošová, M. (2002). Real-time quantitative PCR detection of WT1 gene expression in children with AML: prognostic significance, correlation with disease status and residual disease detection by flow cytometry. *Leukemia* 16(7): 1381.

Tsukimoto, I., Tawa, A., Horibe, K., Tabuchi, K., Kigasawa, H., Tsuchida, M., Yabe, H., Nakayama, H., Kudo, K. and Kobayashi, R. (2009). Risk-stratified therapy and the intensive use of cytarabine improves the outcome in childhood acute myeloid leukemia: the AML99 trial from the Japanese Childhood AML Cooperative Study Group. *Journal of Clinical Oncology* 27(24): 4007-4013.

Vardiman, J. W., Harris, N. L. and Brunning, R. D. (2002). The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* 100(7): 2292-2302.

Vardiman, J. W., Thiele, J., Arber, D. A., Brunning, R. D., Borowitz, M. J., Porwit, A., Harris, N. L., Le Beau, M. M., Hellstrom-Lindberg, E., Tefferi, A. and Bloomfield, C. D. (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 114(5): 937-951.

Viale, A., De Franco, F., Orleth, A., Cambiaghi, V., Giuliani, V., Bossi, D., Ronchini, C., Ronzoni, S., Muradore, I., Monestiroli, S., Gobbi, A., Alcalay, M., Minucci, S. and Pelicci, P. G. (2009). Cell-cycle restriction limits DNA damage and maintains self-renewal of leukaemia stem cells. *Nature* 457(7225): 51-56.

Vidal, E., Sayols, S., Moran, S., Guillaumet-Adkins, A., Schroeder, M. P., Royo, R., Orozco, M., Gut, M., Gut, I., Lopez-Bigas, N., Heyn, H. and Esteller, M. (2017). A DNA methylation map of human cancer at single base-pair resolution. *Oncogene* 36: 5648.

Vryer, R. and Saffery, R. (2017). What's in a name? Context-dependent significance of 'global' methylation measures in human health and disease. *Clinical Epigenetics* 9: 2.

Walker, B. A., Wardell, C. P., Chiecchio, L., Smith, E. M., Boyd, K. D., Neri, A., Davies, F. E., Ross, F. M. and Morgan, G. J. (2011). Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood* 117(2): 553.

Wang, G. G. and Kamps, M. P. (2007). Survival signaling in HoxA9/Meis1 AML. *Blood* 109(9): 3619.

Wang, X., Chen, H., Bai, J. and He, A. (2017). MicroRNA: an important regulator in acute myeloid leukemia. *Cell Biol Int* 41(9): 936-945.

Wang, Y. X., Zhang, T. J., Yang, D. Q., Yao, D. M., Yang, L., Zhou, J. D., Deng, Z. Q., Ma, J. C., Guo, H., Wen, X. M., Lin, J. and Qian, J. (2016). Reduced miR-215 expression predicts poor prognosis in patients with acute myeloid leukemia. *Jpn J Clin Oncol*.

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J. M. and Network, C. G. A. R. (2013). The cancer genome atlas pan-cancer analysis project. *Nature Genetics* 45(10): 1113.

Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., Wartman, L. D., Lamprecht, T. L., Liu, F., Xia, J., Kandoth, C., Fulton, R. S., McLellan, M. D., Dooling, D. J., Wallis, J. W., Chen, K., Harris, C. C., Schmidt, H. K., Kalicki-Veizer, J. M., Lu, C., Zhang, Q., Lin, L., O'Laughlin, M. D., McMichael, J. F., Delehaunty, K. D., Fulton, L. A., Magrini, V. J., McGrath, S. D., Demeter, R. T., Vickery, T. L., Hundal, J., Cook, L. L., Swift, G. W., Reed, J. P., Alldredge, P. A., Wylie, T. N., Walker, J. R., Watson, M. A., Heath, S. E., Shannon, W. D., Varghese, N., Nagarajan, R., Payton, J. E., Baty, J. D., Kulkarni, S., Klco, J. M., Tomasson, M. H., Westervelt, P., Walter, M. J., Graubert, T. A., DiPersio, J. F., Ding, L., Mardis, E. R. and Wilson, R. K. (2012). The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150(2): 264-278.

Wertheim, G. B., Smith, C., Lusk, M., Rager, A., Figueroa, M. E., Carroll, M. and Master, S. R. (2015). Validation of DNA methylation to predict outcome in acute myeloid leukemia by use of xMELP. *Clin Chem* 61(1): 249-258.

WHO (2014). Acute Myelogenous Leukemia and Acute Promyelocytic Leukemia.

Woods, W. G., Neudorf, S., Gold, S., Sanders, J., Buckley, J. D., Barnard, D. R., Dusenbery, K., DeSwarte, J., Arthur, D. C., Lange, B. J. and Kobrinsky, N. L. (2001). A comparison of allogeneic bone marrow transplantation, autologous bone marrow transplantation, and aggressive chemotherapy in children with acute myeloid leukemia in remission: a report from the Children's Cancer Group. *Blood* 97(1): 56.

Wouters, B. J., Jorda, M. A., Keeshan, K., Louwers, I., Erpelinck-Verschueren, C. A., Tielemans, D., Langerak, A. W., He, Y., Yashiro-Ohtani, Y., Zhang, P., Hetherington, C. J., Verhaak, R. G., Valk, P. J., Lowenberg, B., Tenen, D. G., Pear, W. S. and Delwel, R. (2007). Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood* 110(10): 3706-3714.

Wu, D. H., Yao, D. M., Yang, L., Ma, J. C., Wen, X. M., Yang, J., Guo, H., Li, X. X., Qian, W., Lin, J. and Qian, J. (2017). Hypomethylation of let-7a-3 is associated with poor prognosis in myelodysplastic syndrome. *Leuk Lymphoma* 58(1): 96-103.

Wu, P., Farrell, W. E., Haworth, K. E., Emes, R. D., Kitchen, M. O., Glossop, J. R., Hanna, F. W. and Fryer, A. A. (2018). Maternal genome-wide DNA methylation profiling in gestational diabetes shows distinctive disease-associated changes relative to matched healthy pregnancies. *Epigenetics* 13(2): 122-128.

Wu, X., Rauch, T. A., Zhong, X., Bennett, W. P., Latif, F., Krex, D. and Pfeifer, G. P. (2010). CpG island hypermethylation in human astrocytomas. *Cancer Res* 70(7): 2718-2727.

Xu, H., Beasley, M. D., Warren, W. D., van der Horst, G. T. and McKay, M. J. (2005). Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis. *Developmental cell* 8(6): 949-961.

Yagi, T., Morimoto, A., Eguchi, M., Hibi, S., Sako, M., Ishii, E., Mizutani, S., Imashuku, S., Ohki, M. and Ichikawa, H. (2003). Identification of a gene expression signature associated with pediatric AML prognosis. *Blood* 102(5): 1849-1856.

Yan-Fang, T., Jian, N., Jun, L., Na, W., Pei-Fang, X., Wen-Li, Z., Dong, W., Li, P., Jian, W. and Xing, F. (2013). The promoter of miR-663 is hypermethylated in Chinese pediatric acute myeloid leukemia (AML). *BMC medical genetics* 14(1): 74.

Yanada, M. (2015). Allogeneic hematopoietic cell transplantation for acute myeloid leukemia during first complete remission: a clinical perspective. *Int J Hematol* 101(3): 243-254.

Yang, H., Ye, D., Guan, K.-L. and Xiong, Y. (2012). IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives. *Clin Cancer Res* 18(20): 5562-5571.

Yang, L., Han, Y., Saiz, F. S. and Minden, M. (2007). A tumor suppressor and oncogene: the WT1 story. *Leukemia* 21(5): 868.

Yin, J. A. L., O'Brien, M. A., Hills, R. K., Daly, S. B., Wheatley, K. and Burnett, A. K. (2012). Minimal residual disease monitoring by quantitative RT-PCR in core binding factor AML allows risk stratification and predicts relapse: results of the United Kingdom MRC AML-15 trial. *Blood* 120(14): 2826-2835.

Yu, J., Liang, Q., Wang, J., Wang, K., Gao, J., Zhang, J., Zeng, Y., Chiu, P. W. Y., Ng, E. K. W. and Sung, J. J. Y. (2017). REC8 functions as a tumor suppressor and is epigenetically downregulated in gastric cancer, especially in EBV-positive subtype. *Oncogene* 36(2): 182-193.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V. and Stewart, R. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318(5858): 1917-1920.

Zampini, M., Tregnago, C., Bisio, V., Simula, L., Borella, G., Manara, E., Zanon, C., Zonta, F., Serafin, V., Accordi, B., Campello, S., Buldini, B., Pession, A., Locatelli, F., Basso, G. and Pigazzi, M. (2018). Epigenetic heterogeneity affects the risk of relapse in children with t(8;21)RUNX1-RUNX1T1-rearranged AML. *Leukemia*.

Zangenberg, M., Grubach, L., Aggerholm, A., Silkjaer, T., Juhl-Christensen, C., Nyvold, C. G., Kjeldsen, E., Ommen, H. B. and Hokland, P. (2009). The combined expression of HOXA4 and MEIS1 is an independent prognostic factor in patients with AML. *Eur J Haematol* 83(5): 439-448.

Zeng, Y., Yao, B., Shin, J., Lin, L., Kim, N., Song, Q., Liu, S., Su, Y., Guo, J. U., Huang, L., Wan, J., Wu, H., Qian, J., Cheng, X., Zhu, H., Ming, G. L., Jin, P. and Song, H. (2016). Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. *Mol Cell* 61(1): 153-160.

Zhang, L. Y., Yuan, Y. Q., Zhou, D. M., Wang, Z. Y., Ju, S. G., Sun, Y., Li, J. and Fu, J. X. (2016). Impact of Global and Gene-Specific DNA Methylation in de Novo or Relapsed Acute Myeloid Leukemia Patients Treated with Decitabine. *Asian Pac J Cancer Prev* 17(1): 431-437.

Zhang, T. J., Lin, J., Zhou, J. D., Li, X. X., Zhang, W., Guo, H., Xu, Z. J., Yan, Y., Ma, J. C. and Qian, J. (2017). High bone marrow miR-19b level predicts poor prognosis and disease recurrence in de novo acute myeloid leukemia. *Gene*.

Zhang, T. J., Wu, D. H., Zhou, J. D., Li, X. X., Zhang, W., Guo, H., Ma, J. C., Deng, Z. Q., Lin, J. and Qian, J. (2017). Overexpression of miR-216b: prognostic and predictive value in acute myeloid leukemia. *J Cell Physiol*.

Zhang, X., Yang, X., Wang, J., Liang, T., Gu, Y. and Yang, D. (2015). Down-regulation of PAX6 by promoter methylation is associated with poor prognosis in non small cell lung cancer. *Int J Clin Exp Pathol* 8(9): 11452-11457.

Zhang, Z. H., Zhang, W., Zhou, J. D., Zhang, T. J., Ma, J. C., Xu, Z. J., Lian, X. Y., Wu, D. H., Wen, X. M., Deng, Z. Q., Lin, J. and Qian, J. (2018). Decreased SCIN expression, associated with promoter methylation, is a valuable predictor for prognosis in acute myeloid leukemia. *Mol Carcinog*.

Zheng, X., Zhan, Z., Naren, D., Li, J., Yan, T. and Gong, Y. (2017). Prognostic value of SRSF2 mutations in patients with de novo myelodysplastic syndromes: A meta-analysis. *PLoS One* 12(9): e0185053.

Zhou, J. D., Lin, J., Zhang, T. J., Ma, J. C., Yang, L., Wen, X. M., Guo, H., Yang, J., Deng, Z. Q. and Qian, J. (2016). GPX3 methylation in bone marrow predicts adverse prognosis and leukemia transformation in myelodysplastic syndrome. *Cancer Med*.

Zhou, J. D., Zhang, T. J., Li, X. X., Ma, J. C., Guo, H., Wen, X. M., Zhang, W., Yang, L., Yan, Y., Lin, J. and Qian, J. (2017). Epigenetic dysregulation of ID4 predicts disease progression and treatment outcome in myeloid malignancies. *J Cell Mol Med*.

Zhou, J. D., Zhang, T. J., Wang, Y. X., Yang, D. Q., Yang, L., Ma, J. C., Wen, X. M., Yang, J., Lin, J. and Qian, J. (2016). DLX4 hypermethylation is a prognostically adverse indicator in de novo acute myeloid leukemia. *Tumour Biol*.

Zhou, L. Y., Yin, J. Y., Tang, Q., Zhai, L. L., Zhang, T. J., Wang, Y. X., Yang, D. Q., Qian, J., Lin, J. and Deng, Z. Q. (2015). High expression of dual-specificity phosphatase 5 pseudogene 1 (DUSP5P1) is associated with poor prognosis in acute myeloid leukemia. *Int J Clin Exp Pathol* 8(12): 16073-16080.

Zhou, L. Y., Zhai, L. L., Yin, J. Y., Vanessa, M. E., Zhou, J., Zhang, J., Tang, X., Lin, J., Qian, J. and Deng, Z. Q. (2016). Pseudogene BMI1P1 expression as a novel predictor for acute myeloid leukemia development and prognosis. *Oncotarget*.

Zhu, X. W., Yao, D. M., Wu, D. H., Wen, X. M., Yang, J., Guo, H., Yang, L., Deng, Z. Q., Zhang, Y. Y., Qian, W., Lin, J. and Qian, J. (2016). Let-7a-3 hypomethylation is associated with favorable/intermediate karyotypes but not with survival in acute myeloid leukemia. *Tumour Biol* 37(1): 491-501.

Appendices

Appendix I - Primer sequences and amplicon lengths

Appendix II - Principal components analysis heatmaps

Appendix III - Prognostic testing of probes from select genes with potential functional relevance to progression to relapse - Discovery analysis

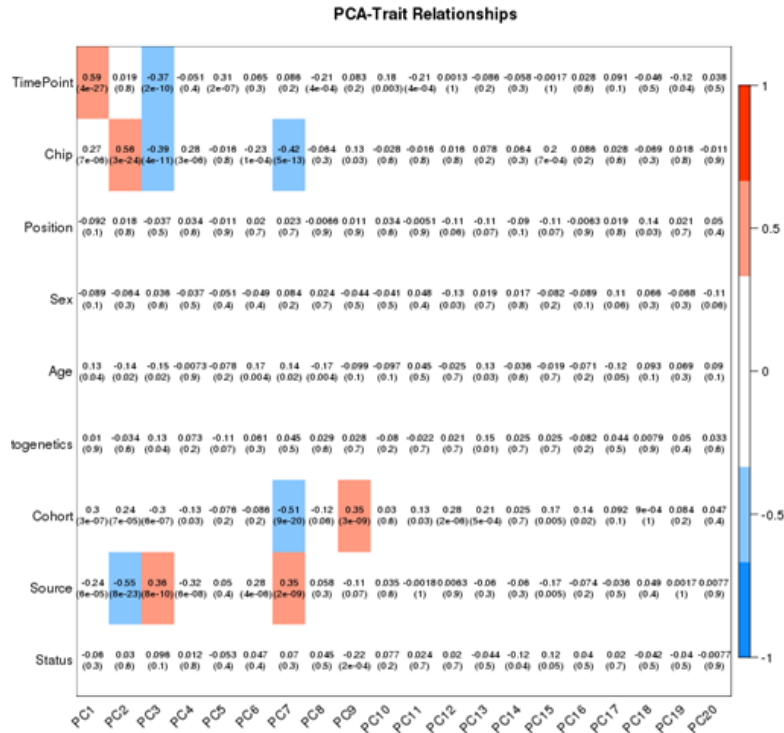
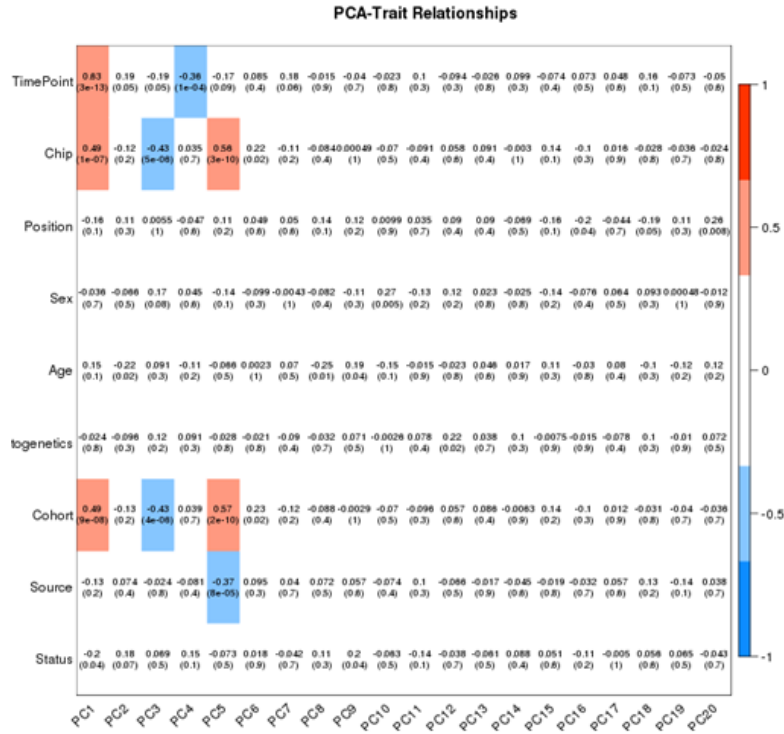
Appendix IV - Prognostic testing of probes from select genes with potential functional relevance to progression to relapse - Pooled-analysis

Appendix I - Primer sequences and amplicon lengths

Gene	Target Size (bp)	Primer Sequence (Uppercase) (Primer tags required for SEQUENOM are in lowercase)
<i>LINE1</i>	103	F: <u>aggaagagag</u> TTTATATTTTGGTATGATTTTGTAG R: <u>cagtaatacgactcactatagggaaggct</u> ATCACCACCAAACCTACCCTAAA
<i>Alu</i>	261	F: <u>aggaagagag</u> GYGGTGGTTTAYGTTTGTAATTTTAGTAT R: <u>cagtaatacgactcactatagggaaggct</u> RCTCTATCRCCCAAACCTAAAATACAATA
<i>WT1 (1)</i>	328	F: <u>aggaagagag</u> GTAGGGGTTAGGGGAGGTAAAGT R: <u>cagtaatacgactcactatagggaaggct</u> CCCAATCACAATACAACCTACAATCA
<i>WT1 (2)</i>	454	F: <u>aggaagagag</u> TTTAGGAGTAGGTAGGGATTTAGGG R: <u>cagtaatacgactcactatagggaaggct</u> CCTCCAAAACCTATACAAAAAACCT
<i>ZNF876P</i>	452	F: <u>aggaagagag</u> TAGGTTTTTGTITGAGTTCGGTTAG R: <u>cagtaatacgactcactatagggaaggct</u> AATAAACTACTCCTCACCCACAAC
<i>ZSCAN1</i>	398	F: <u>aggaagagag</u> TTTGGATATTGTATAAAACGGTTGC R: <u>cagtaatacgactcactatagggaaggct</u> ATCATAACGACCCCTAAACTAACGA
<i>DGKG</i>	262	F: <u>aggaagagag</u> ATTTAGGATTTGGTTAGTAGGTTGTTT R: <u>cagtaatacgactcactatagggaaggct</u> TTATCCTCCCAAAAAAACACAAT
<i>MN1</i>	257	F: <u>aggaagagag</u> GTTGCTGGGGATGATGATTTT R: <u>cagtaatacgactcactatagggaaggct</u> CGAAACCGAACTAAACATAAACAC
<i>MEIS1</i>	452	F: <u>aggaagagag</u> TAATTTAGAATTGGATAATTTGGTAAGAG R: <u>cagtaatacgactcactatagggaaggct</u> CTACCTTCCTTACACCTTTCAAACA

Appendix II – Principal component analysis heatmaps for determination of data variation

Chapter 4 – Discovery & Pooled Analysis Heatmaps Respectively



Appendix III – Prognostic testing of probes from select genes with potential functional relevance to progression to relapse – Discovery analysis

CpG	Alias	UCSC RefGene Name	UCSC RefGene Group	Chr	$\Delta\beta$	p
cg15712177	cpG1	RUNX3	Body	chr1	-0.205	0.009774
cg01103597	cpG2	RUNX3	Body		-0.087	0.002596
cg19590532	cpG3	RUNX3	Body		-0.028	0.006646
cg20695936	cpG4	RUNX3	TSS1500		0.018	0.003921
cg18087266	cpG5	RUNX3	TSS1500		0.016	0.005873
cg18633684	cpG6	TXNRD1	5'UTR	chr12	-0.22772	0.003481
cg03817911	cpG7	TXNRD1	5'UTR		-0.259	0.011594
cg20923245	cpG8	TXNRD1	5'UTR		-0.23604	0.001998
cg27205904	cpG9	TXNRD1	5'UTR		-0.21662	0.012151
cg21234561	cpG10	TXNRD1	5'UTR		-0.23848	0.003651
cg26614816	cpG11	TXNRD1	5'UTR		-0.25027	0.004289
cg09477407	cpG12	TXNRD1	5'UTR		-0.22427	0.002887
cg05057777	cpG13	TXNRD1	1stExon		-0.22911	0.010608
cg08065092	cpG14	TXNRD1	3'UTR		-0.01687	0.002059
cg27300804	cpG15	LAMP1	Body	chr13	-0.14535	0.000399
cg01330448	cpG16	LAMP1	Body		-0.206	0.001697
cg20928950	cpG17	LAMP1	Body		-0.10429	0.01236
cg27055601	cpG18	LAMP1	Body		-0.15341	0.00443
cg13804476	cpG19	LAMP1	Body		-0.12044	0.010784
cg19236493	cpG20	HOXB3	Body	chr17	-0.031	0.008033
cg03260240	cpG21	HOXB3	5'UTR		-0.182	0.000567
cg24095592	cpG22	HOXB3	5'UTR		-0.170	0.009461
cg21676780	cpG23	BAHCC1	Body		-0.022	0.004945
cg11214532	cpG24	BAHCC1	Body		-0.057	0.00601
cg13309415	cpG25	BAHCC1	Body		-0.085	0.011204
cg05041061	cpG26	BAHCC1	Body		-0.132	0.004024
cg22515589	cpG27	BAHCC1	Body		-0.115	0.008725
cg06833110	cpG28	MEIS1	Body	chr2	-0.218	0.005374
cg04751149	cpG29	MEIS1	Body		-0.174	0.007431
cg10010036	cpG30	MEIS1	Body		-0.074	0.008828
cg22156632	cpG31	WNT6	Body		-0.143	0.010621
cg11511777	cpG32	WNT6	Body		-0.220	0.012655
cg23595222	cpG33	C22orf26	Body	chr22	0.116162	0.008258
cg26394940	cpG34	C22orf26	Body		0.129468	0.004703
cg11153400	cpG35	C22orf26	Body		0.151	0.003359
cg01080924	cpG36	C22orf26	Body		0.091251	0.002371
cg24349631	cpG37	C22orf26	1stExon		0.094028	0.008969
cg06976589	cpG38	MAEA	Body	chr4	-0.076	0.006455
cg18515868	cpG39	MAEA	Body		-0.10546	0.000664
cg24973755	cpG40	MAEA	Body		-0.15825	0.000916

cg10133462	cpg41	MAEA	Body		-0.10709	0.000265
cg18299068	cpg42	MAEA	Body		-0.10004	0.007475
cg02790486	cpg43	FGFR3	TSS1500		0.053841	0.00673
cg21339581	cpg44	FGFR3	Body		0.069	9.37E-05
cg13131329	cpg45	FGFR3	Body		0.049228	0.005154
cg05203875	cpg46	FGFR3	Body		0.031175	0.007617
cg06897306	cpg47	PTPRN2	Body	chr7	-0.015	0.005788
cg18004701	cpg48	PTPRN2	Body		-0.120	0.003255
cg03262885	cpg49	PTPRN2	Body		-0.030	0.004203
cg10552603	cpg50	PTPRN2	Body		-0.026	0.011951
cg02470440	cpg51	PTPRN2	Body		-0.051	0.010692
cg21649205	cpg52	PTPRN2	Body		-0.028	0.002487
cg23841288	cpg53	PTPRN2	Body		0.097	0.007425

Abbreviations: CpG - Cytosine-phosphate-guanine dinucleotide; Chr - Chromosome; $\Delta \beta$ - Delta beta.

Appendix IV - Prognostic testing of probes from select genes with potential functional
relevance to progression to relapse - Pooled-analysis

CpG	Alias	UCSC RefGene Name	UCSC RefGene Group	Chr	$\Delta\beta$	p
cg16557061	cg54	PTPRN2	Body	chr7	-0.042	0.006
cg18004701	cg55	PTPRN2	Body	chr7	-0.111	0.004
cg07529930	cg56	PTPRN2	Body	chr7	0.011	0.003
cg03784375	cg57	PTPRN2	Body	chr7	0.023	0.006
cg02323550	cg58	PTPRN2	Body	chr7	0.022	0.006
cg02331348	cg59	PTPRN2	Body	chr7	-0.038	0.006
cg02868521	cg60	PTPRN2	Body	chr7	-0.055	0.002
cg03262885	cg61	PTPRN2	Body	chr7	-0.022	0.007
cg21521989	cg62	PTPRN2	Body	chr7	0.033	0.007
cg01224451	cg63	PTPRN2	Body	chr7	-0.026	0.000
cg26211126	cg64	PTPRN2	Body	chr7	0.015	0.007
cg12364136	cg65	PTPRN2	Body	chr7	0.018	0.005
cg23841288	cg66	PTPRN2	Body	chr7	0.055	0.004
cg21608500	cg67	FGFR3	TSS1500	chr4	0.011	0.004
cg02790486	cg68	FGFR3	TSS1500	chr4	0.036	0.000
cg04742135	cg69	FGFR3	Body	chr4	0.017	0.005
cg02350535	cg70	FGFR3	Body	chr4	0.037	0.000
cg21339581	cg71	FGFR3	Body	chr4	0.030	0.000
cg13131329	cg72	FGFR3	Body	chr4	0.039	0.003
cg05203875	cg73	FGFR3	Body	chr4	0.022	0.001
cg19625524	cg74	FGFR3	Body	chr4	0.015	0.002
cg11545838	cg75	FGFR3	Body	chr4	0.018	0.001
cg03262295	cg76	FGFR3	Body	chr4	0.014	0.007
cg23595222	cg77	C22orf26	Body	chr22	0.096	0.006
cg26394940	cg78	C22orf26	Body	chr22	0.104	0.003
cg11153400	cg79	C22orf26	Body	chr22	0.127	0.002
cg01080924	cg80	C22orf26	Body	chr22	0.091	0.000
cg24349631	cg81	C22orf26	1stExon	chr22	0.105	0.005
cg09693106	cg82	C22orf26	1stExon	chr22	0.122	0.008
cg27131176	cg83	C22orf26	TSS200	chr22	0.055	0.006
cg15702185	cg84	MIRLET7A3	TSS1500	chr22	-0.107	0.002
cg24911721	cg85	MIRLET7A3	TSS200	chr22	-0.209	0.001
cg13251842	cg86	MIRLET7A3	TSS200	chr22	-0.125	0.001
cg03841312	cg87	MIRLET7A3	TSS200	chr22	-0.094	0.002
cg04063235	cg88	MIRLET7A3	TSS200	chr22	-0.121	0.002
cg22121941	cg89	MIRLET7A3	TSS200	chr22	-0.106	0.004
cg14800351	cg90	PAX6	Body	chr11	-0.162	0.010
cg04598774	cg91	PAX6	Body	chr11	-0.120	0.006
cg22272457	cg92	PAX6	TSS200	chr11	-0.062	0.008
cg02684104	cg93	FLOT1	Body	chr6	-0.095	0.005

cg18760837	cg94	FLOT1	Body	chr6	-0.107	0.002
cg17988780	cg95	FLOT1	Body	chr6	-0.089	0.003
cg19110902	cg96	FLOT1	Body	chr6	-0.086	0.006
cg21568145	cg97	GRLF1	3'UTR	chr19	-0.190	0.003
cg17431280	cg98	GRLF1	3'UTR	chr19	-0.237	0.004
cg08666831	cg99	GRLF1	3'UTR	chr19	-0.232	0.001

Abbreviations: CpG - Cytosine-phosphate-guanine dinucleotide; Chr - Chromosome; $\Delta \beta$ - Delta beta.