

Genome in 3D – regulation of lymphocyte development by genome architecture

Wing Fuk Chan
ORCID ID: 0000-0002-8719-7787

Submitted in total fulfilment for the degree of Doctor of Philosophy

September, 2020

The Walter and Eliza Hall Institute of Medical Research

Department of Medical Biology
Faculty of Medicine, Dentistry and Health Science
The University of Melbourne

Abstract

The functionally and phenotypically diverse cell populations that make up the immune system arise from the expression of a select part of a genome at a given time. The mechanisms governing such differential transcription are still, surprisingly, not fully understood. Only recently has the three-dimensional organisation of chromatin in the interphase nucleus been acknowledged to play a crucial role in modulating transcription. For instance, distal cis-regulatory elements like enhancers can form long-range chromatin loops with promoters to drive transcription, and these chromatin loops are in turn harboured in topologically associating domains (TADs), shielded from interference by outside elements. These three-dimensional structures can be lineage-specific and their roles during cellular differentiation are beginning to be uncovered.

The chromosome conformation, or genome architecture, in B- and T-lymphocytes, captured previously as Hi-C data in the lab, has provided immense information about lineage-specific DNA interactions that might be critical during differentiation. Based on this resource, the work herein aimed to develop an approach to identify, characterise and functionally dissect any novel and critical regulatory elements. Using this strategy, I have identified several putative T- and B-cell specific elements and subsequently adopted the CRISPR/Cas9 platform in generating large deletions as to dissect these elements. The approach has identified and confirmed the enhancers of T cell-specific transcription factors Bcl11b and Gata3.

Upon closer inspection an uncharacterised long non-coding RNA (lncRNA) Gm13218 was uncovered to associate with the enhancer of Gata3. Given the recent recognition of lncRNAs as important regulator of the 3D genome, I have retrieved the full-length sequence and characterised its expression pattern. It was found that expression of Gm13218 is highly correlated with that of Gata3 during early T cell development in thymus as well as T helper 2 (T_H2) cell differentiation. Knockdown and overexpression of Gm13218 transcripts, CRISPR-mediated silencing, activation, demethylation of the locus as well as interference of transcription elongation suggest that Gm13218 may be involved in the establishment, but not the maintenance of Gata3 expression.

By utilising Hi-C, RNA-seq, cell division and cell cycle indicators, the spatiotemporal dynamics of genome architecture during B cell activation and terminal differentiation into antibody-secreting cells was examined. It was revealed that genome organisation exhibit two discrete waves of restructuring – the first occurs just prior to the first cell division, with the resulting genome architecture being inherited through the subsequent rapid clonal expansion for many days until the second wave of restructuring upon differentiation into plasmablast. In addition, the first restructuring event was shown to precede the first DNA replication phase, suggesting that genome reorganisation is independent of, and well partitioned from, DNA synthesis and mitosis. In contrast, transcription underwent very early burst and was altered throughout the entire differentiation process. Further analysis suggests that transcription is intricately intermingled with genome organisation in a reciprocal fashion.

Overall, the work in this thesis has revealed a number of important findings regarding how the 3D genome controls the development and function of the immune system.

Declaration

This is to certify that,

The thesis comprises only my original work towards PhD except where indicated in the preface;

Due acknowledgement has been made in the text to all other material used;

The thesis is fewer than 100,000 words in length, exclusive of figures, tables and references.

Wing Fuk Chan (Luke)

Preface

Pursuant to the regulations governing the degree of Doctor of Philosophy at the University of Melbourne, this thesis comprises only my original work apart from the bioinformatic analyses of Hi-C and RNA-seq data (chapter 2.26-2.35), which were performed by Dr. Hannah Coughlan, and the pre-existing data – Hi-C, ATAC-seq, DNA methylation and CAGE – that was retrieved and duly referenced in thesis chapters.

Part of the material presented in this thesis has been submitted or published, and adapted, as follows:

Chan, W. F., Johanson, T. M., & Allan, R. S. (2020). Three-dimensional genome rewiring during the development of antibody-secreting cells. *Biochemical Society Transactions*.

Excerpts from this article have been directly adapted and referenced accordingly in p.40, 48-55, 155, 184, 186, 194.

Chan, W.F., Coughlan, H.D., Iannarella, N., Smyth, G.K., Johanson, T.M., Keenan, C.R. & Allan, R.S. (2020). Identification and characterization of the long non-coding RNA Dreg1 as a novel regulator of Gata3. *Immunology and Cell Biology*.

Excerpts from this article have been directly adapted and referenced accordingly in p.64-67, 70, 77, 82, 149, 150, 153, 154.

Chan, W.F., Coughlan, H.D., Zhou, J.H., Keenan, C.R., Bediaga, N.G., Hodgkin, P.D., Smyth, G.K., Johanson, T.M. and Allan, R.S., 2019. (2020). Pre-mitotic genome reorganisation bookends the cellular differentiation process. (In revision at *Nature Communications*, preprint available at bioRxiv.)

Excerpts from this article have been directly adapted and referenced accordingly in p.65, 83-89.

Acknowledgement

PhD is no easy task, there are many people who have helped me through this journey. I would like to take this opportunity to thank my supervisor Dr. Rhys Allan, who took me as his first, and so far still his only, PhD student, bringing me to Australia from the other side of the globe across the Pacific Ocean. The inspiring discussions regarding various topics in science throughout the years were those that I treasured and enjoyed. The astonishing depth of knowledge and admirable wealth of ideas have thus encouraged me to think critically and logically, and have definitely supported me in pursuing the known unknowns in the field. The leisure, less serious chat during lunch and other times has also let me quickly settle in this new environment, and very much feel respected and valued. I am as well grateful for the freedom that I enjoyed in the lab to independently tackle questions that most interest me. All these have made my PhD research a feasible and enjoyable pursuit of knowledge. I am much indebted.

I would also like to thank Dr. Tim Johanson for his support and guidance through different techniques, advice and tips for presentation, as well as frequent sharing of thoughts and views towards science, research environment and past experiences. I also thank Prof. Stephen Nutt, who provided important suggestions to my PhD project, and whose immense knowledge in immunology is what everyone can confidently count on. I would also like to thank my PhD committee, Prof. Gordon Smyth and Dr. Menno van Zelm, who oversaw my project and kindly provided advice in the progress of the project. I particularly feel thankful for the collaboration of our lab with Prof. Gordon Smyth who supervised all the bioinformatic analyses in the project.

I would then extend my thanks to members of the Allan lab. Especially Dr. Hannah Coughlan, who actively took part in the project in the dry lab side, analysing all the big sequencing data and generating lots of impressive plots. Thanks must also be conveyed to Ms Nadia Iannarella who incessantly keeps the lab and mouse colonies in a most organised, well-stocked and tidiest fashion which is not seen in other labs. The absolute assurance has definitely facilitated

my research in many ways and is thus most appreciated. I also thank Dr. Christine Keenan for her active participation in discussing various aspects in the project, and for the recent help with the preparation of the manuscript.

I would also like to express my gratitude towards various labs and facilities in the institute that have contributed to the work. Genomics lab and in particular Dr. Stephen Wilcox who has run all of my sequencing experiments. The FACS lab and all the flow cytometry technicians for sorting all of the precious samples. The Bio-services department and the animal technicians – Ms. Teisha Mason and Ms. Rhiannan Crawley, who have helped with the mouse work and husbandry. I would also like to thank Dr. Macro Herold, Ms Lin Tai, Dr. Chris Tonkin, Prof. Phil Hodgkin and Dr. Melissa Holmes for generous provision of reagents, equipment and mouse strain.

I am as well very grateful to be within the community of the Immunology Division, and probably the entire institute. Having past research experiences in isolated labs across the Pacific half-buried in the Canadian snow, I have never been exposed to such large and interactive community full of diverse knowledge and expertise. The power of such – scientific collaborations and connections that were extensively fostered and built up – has no doubt impressed and benefited me a lot. I am extremely fortunate to get along with several friends and colleagues and tea buddies. Simon, Miles, Kylie and most importantly Izzie, Steph, Ange and Wil. For the most enjoyable chats, whinges, distractions, mental supports and science discussions. As a semi-antisocial international student, this is very much almost all the socialising that I have, and this meant a lot to me. I sincerely hope that we are still able to keep it that way even after PhD, perhaps even when we are scattered in different places around the globe.

Lastly, to my family, Fiona and friends outside academia, for the patience and understanding, as well as the willingness to keep in touch with me despite being thousands of miles away. The long-range communication, through the air instead of 3D looping which you shall see in this thesis, has for sure supported and motivated me from time to time.

Table of Contents

1. Introduction	1
1.1. Genome in 3D	1
Non-coding DNAs.....	1
Promoters and enhancers.....	2
Silencers.....	3
Epigenetic modifications to chromatin.....	5
Repressive loops	6
Gene loops	6
CTCF and topologically associating domains (TADs)	7
Phase separation and compartments.....	11
Nuclear space and chromosome territories	12
Nuclear envelope and lamina associated domain	12
Nuclear pores – an exception to peripheral silencing	13
Nuclear bodies, transcription hubs and phase separation	15
Loop extrusion vs phase separation	16
Non-coding RNAs in chromatin biology	17
Alterations of chromatin landscape by non-coding RNAs	18
Non-coding RNAs in modulating genome conformation	21
Genome architecture during cell cycle	22
1.2. Methods of investigating genome architecture	24
Microscopy-based.....	24
Proximity ligation – chromosome conformation capture.....	24
High throughput chromosome conformation capture – Hi-C.....	25
Manifestation of chromatin structures in Hi-C data	28
Improved resolution C-technology – micro-C.....	31
Ligation independent method – genome architecture mapping.....	31
Multi-way assessment of DNA interactions – SPRITE	32
1.3. Methods to manipulate the genome	33
Genome editing	33
Engineering the epigenome.....	35
Constructing the 3D genome	38
1.4. Genome reorganisation during B cell development and differentiation.....	40
Early B cell development	40
V(D)J recombination	43
B cell activation.....	48
Class switch recombination	49
Terminal differentiation into antibody-secreting cells	53
1.5. Genome reorganisation during T cell development and differentiation	56
Early T cell development.....	56
T cell activation and terminal differentiation	57
1.6. Research aims	62
2. Materials and methods	63
2.1. Mice.....	63
2.2. Cell culture	63
2.3. B cells activation and differentiation	63
2.4. Isolation of thymic T cells	64
2.5. T_H1 and T_H2 cell polarisation.....	64

2.6.	In situ Hi-C	65
2.7.	RNA-seq	65
2.8.	Plasmids construction	66
2.9.	gRNA design	70
2.10.	gRNA cloning.....	70
2.11.	In vitro transcription (IVT) of sgRNA	71
2.12.	Ribonucleoprotein (RNP) assembly	73
2.13.	Electroporation	73
2.14.	Screening and isolation of knockout clones	74
2.15.	Polyadenylation signal insertion	76
2.16.	Lentivirus production	77
2.17.	Transduction	77
2.18.	Droplet digital PCR (ddPCR).....	78
2.19.	T7 endonuclease I (T7EI) assay	79
2.20.	End point reverse transcription PCR (RT-PCR).....	79
2.21.	Quantitative reverse transcription PCR (RT-qPCR)	80
2.22.	5' RACE (Rapid amplification of cDNA ends)	81
2.23.	3' RACE	82
2.24.	Antisense oligonucleotide (ASO) mediated knockdown.....	82
2.25.	<i>PiggyBac</i> non-coding RNA overexpression	83
2.26.	Hi-C pre-processing	83
2.27.	Detecting differential interactions (DIs).....	83
2.28.	Detecting TAD boundaries.....	84
2.29.	Detecting differential TAD boundaries	85
2.30.	Detection of A/B compartments	85
2.31.	Detecting differentially interacting promoters (DIPs)	86
2.32.	Detecting correlation between structures and gene expression	87
2.33.	Visualisation of Hi-C	88
2.34.	Differential expression analysis.....	88
2.35.	Gene set enrichment.....	89
3.	Exploring and identifying functional cis-regulatory elements of lymphocytes ...	90
3.1.	Results	92
	<i>Immune lineage-restricted TADs coincide with lineage-specific genes.....</i>	<i>92</i>
	<i>Differential chromatin accessibility profiles enable discoveries of novel cis-elements</i>	<i>95</i>
	<i>Genome structures are preserved in lymphoma cell line EL4 and A20</i>	<i>97</i>
	<i>CRISPR/Cas9 RNP intracellular concentration does not influence deletion efficiency</i>	<i>99</i>
	<i>The effects of the removal length to deletion efficiency</i>	<i>101</i>
	<i>In vitro transcribed sgRNA confers similar efficiency to chemically synthesised crRNA system</i>	<i>103</i>

<i>Removal of cis-element B1 leads to downregulation of Bcl11b</i>	105
<i>Loop anchors confers no regulatory functions in steady state Ebf1 expression</i>	108
<i>Multiple cis-regulatory elements modulate Gata3 transcription</i>	110
3.2. Discussion	112
4. Identification and characterisation of the lncRNA Gm13218 at enhancer of Gata3	117
4.1. Results	119
<i>A long non-coding RNA Gm13218 is located in an enhancer of Gata3</i>	119
<i>Gm13218 locus DNA is differentially methylated and differentially accessible</i>	121
<i>Gm13218 is co-regulated with Gata3 in thymus</i>	123
<i>Gm13218 is co-regulated with Gata3 during T_H2 differentiation</i>	125
<i>Gm13218 is co-regulated with Gata3 during early T cell lineage commitment</i>	127
<i>T cell specific TAD and non-coding transcription are conserved between mouse and human</i>	129
<i>Antisense oligonucleotide knockdown of Gm13218 does not affect steady state Gata3</i>	131
<i>CRISPRi via dCas9-KRAB silence Gm13218 and prolonged targeting downregulate Gata3</i>	133
<i>Introduction of polyadenylation signal near Gm13218 TSS reduces Gm13218 but not Gata3</i>	135
<i>The promoter of Gm13218 does not affect steady state Gata3 transcription in EL4</i>	140
<i>Exogenous source of Gm13218 modestly upregulates Gata3 in A20 B cells</i>	142
<i>CRISPRa via SunTag-VP64 and SAM, and DNA demethylation failed to activate Gm13218</i>	145
<i>Combining CRISPRa with DNA demethylating agent activates Gm13218 but not Gata3</i>	147
4.2. Discussion	149
5. Characterising genome reorganisation during B cell clonal expansion and differentiation	155
5.1. Results	157
<i>B cells undergo first cell division after 30-33 hours upon LPS stimulation</i>	157
<i>Different stages of activated B cells were harvested for Hi-C and RNA-seq</i>	159
<i>Two discrete waves of genome restructuring across B cell activation and differentiation</i>	161
<i>The first major genome reorganisation event occurs prior to first cell division</i>	163
<i>The second major genome reorganisation occurs upon differentiation into plasmablast</i>	165
<i>TADs increase in size immediately prior to the first cell division</i>	167
<i>Compartments are largely invariant across B cell activation and differentiation</i>	169
<i>Dynamic transcriptional changes throughout the course of B cell activation and differentiation</i>	171
<i>Gene ontology unveils the stage-specific cellular changes across B cell activation</i>	173
<i>Remodelling of interactions around promoters correlates with transcriptional changes</i>	175
<i>Genome reorganisation occurs before S phase prior to first cell division</i>	177
<i>Terminal differentiation into plasmablast associates with prolonged time in G₁ phase</i>	179
5.2. Discussion	182
6. General Discussion	187
<i>Genomic structures and the relationship with gene transcription</i>	187
<i>Requirement for precision genome editing to examine mediators of transcription</i>	190
<i>Waves of cell cycle related genome reorganisation during B cell differentiation</i>	191
<i>Overall conclusions</i>	196
Reference	197

List of Figures

Figure 1.1. Different chromatin loops in genome at kilobase to sub-megabase level.	4
Figure 1.2. Two major molecular mechanisms of genome organisation.	10
Figure 1.3. Nuclear space in interphase nucleus.	14
Figure 1.4. Mechanisms of non-coding RNA in remodelling 3D genome.....	20
Figure 1.5. Methodology of high throughput chromosome conformation capture (Hi-C).	27
Figure 1.6. Manifestation of different genome structures in Hi-C contact matrices.	30
Figure 1.7. The expanding CRISPR toolbox in manipulating genome.	36
Figure 1.8. B lymphocyte development.	42
Figure 1.9. TADs and chromatin loops in IgH locus during V(D)J recombination.	45
Figure 1.10. Chromatin loops during class switch recombination.	52
Figure 1.11. Genome reorganisation during terminal plasma cell differentiation.	55
Figure 1.12. Genome organisation in naïve CD4+ T cells.....	60
Figure 3.1. B cell specific TADs.....	93
Figure 3.2. T cell specific TADs.....	94
Figure 3.3. Differential accessibility profiles allow identification of potential cis-regulatory elements.	96
Figure 3.4. Genome structures are preserved in lymphoma cell line EL4 and A20.	98
Figure 3.5. CRISPR/Cas9 RNP concentration does not affect deletion efficiency.....	100
Figure 3.6. The effects of deletional length to deletion efficiency.	102
Figure 3.7. IVT sgRNA confers similar deletion efficiency to chemically synthesised crRNA system.....	104
Figure 3.8. Deletion of cis-element B1 leads to downregulation of Bcl11b.	107
Figure 3.9. Deletion of putative cis-elements does not significantly affect Ebf1 expression in steady state.	109
Figure 3.10. Identification of functional cis-elements regulating Gata3 transcription.	111
Figure 4.1. Expression of long non-coding RNA Gm13218 at Gata3 enhancer.	120
Figure 4.2. Chromatin accessibility and DNA methylation profiles at the Gm13218 gene.	122
Figure 4.3. Expression pattern of Gata3 and Gm13218 across different tissues.	124
Figure 4.4. Expression of Gata3 and Gm13218 during TH1 and TH2 differentiation.	126
Figure 4.5. Gm13218 is co-regulated with Gata3 during early T cell lineage commitment.	128
Figure 4.6. Conservation of GATA3 TAD, lncRNA expression and synteny in human.....	130
Figure 4.7. Gm13218 knockdown via ASO confers no influence to Gata3 expression.	132

Figure 4.8. CRISPR interference via KRAB domain silences Gm13218 and prolonged targeting leads to downregulation of Gata3.....	134
Figure 4.9. Introduction of SV40 poly(A) signal in EL4 near Gm13218 TSS...	137
Figure 4.10. Introduction of poly(A) signal reduces Gm13218 transcription but not Gata3.....	139
Figure 4.11. Promoter of Gm13218 does not regulate steady state Gata3 in EL4.	141
Figure 4.12. Ectopically overexpression of Gm13218 modestly upregulate Gata3 in A20.....	144
Figure 4.13. CRISPR activation and demethylation failed to activate Gm13218 in A20.	146
Figure 4.14. Combining SAM with SunTag-TET1 activates Gm13218 but not Gata3.....	148
Figure 5.1. B cells undergo cell division after 30-33 hours upon LPS stimulation.	158
Figure 5.2. Isolation of different stages of activated or differentiated B cells. .	160
Figure 5.3. Two discrete waves of genome reorganisation events across B cell activation with first occurs prior to first cell division.	162
Figure 5.4. Genome reorganisation events prior to first cell division.	164
Figure 5.5. Genome reorganisation events upon terminal differentiation.....	166
Figure 5.6. TAD increase in size immediately prior to first cell division.....	168
Figure 5.7. Compartments are largely invariant across B cell activation and differentiation.	170
Figure 5.8. Dynamic transcriptional changes throughout the course of B cell activation.	172
Figure 5.9. Gene ontology unveils the stage-specific cellular changes across B cell activation.	174
Figure 5.10. Remodelling of interactions around promoters correlates with transcriptional changes.....	176
Figure 5.11. The first genome reconfiguration event occurs prior to DNA synthesis phase.....	178
Figure 5.12. Cell cycle analysis of LPS activated FUCCI B cells over a five-day period.....	180
Figure 5.13. Terminal differentiation into plasmablast associates with prolonged G1 phase.	181
Figure 6.1. During differentiation the genome is reorganised prior to mitosis.	193
Figure 6.2. Cellular events during B cell clonal expansion and differentiation.	195

List of Tables

Table 2.1 Antibodies used for flow cytometry	64
Table 2.2 Primers for plasmid construction	69
Table 2.3 Coordinates (mm10) and designation of target genomic regions	70
Table 2.4 Primers for gRNA cloning	71
Table 2.5. Specific primers for generation of in vitro transcription templates	72
Table 2.6. Chemically synthesised crRNA	73
Table 2.7. Electroporation parameters for 4D-Nucleofector	73
Table 2.8. Primers for screening positive clones	75
Table 2.9. Droplet digital PCR primers and probes	78
Table 2.10. RT-qPCR primers and probes	80
Table 2.11. GapmeRs antisense oligos.....	82
Table 3.1. Lineage-specific TADs and the putative cis-regulatory elements.....	95

List of Abbreviations

3C	Chromosome conformation capture
3'RR	3' regulatory region
4C	Chromosome conformation capture-on-chip
53BP1	p53-binding protein 1
5C	Chromosome conformation capture carbon copy
ABA	Absciscic acid
ABI	ABA insensitive
AID	Activation-induced cytidine deaminase
ASO	Antisense oligonucleotide
ATAC	Assay for transposase-accessible chromatin
ATM	Ataxia-telangiectasia mutated
BAC	Bacterial artificial chromosome
BER	Base excision repair
Blimp1	B lymphocyte induced maturation protein 1
bp	Base pair
BRD4	Bromodomain-containing protein 4
CAGE	Cap analysis of gene expression
Cas	CRISPR-associated protein
CAT	Chloramphenicol acetyltransferase
CIB1	Cryptochrome-interacting basic helix-loop-helix 1
CIP	Calf intestinal alkaline phosphatase
circRNA	Circular RNA
CLOuD9	Chromatin loop reorganisation using dCas9
CLP	Common lymphoid progenitor
Cpf1	CRISPR from <i>Prevotella</i> and <i>Francisella</i> 1
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRi	CRISPR interference
CRISPRa	CRISPR activation
CRISPR-GO	CRISPR genome organiser
crRNA	CRISPR RNA
CRY2	Cryptochrome 2
CSR	Class switch recombination

CTCF	CCCTC-binding factor
CTV	CellTrace Violet
dCas9	Catalytically dead/inactive Cas9
ddPCR	Droplet digital PCR
DEG	Differentially expressed gene
DI	Differential interaction
DN	Double negative
DNMT3A	DNA methyltransferase 3A
DP	Double positive
DSB	Double-strand break
E2A	Immunoglobulin enhancer-binding factors E12/E47
EBF1	Early B cell factor 1
eRNA	Enhancer RNA
EST	Expressed sequence tag
Exo I	Exonuclease I
EZH2	Enhancer of zeste homolog 2
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FISH	Fluorescent <i>in situ</i> hybridisation
FKBP	FK506 binding protein 12
FRB	FKBP rapamycin binding
FUCCI	Fluorescent ubiquitination-based cell cycle indicator
G9a	Euchromatic histone-lysine N-methyltransferase 2
GA	Gibberellin
GAI	Gibberellic acid insensitive
GAM	Genome architecture mapping
GCN4	General control protein GCN4
GID1	GA insensitive dwarf 1
GLT	Germline transcription
HDR	Homology directed repair
Hi-C	High-throughput chromosome conformation capture
HOTTIP	HoxA distal transcript antisense RNA
HP1	Heterochromatin protein 1

HPC	Haematopoietic progenitor cell
HSF1	Heat shock factor 1
IDR	Intrinsically Disordered Region
IFN	Interferon
Ig	Immunoglobulin
IGCR1	Intergenic control region 1
Igh	Immunoglobulin heavy chain
Igk	Immunoglobulin κ light chain
Ig λ	Immunoglobulin λ light chain
IL	Interleukin
Immgen	The Immunological Genome
Indel	Insertions and deletions
ITR	Inverted terminal repeat
IVT	In vitro transcription/transcribed
KRAB	Krüppel associated box
LAD	Lamin-associated domain
LADL	Light activated dynamic looping
LCR	Locus control region
LNA	Locked nucleic acid
lncRNA	Long non-coding RNA
LPS	Lipopolysaccharide
IRES	Internal ribosome entry site
LSD1	Lysine-specific histone demethylase 1A
LTR	Long terminal repeat
MeCP2	Methyl CpG binding protein 2
miRNA	MicroRNA
MMR	Mismatch repair
MNase	Micrococcal nuclease
NHEJ	Non-homologous end joining
NK	Natural killer
NPC	Nuclear pore complex
nt	Nucleotide
ORCA	Optical reconstruction of chromatin architecture

p300	Histone acetyltransferase p300
p65	NF- κ B p65 subunit
PAIR	PAX5-activated intergenic repeat
PALM	Photo-activated localisation microscopy
PAM	Protospacer adjacent motif
PAX5	Paired Box 5
PCA	Principal component analysis
PCR	Polymerase chain reaction
PhyB	Phytochrome B
PIC	Preinitiation complex
PIF	Phytochrome interacting factor
piRNA	Piwi-interacting RNA
PML	Promyelocytic leukaemia
PNK	Polynucleotide kinase
Pol II	RNA polymerase II
PRC2	Polycomb Repressive Complex 2
PYL1	Pyrabactin resistance 1-Like 1
RACE	Rapid amplification of cDNA ends
RAG	Recombination activating gene
RC	Recombination centre
RNP	Ribonucleoprotein
RppH	5' Pyrophosphohydrolase
rSAP	Recombinant shrimp alkaline phosphatase
RSS	Recombination signal sequences
Rta	Epstein-Barr virus R trans-activator
RT-qPCR	Quantitative reverse transcription PCR
SABER-FISH	Signal amplification by exchange reaction FISH
SAM	Synergistic activation mediator
scFv	Single-chain variable fragment
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
SID	mSin3 interaction domain
siRNA	Small interfering RNA

SNP	Single nucleotide polymorphism
SPRITE	Split-pool recognition of interactions by tag extension
ssODN	Single-strand oligodeoxynucleotide
STAT	Signal transducer and activator of transcription
STORM	Stochastic optical reconstruction microscopy
Suv39H1	Suppressor of variegation 3-9 homolog 1
T7EI	T7 endonuclease I
TAD	Topologically associating domain
TALEN	Transcription activator-like effector nuclease
TCR	T cell receptor
TET1	Ten-eleven translocation methylcytosine dioxygenase 1
tracrRNA	Trans-activating crRNA
TSS	Transcription start site
UPR	Unfolded protein response
V(D)J	Variable, Diversity, and Joining gene segments
VP16	Herpes simplex virus protein vmw65
VP64	4 tandem repeats of VP16
VPR	VP64-p65-Rta
YY1	Yin Yang 1
ZFN	Zinc-finger nuclease

1. Introduction

The immune system comprises of a variety of cell types, that have developed from a common progenitor, with each performing a distinct type of functions. For instance, B lymphocytes secrete specific antibodies to neutralise invading pathogens, whereas T lymphocytes engage in the clearance of virally infected cells, or release different cytokines to remotely signal and coordinate an adaptive immune response. It is well known that lineage-specific gene expression programmes regulate these different functions. However, interestingly, the genome of these populations is almost identical suggesting that the processes that regulate how and when genes are expressed are critical to the development of a healthy immune system. This thesis explores how the 3D structure of the genome contributes to the regulation of gene expression and immune cell developmental processes.

1.1. Genome in 3D

Non-coding DNAs

The completion of the Human Genome Project in 2003 has allowed scientists access to almost the entire genetic blueprint upon which a human is built (Lander et al., 2001; Schmutz et al., 2004). Before that, on top of the approximately 10,000 mRNAs that had been curated (Lander et al., 2001; Pruitt and Maglott, 2001), the number of expressed sequence tags (ESTs) captured have prompted scientists to anticipate that there are around 50,000 to 100,000 protein-coding genes present in the human genome (Pertea and Salzberg, 2010). To their extreme surprise only approximately 20,000 of such genes were found (Lander et al., 2001; Schmutz et al., 2004). The amount is comparable to that of a nematode worm (Consortium, 1998) or a zebrafish (Howe et al., 2013), which are primitive and much less anatomically complex than humans. Even more astounding is that the majority of the genes in such primitive organisms have their direct equivalences or orthologous counterparts in human (Lai et al., 2000). The lack of correlation between the number of protein-coding genes and the complexity of an organism has puzzled scientists for a very long time (Gall, 1981; Gregory, 2005; Hahn and

Wray, 2002). Pleiotropy and epistasis are suggested to be the main contributing factors to resolving this paradox where a gene has functional influences on many other genes and therefore obtains multiple roles at different place and time (Gibson, 1996). This spatiotemporal gene interaction network, and hence, gene regulation, enables different cell types, tissues and organs to arise from one genetic blueprint. Also, it allows cells and organisms to rapidly respond to the ever-changing environment, for example during pathogen assault. Overall, among the vast genetic landscape of humans, protein-coding genes constitute only about 2% and the remaining 98% of non-coding DNAs (Lander et al., 2001), which were once dubbed as “junk DNA” as they serve no apparent functions, are recently unravelled to be extremely critical in regulating gene transcription, as they are thought to contain regulatory elements that influence cell type specific gene transcription (Gibson, 1996).

Promoters and enhancers

Over the recent decades, a number of important non-coding DNA sequences have been identified. These, termed as cis-regulatory elements based on the fact that they act on the same DNA molecule as they originate, have been characterised in controlling gene transcription. The best defined of these are promoters. Promoters are sequences immediately upstream of the transcription start site where RNA polymerase II (Pol II) and other general transcription factors bind and assemble into transcription preinitiation complex (PIC), which subsequently unwind DNA and start RNA synthesis (Maston et al., 2006). Enhancers are cis-elements that can augment transcription in a position- and orientation-independent manner. They can locate on either side of the promoter over a vast linear genetic distance, or it can be embedded within an intron of a gene. The underlying mechanism has remained enigmatic for many years until it was demonstrated that enhancers of the β -globin at 40-60 kb away from the locus come into close physical proximity with the corresponding promoter in erythroid cells (Fig. 1.1) (Palstra et al., 2003; Tolhuis et al., 2002). The DNA loop is erythroid specific and not found in other cells which do not express β -globin. It is therefore suggested that enhancer, via a three-dimensional loop with promoter, recruits additional transcription factors and thereby stabilises the preinitiation

complex at the transcription start site (Tolhuis et al., 2002). It has been proposed that such interaction between enhancers and promoters can theoretically activate genes in two different ways, either in an instructive or in a permissive manner (de Laat and Duboule, 2013). In the instructive model gene transcription is triggered immediately upon *de novo* enhancer-promoter loop formation; whereas in the permissive model the loops are poised for activation, with transcription delayed until further signal is received.

Silencers

Silencers, on the other hand, can suppress gene transcription through a similar position- and orientation-independent fashion (Fig. 1.1). Despite the first discovery over three decades ago (Brand et al., 1985), the roles of silencers in gene regulation remain poorly explored. There were only a handful of investigations characterising such an important cis-element, with the most extensive ones being the CD4 gene in CD8+ T cell (Donda et al., 1996; Sawada et al., 1994; Siu et al., 1994; Taniuchi et al., 2002a; Taniuchi et al., 2002b). The biased focus on enhancers over silencers is probably attributed to the presumption that the default transcriptional state in a genome is repressive and thereby gene activation is a more specific phenomenon (Maston et al., 2006). In addition, other mechanisms of gene silencing have taken priority over the study of silencers. These epigenetic mechanisms, so named as they work outside of the genetic sequences, are DNA methylation and histone modifications.

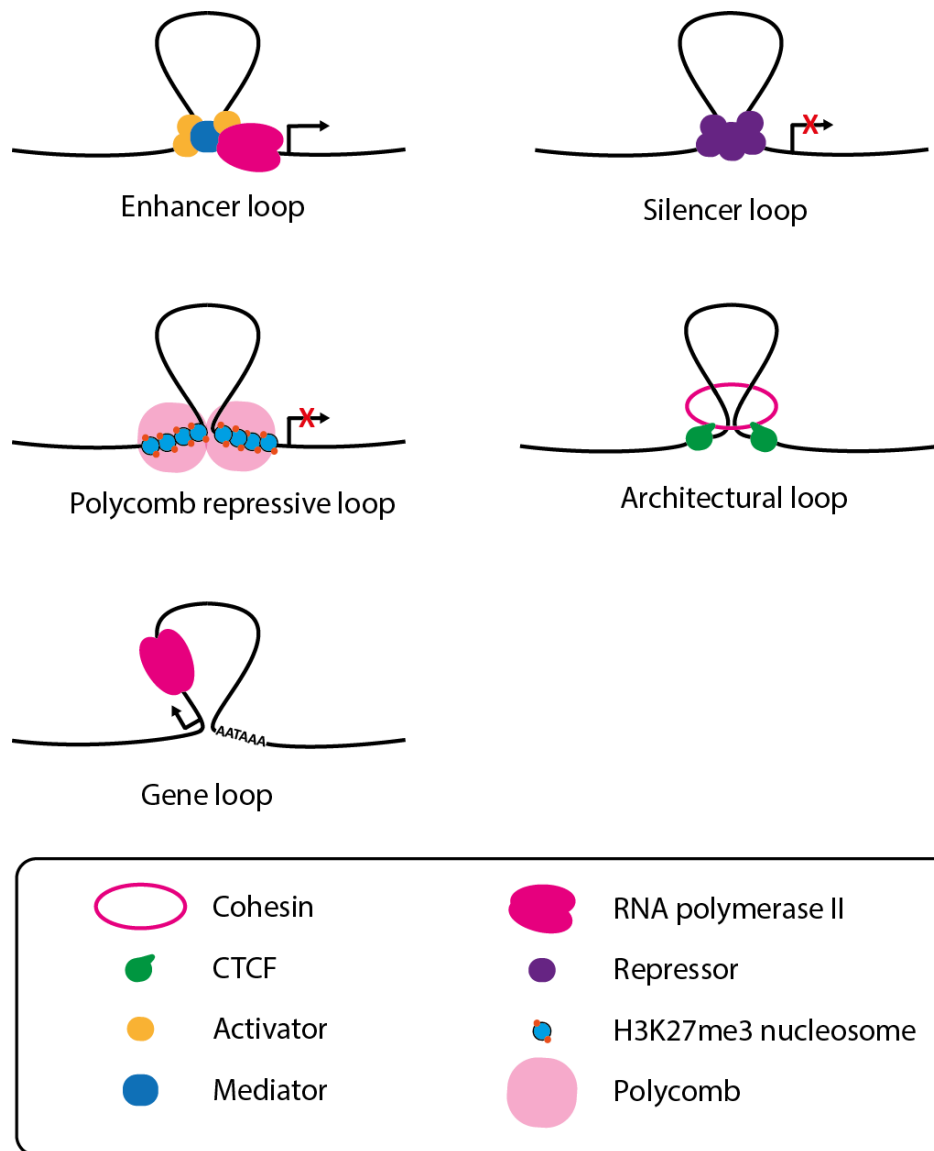


Figure 1.1. Different chromatin loops in genome at kilobase to sub-megabase level.

Enhancer-promoter interactions mediate gene transcription, whereas silencer-promoter loops suppress gene expression. Repressive regions bound by Polycomb repressive complex can also form long-range chromatin loops. Loops can also be established by the action of cohesin and CTCF, or formed between the transcription start site and the termination site as gene loop.

Epigenetic modifications to chromatin

DNA sequences can be chemically modified to alter the transcriptional status of a gene. One nucleotide, cytosine, is subject to methylation which converts it to 5-methylcytosine (Deaton and Bird, 2011). In vertebrates, this occurs almost exclusively at the CpG dinucleotide motif. DNA methylation in general is associated with transcriptional repression and in vertebrate genomes some CpG dinucleotides happen to aggregate in high density, forming the so-called CpG islands. These CpG islands very often coincide with gene promoters and methylation of this site is commonly associated with gene silencing. While methylation of cytosine can itself potentially block transcription factors from binding (Eden and Cedar, 1994), it can also be bound by proteins such as methyl CpG binding protein 2 (MeCP2) resulting in the condensation of the surrounding chromatin (Jones et al., 1998), further preventing underlying genes from access to transcription machinery.

Chromatin is a complex composed of 146 base pairs of DNA wrapped around histone proteins to form nucleosomes. Histones are mostly globular in structure except for the N-terminal peptide tails, which protrude out and float around the nucleosomes (Kouzarides, 2007). These tails are subject to a variety of chemical modifications, such as acetylation, methylation, phosphorylation, sumoylation and ubiquitylation, which subsequently enable the recruitment of various chromatin-modifying enzymes to loosen or tighten chromatin, hence regulate gene expression. Chromatin bearing marks such as trimethylation of lysine 4 of histone-H3 (H3K4me3) and H3K27ac are associated with euchromatin, in which the chromatin state is generally more relaxed, allowing access by transcription factors and therefore permissive to transcription. On the other hand, H3K9me3 and H3K27me3 is associated with heterochromatin in which chromatin is generally more compacted and transcription is prohibited (Grewal and Jia, 2007; Kouzarides, 2007). There are two variants of heterochromatin, constitutive heterochromatin is found around centromeres and telomeres, is highly condensed and contains a high proportion of DNA repeat elements, such as satellite repeats and transposable elements. Constitutive heterochromatin is relatively invariable throughout cell cycle and between cell types, with its formation contributed mainly by H3K9me3 and HP1 (heterochromatin protein 1).

Contrarily, the repressive state in facultative heterochromatin can be altered depending on cell types, cellular signal and activity, and it exhibits histone mark H3K27me3, which is mainly laid down by Polycomb Repressive Complex 2 (PRC2).

Repressive loops

Despite the speculation that DNA loops are limited to accessible DNA regions like enhancers and promoters, surprisingly it has been demonstrated that the Polycomb-repressed regions can also mediate such long-range loop formation (Fig. 1.1) (Bantignies et al., 2011; Ogiyama et al., 2018; Tiwari et al., 2008). Two recent studies have further delved into these repressive loops by a genome-wide approach (Cai et al., 2019; Ngan et al., 2020), it is found that promoters are quite often involved in loops that are bound by PRC2 or rich in H3K27me3. Some of these DNA interactions are tissue specific and in some cases deletion of one of the anchor points resulted in transcriptional upregulation of the interacting genes, demonstrating an apparent silencer nature (Cai et al., 2019; Ngan et al., 2020). As mentioned above, the nature, or exact identity, of long-range silencers is very obscure, owing to this overlapping role of repressive histone marks. Two very recent reports have tried to address this problem by uncovering some of the *bona fide* silencers, which lies in accessible regions and harbours specific DNA motifs for repressor protein binding (Doni Jayavelu et al., 2020; Pang and Snyder, 2020). They have identified thousands of silencer elements which extensively interact with promoters via long-range chromatin contacts (Fig. 1.1), with some of them being able to act on multiple genes. These silencers are shown to associate with specific transcription factors and thus presumably operate in a tissue specific manner. Overall, the studies demonstrated that silencers, akin to enhancers, can work via DNA looping, and their frequent occurrence supports them as being important cis-regulatory elements.

Gene loops

Loops can also form between promoters and the 3'-ends of genes, which are referred as gene loops (Fig. 1.1). First identified in yeast (O'Sullivan et al., 2004), gene loops are thought to regulate the directionality of transcription, by restricting

unproductive transcription from an otherwise bidirectional promoter (Tan-Wong et al., 2012). In addition, gene loops can confer transcriptional memory by promptly recycling RNA polymerase II from the termination sites back to the promoters, hence enabling a more rapid transcription re-initiation (Laine et al., 2009; Shandilya and Roberts, 2012; Tan-Wong et al., 2009). The looping was shown to be mediated by the Pol II itself as well as a general transcription factor TFIIB (Allepuz-Fuster et al., 2019; Singh and Hampsey, 2007), suggesting a self-regulatory role of PIC in modulating gene expression. Albeit being largely documented in yeast, gene loops were recently unveiled to be a significant phenomenon in human and mouse as well, for instance, in *BRCA1* gene in human and mouse breast tissue (Tan-Wong et al., 2008). Furthermore, it was also demonstrated in human germinal centre B cells extensive 5'-3' gene loops are formed genome-wide (Bunting et al., 2016). These loops, absent in naïve B cells, were shown to associate with gene upregulation, suggesting a potential global regulatory role in gene expression.

CTCF and topologically associating domains (TADs)

Other than promoters, enhancers and silencers, traditional genetics has further described another cis-regulatory element known as insulator, that can disrupt the enhancer-promoter communication when it is located between the two elements (Bell et al., 1999). Over decades of active research by scientists the CCCTC-binding factor (CTCF) is the only known example of insulator element in vertebrate (Merkenschlager and Nora, 2016). Unlike enhancers, CTCF operates in an orientation- and position-dependent manner and in certain cases the methylation status of DNA is critical to the function (Bell and Felsenfeld, 2000; Hark et al., 2000). Again, the mode of action of CTCF remained unclear until the recent development of chromosome conformation capture (3C) technique, which exploits a proximity ligation assay to determine the physical distance between two DNA elements in three-dimension (Dekker et al., 2002). With the adoption of next generation sequencing (Hi-C) in the platform a high-throughput approach was later developed to evaluate all the DNA interactions in the genome (Lieberman-Aiden et al., 2009; Rao et al., 2014). It revealed that the genome is spatially segregated into numerous discrete chromatin globules known as

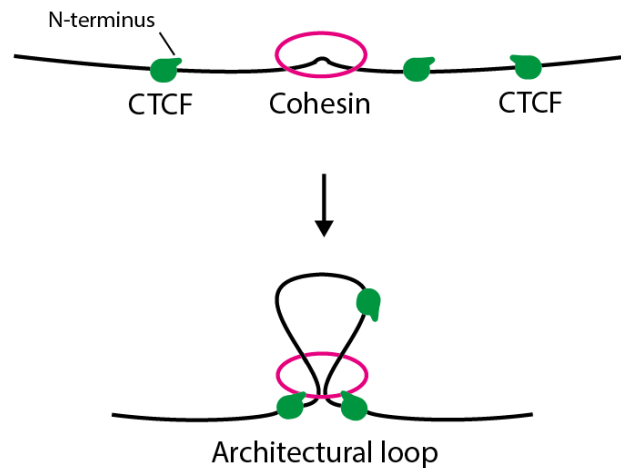
topologically associating domains (TADs), where DNA would preferentially interact within the same TADs, but not across different ones (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). Intriguingly, CTCF-bound sites are enriched at the TAD boundaries, or the margins of TADs. It has been shown that deletion or disruption of such sites would dramatically alter the TAD structures (Narendra et al., 2015; Nora et al., 2012; Sanborn et al., 2015), whereas acute degradation of CTCF would lead to complete abolishment of TADs (Busslinger et al., 2017; Nora et al., 2017). This indicates that CTCF is essential in the formation of TADs.

By using Hi-C data and mathematical modelling, it was recently shown that CTCF works together with cohesin as architectural proteins and thus have a direct role in the packaging of TADs, via a mechanism known as loop-extrusion (Fig. 1.1, 1.2A) (Davidson et al., 2019; Fudenberg et al., 2016; Kim et al., 2019b; Sanborn et al., 2015). In this model, cohesin, which is a protein complex that forms a ring-like structure, randomly loads onto the chromatin and the DNA is extruded through. The extrusion will continue to progress until the cohesin encounters, on both sides of DNA, a pair of CTCF molecules being bound at sites that are in a convergent orientation (i.e. 5'-forward-reverse-3'). At this stage, cohesin will come to a halt and a stable loop is thus formed. The orientation of the CTCF motif is very critical as inversion of the orientation of one of the CTCF anchors disrupts the existing loop, resulting in the formation of an ectopic loop (de Wit et al., 2015; Guo et al., 2015; Sanborn et al., 2015). The molecular mechanism of this polarity is very enigmatic: how can cohesin skip past CTCF at one orientation while being retained at the other? Two very recent studies involving crystallography and CTCF mutants have shed light on this phenomenon (Li et al., 2020; Nora et al., 2019). The N-terminus of CTCF, which is not involved in DNA binding, was shown to interact with cohesin and thus is crucial in holding cohesin in position. It was found that CTCF lacking the N-terminus fails at cohesin retention and leads to improper TAD formation. In addition, the position of N-terminus in the CTCF molecule is also critical as swapping of N- and C-terminus did not rescue CTCF from retention failure. Given the fact that in a convergent setting, loop extruding cohesin would first run head-on into the N-termini of CTCF, it has been proposed

that this engagement allows CTCF to be at a correct position to ensnare cohesin and stabilise the loop.

Overall, this new data reconciles with the original understanding of an insulator protein. CTCF can prevent crosstalk between promoter and enhancer by partitioning each cis-element into two separate TADs. Therefore, TAD boundaries play a crucial role in regulating gene expression by facilitating appropriate enhancer-promoter interactions, while insulating against aberrant ones. For instance, several recent reports have demonstrated that disruption of TADs can result in the formation of inappropriate enhancer-promoter loops, leading to gene deregulation and diseases (Flavahan et al., 2016; Lupianez et al., 2015). Albeit important, the disruption of TADs does not always lead to absolute and global gene deregulation – majority of gene transcription does persist upon the degradation of CTCF and cohesin (Nora et al., 2017; Rao et al., 2017; Stik et al., 2020) or removal of CTCF sites (Despang et al., 2019). This indicates there are other factors regulating enhancer-promoter contact hence gene transcription, and one of such is phase separation.

A)



B)

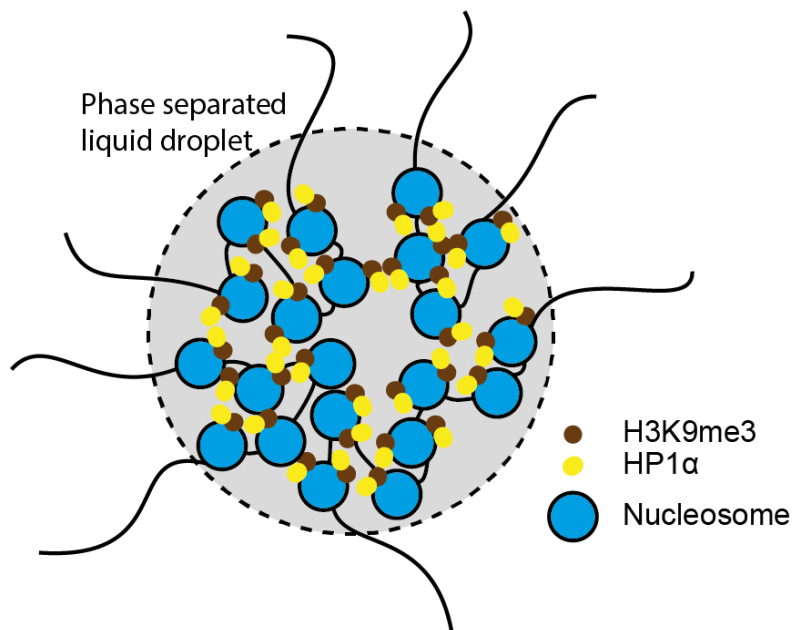


Figure 1.2. Two major molecular mechanisms of genome organisation.

A) Loop extrusion model proposes that the cohesin in a ring-like structure would load onto chromatin in a random manner, with chromatin subsequently traverse through the cohesin complex and extrudes along. Such loop extrusion in both directions will continue to progress until the cohesin complex encounters head-on to the N-terminus of CTCF. B) Compartments are suggested to develop through a physical process known as phase separation. For instance, specific histone modification such as H3K9me3, with its associating chromatin-binding proteins HP1 α , due to the unique physical properties they possess, would tend to aggregate and eventually form a phase-separated environment.

Phase separation and compartments

At a broader scale, chromatin that are of similar epigenetic state would tend to agglomerate, forming compartments (Lieberman-Aiden et al., 2009). Initial Hi-C studies described that the mammalian genome is partitioned into two compartments, A and B, whereby crosstalk among the same compartment is common but interactions between are at minimum. In general, compartment A tend to contain euchromatic regions which harbour active genes, whereas compartment B contrarily associates with inactive heterochromatin. Later advanced Hi-C assays with finer resolution further divided the genome into six different sub-compartments with unique histone modifications and epigenetic landscapes (Rao et al., 2014). It is thus suggested that the formation of compartments is facilitated by the similar physical affinities of certain epigenetic states of chromatin, potentially through phase separation, where an immiscible liquid medium, or phase, is isolated or created from the rest of the nucleoplasm (Falk et al., 2019; Hnisz et al., 2017; Lesne et al., 2019; Sabari et al., 2018; Strom et al., 2017). Supporting this, a number of studies have shown that HP1 α and other complexes recognising the heterochromatic H3K9me3 modification are able to aggregate and lead to the formation of phase-separated liquid droplets, with macromolecules such as DNA being retained in such droplets (Falk et al., 2019; Larson et al., 2017; Sanulli et al., 2019; Wang et al., 2019). This thus highlights the role of histone marks in regulating compartment formation. Intriguingly, other studies have demonstrated that the A- and B-compartments remain totally unaffected upon acute CTCF or cohesin degradation, which would otherwise completely abolish TADs (Busslinger et al., 2017; Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017). This indicates that the establishment of compartments is entirely independent of CTCF and the associating mechanism of loop-extrusion.

In general, TADs are suggested to be fairly conserved between different cell types and even across different vertebrate species, whereas compartments and loops display a higher degree of lineage-specificity (Dixon et al., 2015; Dixon et al., 2012; Lieberman-Aiden et al., 2009; Rao et al., 2014). When stimulated or during differentiation genes are able to switch between compartments and such switch generally involves the flipping of TADs as a whole unit from A- to B-

compartment or vice versa. A to B switch of compartment has been shown to almost always associate with silencing of transcription (Miura et al., 2019). However, B to A compartmental switch does not necessarily associate with transcriptional upregulation, as it conceivably provides an environment that is permissive to looping (Buecker et al., 2014; Miura et al., 2019), which in turn relies on the presence of appropriate transcriptional activators. Thus enhancer- and silencer-promoter contacts are presumably very lineage-specific and are believed to be the most critical factors in regulating cell type and developmental stage specific transcription.

Nuclear space and chromosome territories

During mitosis all transcription ceases and chromatin compacts into highly condensed chromosomes resulting in the temporary loss of loops, TADs and compartments (Naumova et al., 2013). Only shortly after cell division do these 3D structures re-emerge again. The re-establishment of structures are confined to their own chromosomes (Fig. 1.3), which individually occupy a discrete and isolated region in the three-dimensional nuclear space, instead of scattering across the entire nucleus (Cremer and Cremer, 2001). This distinct localisation in the nucleus is then referred as chromosome territory. Intermingling of chromosome territories is very infrequent but does occur, which is implicated in rare translocation events typically leading to cancers (Branco and Pombo, 2006; Roix et al., 2003). The maintenance of chromosome territory is therefore important to safeguard genome stability and functions, yet very little is known about the mechanism. Nonetheless, lamin, a critical component of the nuclear envelope, appears to be one of the factors in upholding the chromosome territory – depletion of lamin A in nucleus renders a much more dynamic chromatin movement (Bronshtein et al., 2016).

Nuclear envelope and lamina associated domain

In addition to securing chromosome territories, the nuclear envelope is also directly implicated in gene regulation. It has long been known that the location of genes in the 3D nuclear space could influence their transcription activity, in which active genes normally reside at the nuclear interior, whereas nuclear periphery is

generally associated with gene silencing (Dechat et al., 2010). At periphery, lamins coat the inner layer of nuclear envelope, forming a meshwork of nuclear lamina (Fig. 1.3). Nuclear lamina is adhered by a considerable amount of chromatin, constituting the lamina-associated domains (LADs), which was found to be enriched in heterochromatin bearing the repressive marks H3K9me2 and H3K9me3 (van Steensel and Belmont, 2017). These histone marks are crucial for the attachment as the disruption of G9a and Suv39H1, responsible for the H3K9me2 and H3K9me3 deposition respectively, resulted in dissociation of LADs from the nuclear lamina (Bian et al., 2013; Chen et al., 2014a; Harr et al., 2015; Kind et al., 2013).

Nuclear pores – an exception to peripheral silencing

Though nuclear periphery normally presents a repressive environment, nuclear pores appear to be an exception where gene activation is favoured (Fig. 1.3) (Blobel, 2010; Buchwalter et al., 2019). Early cytological observations documented chromatin de-condensation around nuclear pores, suggesting a different role in gene regulation as compared to the nuclear lamina (Blobel, 1985; Watson, 1955). Nuclear pore complex (NPC) was later revealed to interact with chromatin and exert a gene regulating function (Ptak and Wozniak, 2016; Sood and Brickner, 2014). For example, NPC was demonstrated to facilitate gene activation by mediating long-range enhancer-promoter looping (Ibarra et al., 2016; Pascual-Garcia et al., 2017). Additionally, the aforementioned 5'-3' gene loop was shown to form and maintain via NPC tethering (Tan-Wong et al., 2009). In plasma cells, the highly active immunoglobulin (Ig) genes are localised at periphery, presumably near the nuclear pores for maximal transcript export (Park et al., 2014). Altogether, this illustrates the unique role of nuclear pores in regulating genome architecture and gene expression.

Chromosomal territories in nucleus

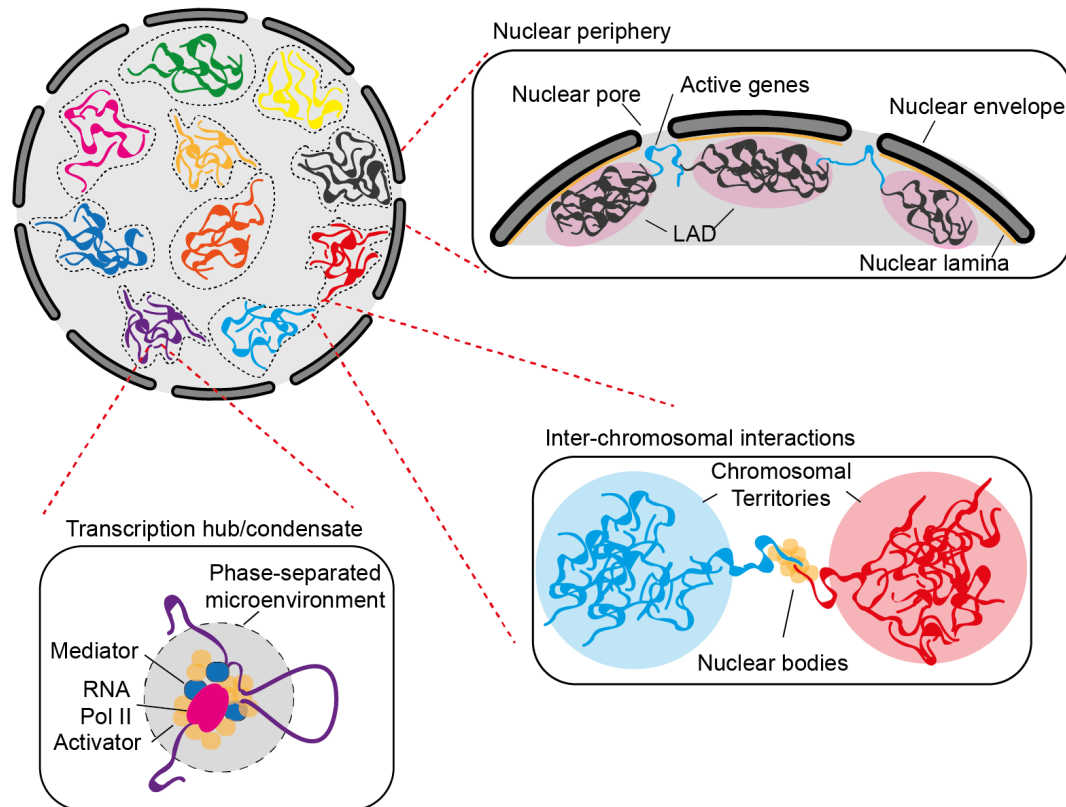


Figure 1.3. Nuclear space in interphase nucleus.

Interphase chromosomes in nucleus all occupy their own chromosomal territories. At the periphery heterochromatin associates with nuclear lamina, forming the lamina-associated domain (LAD), whereas active genes are located near nuclear pores. Inter-chromosomal interactions can form when chromatin extends beyond the territories, such interactions are often found associating with nuclear bodies such as nuclear speckles and nucleolus. Bearing the Intrinsically Disordered Region (IDR), transcription activators, Mediators and RNA polymerase II are able to interact and form phase-separated, liquid-like droplet. These are often referred as transcription condensates, transcription hubs or transcription factories.

Nuclear bodies, transcription hubs and phase separation

Inside the nucleus, there are other landmarks that are associated with a distinct gene regulatory function, and intriguingly, involve inter-chromosomal interactions (Spector, 1993). The nucleolus and nuclear speckles are nuclear bodies that are prominent enough to be observed and functionally characterised in early days through conventional microscopy (Fig. 1.3). While nucleoli are specialised in ribosomal RNA transcription and ribosome assembly, nuclear speckles are responsible for mRNA processing and splicing. Phase separation is believed to drive the formation of these nuclear bodies (Hnisz et al., 2017; Lesne et al., 2019; Palikyras and Papanonis, 2019). Being membrane-less organelles, they can withhold and concentrate certain protein factors for the required biochemical reactions, and undergo fission and fusion, which are all properties of phase-separated liquid droplets. More importantly, some of the protein factors within these organelles contain an Intrinsically Disordered Region (IDR), a peptide segment of irregular secondary structure, which is indispensable in the formation of liquid droplets through multivalent hydrophobic interactions (Li et al., 2012b). Attributed to the coalescing capability of liquid droplets, nucleoli and speckles are able to recruit genes from different chromosomes, forming the inter-chromosomal hubs (Quinodoz et al., 2018). Extending beyond their own chromosome territories, these genes colocalise into the nuclear bodies as to coordinate and facilitate the rRNA transcription and the post-transcriptional processing of mRNA.

Similarly, phase separation assists the assembly of transcription factories and other foci. Rather than diffusely occurring across the nucleus, Pol II transcription surprisingly manifests in discrete and solitary clusters (Fig. 1.3) (Buckley and Lis, 2014). Pol II, and other general transcription factors such as Mediators and BRD4, contain IDR and are thereby capable of forming liquid droplets (Boehning et al., 2018; Cho et al., 2018; Sabari et al., 2018). It was further demonstrated that the activation domains of other transcription factors are also IDR *per se*, which are able to interact with Mediators and Pol II (Boija et al., 2018; Chong et al., 2018). Consequently, the aggregation between transcription factors, Mediators and Pol II can draw enhancers from afar to a close proximity to promoters, with the interactions accommodated in a liquid-like condensate (Boija et al., 2018; Chong et al., 2018; Sabari et al., 2018; Shrinivas et al., 2019). The assemblage is even

more prominent in super-enhancers, being clusters of enhancers arranged at a high density along genetic loci, they can recruit a substantial amount of transcription factors and thereby conglomerate into a stable, huge and fully phase-separated liquid droplet. Interestingly, interactions associated with certain super-enhancers persist and even strengthen in the absence of cohesin and loop extrusion, further suggesting phase separation as the primary driving force (Rao et al., 2017; Schwarzer et al., 2017). Moreover, this suggests phase separation, in addition to loop extrusion, can also mediate local enhancer-promoter interactions. In contrast to loop extrusion, which can only execute along a single chromosome, phase separation has the potential to recruit genes from different chromosomes into a single transcription hub. The cohesin-independent super-enhancers interactions were shown to harbour some inter-chromosomal contacts (Rao et al., 2017; Schwarzer et al., 2017). Although not been completely elucidated, the inter-chromosomal interactions in the olfactory receptors were implicated to be orchestrated by phase separation as well (Bashkirova and Lomvardas, 2019).

Loop extrusion vs phase separation

All in all, loop extrusion and phase separation represent two independent mechanisms in shaping genome architecture. Integrating both mechanisms by mathematical polymer simulation revealed competition between the two processes with extrusion of cohesin overriding compartmental segregation (Nuebler et al., 2018). Manifesting in Hi-C data, the CTCF-mediated TAD would then mask over the compartments mediated by phase-separation. This interference therefore argues against the conventional claim about the organisational hierarchy where compartments are larger structures holding interacting TADs. In lieu of that, it suggests an intricate interplay where loop extrusion modulates the local basal architecture established by phase-separating epigenetic states. Coherent with this, the mammalian genome, upon cohesin removal, was revealed to possess more finely segregated compartments that was once merged as a single TAD (Rao et al., 2017; Schwarzer et al., 2017). Additionally, genome of *Drosophila*, which lacks the loop extrusion mechanism, consists of fine compartments that are of comparable size to TADs in mammals

(Rowley et al., 2017). As such, it appears that loop extrusion was a more recent evolutionary innovation. Accordingly, it would be interesting to speculate that perhaps loop extrusion could be a contributing factor towards the increasing complexity of higher organisms, as it allows a local adjustment of structure and the establishment of more refined point-to-point interactions. One revolutionary consequence of this could be the programmed DNA rearrangement – the V(D)J recombination and class switch recombination (CSR) – in vertebrate lymphocytes, where Ig genes are rearranged to diversify the antibody specificities and functions (Hozumi and Tonegawa, 1976; Muramatsu et al., 2000). It was recently demonstrated that both V(D)J recombination and CSR are mediated via CTCF and the loop extrusion process (Zhang et al., 2019a; Zhang et al., 2019b).

Non-coding RNAs in chromatin biology

While 98% of the human genome does not encode for proteins, an even more unexpected discovery was that transcription is pervasive and probably occurs along more than 90% of the human genome (Consortium et al., 2007; Djebali et al., 2012; Wilusz et al., 2009). Following this discovery it was speculated that the majority of these non-coding RNAs are simply transcriptional noise without any functions (Ebisuya et al., 2008; Struhl, 2007; Wang et al., 2004). However, the fact that many well-characterised, essential components of cellular processes – such as transfer RNAs and ribosomal RNAs – are non-coding transcripts, it was suggested that it was very likely that other non-coding RNAs are of biological relevance but yet to be characterised (Mattick, 2004). Indeed, recent studies have uncovered a myriad of other non-coding RNA classes which serve regulatory functions, for instance small interfering RNA (siRNA), microRNA (miRNA), Piwi-interacting RNA (piRNA), circular RNA (circRNA) and long non-coding RNA (lncRNA) (Costa, 2010; Salzman, 2016; Wilusz et al., 2009). Some of these types such as siRNA and miRNA are functionally and mechanistically distinct. In contrast, lncRNAs arbitrarily represent any non-coding transcripts that are over 200 nt in length, with functions of many of which remain largely unknown (Wilusz et al., 2009). Nonetheless, their spatiotemporal and cell-type specific expression pattern suggests a regulatory potential, and henceforth a growing number of lncRNAs have recently been functionally characterised. It appears that lncRNAs

regulate transcription in a plethora of diverse ways, among which includes altering chromatin structures and loop formation.

Alterations of chromatin landscape by non-coding RNAs

LncRNAs have been shown to directly recruit chromatin modifiers, either in cis or in trans (Fig. 1.4 Top). In human a lncRNA *HOTTIP* would remain in cis and recruit WDR5, which helps lay down H3K4me3 marks at the promoters of its flanking genes, and thereby drive gene expression (Wang et al., 2011). LncRNA *Air* and *Kcnq1ot1*, critical players during genomic imprinting, can interact with G9a and Polycomb complex to enforce epigenetic silencing in the paternal allele (Nagano et al., 2008; Pandey et al., 2008). In general, it is thought that the cis-acting behaviour is achieved by tethering, in which the nascent transcript is still attached to the elongating Pol II while delivering the chromatin modifiers (Guttman and Rinn, 2012; Magistri et al., 2012). On the other hand, a trans-acting lncRNA is exemplified by *HOTAIR*, which relocates to another chromosome and recruits PRC2, resulting in the laying down of H3K27me3 to repress the target locus (Rinn et al., 2007).

While some of the lncRNAs are important in recruiting histones modifiers, in other cases the act of transcription *per se* appears to be more essential than the RNA transcript that is produced (Kornienko et al., 2013). The unwinding of DNA and creation of transcription bubble were shown to directly displace nucleosomes (Fig. 1.4 Top) (Hughes et al., 2012; Valouev et al., 2011; Weiner et al., 2010). As a result, transcription elongation can create a different nucleosome landscape along the chromatin fibre. For example, lncRNA transcription of *SRG1* in yeast can increase nucleosome density at the promoter of target gene, and subsequently achieve gene silencing (Martens et al., 2004). Moreover, different modulators such as histone acetyltransferases, histone methyltransferases and ATP-dependent chromatin remodelling complexes have been shown to interact with elongating Pol II (Ng et al., 2003; Wilson et al., 1996). Accordingly, progression of Pol II is capable of modulating histones and hence creating a permissive or repressive environment (Fig. 1.4 Top). For example, the act of transcription has been implicated in recruiting histone deacetylases, which results

in the suppression of transcription initiation in the overlapping coding genes (Camblong et al., 2007). In some cases, the non-coding transcription can even induce the formation of heterochromatin (Yu et al., 2008). Contrarily, in Hox clusters lncRNA transcription was suggested to prevent certain Hox genes from silencing by Polycomb complex (Rank et al., 2002; Schmitt et al., 2005). In other cases, continuous Pol II progression has been implicated in keeping the DNA region open and accessible (Corcoran, 2010; Yewdell and Chaudhuri, 2017). Of notable significance is during V(D)J recombination and CSR, where non-coding transcription prevails to control regional accessibilities of the Ig locus, which will be further discussed in the later sections.

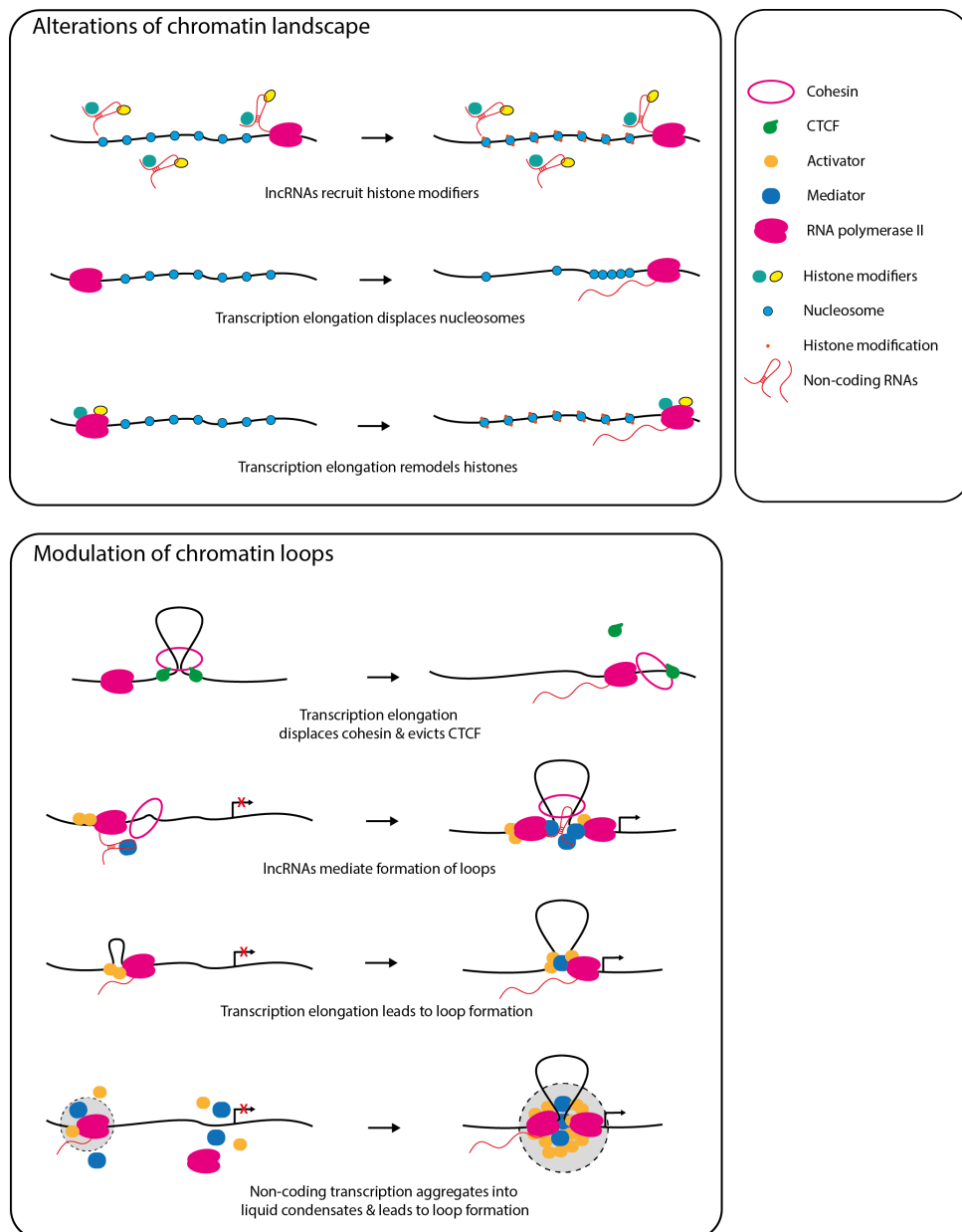


Figure 1.4. Mechanisms of non-coding RNA in remodelling 3D genome.

Top) Non-coding RNA can alter local chromatin environment. Transcripts can recruit histone modifiers to deposit histone marks in cis or trans. Transcription elongation can displace nucleosomes or recruit histone modifiers to alter the local chromatin landscape. Bottom) Non-coding RNA can modulate chromatin loops. Transcription elongation is capable of displacing cohesin and evicting CTCF. RNA transcripts can interact with Mediator and hence aid in loop formation. RNA polymerases initiate transcription from enhancers and remain bound to, and therefore escort, enhancers towards promoters. Enhancer transcription aggregates into transcription condensates and mediate long-range interaction through further coalescence.

Non-coding RNAs in modulating genome conformation

The process of transcription of non-coding RNAs can also affect chromatin structure three-dimensionally. Recent high-throughput sequencing has unveiled widespread bidirectional non-coding transcription at enhancers (Kim et al., 2010). These enhancer RNAs (eRNAs) are largely non-polyadenylated and therefore deemed unstable, unlike most of the other lncRNAs. Yet, expression level of these eRNAs are very well correlated with the genes that the enhancers regulate. This has led to the speculations that eRNAs are crucial transcription regulators and, apart from the aforementioned mechanisms, are perhaps able to regulate three-dimensional looping (Lam et al., 2014; Li et al., 2016; Meng and Bartholomew, 2018). Indeed, Pol II has been shown to evict CTCF and displace cohesin, thereby interfering with loop extrusion (Fig. 1.4 Bottom) (Busslinger et al., 2017; Heinz et al., 2018; Lefevre et al., 2008; Lengronne et al., 2004). As such, transcription elongation thus appears as an additional layer in shaping genome conformation. In contrast, although the transcripts *per se* were thought to be unstable and prone to degradation, in some cases eRNA transcripts were suggested to be critical. Enhancer RNAs have been implicated in physically interacting with cohesin and Mediator complex (Fig. 1.4 Bottom), in which both are important architectural proteins, and assist in loop formation (Hsieh et al., 2014; Lai et al., 2015; Lai et al., 2013; Li et al., 2013b). Additionally, while the identities between eRNAs and other lncRNAs are somewhat blurred and partially overlapped, a number of lncRNAs were subsequently shown to associate with chromosomal looping, including *HOTTIP* (Isoda et al., 2017; Trimarchi et al., 2014; Wang et al., 2011; Xiang et al., 2014). However, the causal relationship between transcription and looping is, if not obscure, context dependent. Transcription of enhancers or eRNAs *per se* can indeed cause loop formation, yet in other instances transcription can be inconsequential to, or entirely independent of, chromosomal looping.

Apart from interfering with loop extrusion or being actively involved in looping, eRNAs can mediate enhancer-promoter interactions in two other ways. In one of the models, Pol II emanating from enhancer is suggested to track towards promoter (Fig. 1.4 Bottom), escorting the enhancer along the way until reaching the destination where a stable loop is formed (Blackwood and Kadonaga, 1998;

Furlong and Levine, 2018). The enhancer-promoter interaction within the β -globin locus was proposed to arise in this manner (Kong et al., 1997). Secondly, transcription at enhancers has been implicated in forming transcription condensates or relocation to transcription factories (Fig. 1.4 Bottom) (Furlong and Levine, 2018). The transcription of enhancer RNAs can emerge as a focal point, then aggregate with Mediator or other protein complexes, and/or coalesce with other existing transcription condensates, thereby bridging promoter and enhancer within the same droplet-like microenvironment. Under such circumstances, transcription could potentially lead to repositioning of DNA to a different compartment. In line with this, global transcription readthrough past the polyadenylation signals coerced by influenza virus NS1 protein has resulted in an extensive A/B compartment switch (Heinz et al., 2018). Likewise, during T cell development the transcription of the non-coding RNA *ThymoD* has been demonstrated to relocate the *Bcl11b* locus away from the nuclear lamina to a more permissive environment at the nuclear centre (Isoda et al., 2017). Overall, there is an intricate interplay between transcription and chromatin conformation – while 3D structures have an immense influence on gene regulation, transcription, on the other hand, is able to feedback into the system to remodel genome architecture.

Genome architecture during cell cycle

The various chromatin structures described thus far are features in the interphase nucleus. During mitosis, however, due to condensation and compaction of chromatin, 3D structures such as loops, TADs and compartments are lost, along with the stalling of transcription (Gibcus et al., 2018; Nagano et al., 2017; Naumova et al., 2013). Intriguingly, upon mitotic exit the precedent chromatin structures are faithfully restored. As such, the genome architecture as well as transcription programme are inherited through successive cell division. This inheritance is likely contributed by an ill-defined process known as mitotic bookmarking (Oomen and Dekker, 2017), in which adherence of various factors – histone modifications, histone variants, methylated CpG motifs, non-coding RNAs and transcription factors – to the mitotic chromosome is believed to play an important role. On the other hand, differentiation contrasts with normal cellular

turnover as reorganisation of chromatin conformation takes place. While still unclear, the re-emergence of 3D structures upon mitotic exit (Gollosi et al., 2017; Oomen and Dekker, 2017), or during DNA replication where chromatin is unwound (Bird et al., 1998; Richter et al., 1999; Weintraub et al., 1978; Wolffe, 1991), are thought to represent times whereupon genome is reorganised to support cellular differentiation.

1.2. Methods of investigating genome architecture

Microscopy-based

In early days the study of chromatin conformation was predominantly based on microscopy. Fluorescent *in situ* hybridisation (FISH) exploits a stretch of chemically synthesised nucleotide with fluorescent tags, referred as probes, to hybridise and visually label the region of interest in the nuclear space (Rudkin and Stollar, 1977). With multiple probes, the proximity between regions can then be inspected with respect to time and different perturbations. In other cases, it can be used to label, or paint, the entire chromosome (Pinkel et al., 1988). These approaches were utilised in the initial descriptions of chromosome territories and inter-chromosomal interactions (Branco and Pombo, 2006; Lichter et al., 1988; Pinkel et al., 1988).

With recent advances in probe design and more cost-efficient probe synthesis, novel approaches such as oligopaints (Beliveau et al., 2012), sequential FISH (Bolzer et al., 2005; Schrock et al., 1996) and signal amplification by exchange reaction (SABER)-FISH (Kishi et al., 2019) allow the development of high-coverage, high-resolution and high-throughput visualisation of chromatin in the nucleus. Based on these approaches, several methods such as Hi-M (Cardozo Gizzi et al., 2019) and optical reconstruction of chromatin architecture (ORCA) (Mateo et al., 2019) were developed to reconstruct the entire genome architecture in a cell. The resolution can be further improved by super-resolution microscopic techniques such as photo-activated localisation microscopy (PALM) (Betzig et al., 2006; Hess et al., 2006) and stochastic optical reconstruction microscopy (STORM) (Rust et al., 2006), which allow the visualisation and tracking of RNA pol II condensates (Cho et al., 2016; Cho et al., 2018) and cohesin complex during loop extrusion process (Hansen et al., 2017; Kim et al., 2019b).

Proximity ligation – chromosome conformation capture

Recently a molecular approach was embarked upon to explore genome architecture in a more quantitative manner. Broadly categorised as chromosome conformation capture (3C), 3C and its variants are in principle a proximity ligation

assay of a fixed genome, which exploits the tendency of DNA fragments to be more efficiently ligated to others in close physical proximity than those located far away (Dekker et al., 2002). In brief, during a 3C reaction the chromatin of a population of cells is chemically cross-linked to capture, or preserve, the relative position of all DNA and protein complexes in three dimensions (Dekker et al., 2002). The fixed genome is thereafter subjected to restriction digestion as to release millions of free DNA ends. Subsequently proximity ligation is performed on the fragmented genome, with the interacting DNA segments covalently joined into a continuous chain. Since the ligation efficiency is inversely proportional to the physical distance, the proximity of any two DNA regions can then be evaluated by amplifying the interaction of interest via PCR. However, the pairwise inspection (one-to-one) reveals limited information about the global interaction landscape, for which a number of later technical extensions from 3C were designed to address. The first extension is chromosome conformation capture-on-chip (4C), which utilises inverse PCR to amplify all the interacting fragments from a target region (one-to-all) (Simonis et al., 2006; Zhao et al., 2006). Herein the ligated 3C library is subjected to a second round of digestion and ligation to generate self-ligated circular DNA fragments. Inverse PCR with a pair of outward facing primers from the region of interest would then capture all the interacting DNA sequences. Another derivative, chromosome conformation capture carbon copy (5C), relies on a huge set of specific primers pre-designed to target multiple sites of interest that are immediately adjacent to the restriction cut site in the genome (Dostie et al., 2006). These primers can then potentially settle right next to each other on the 3C library, which are then ligated together by *Taq* ligase. With a universal sequence overhang at the 5' end of those primers, one can then apply microarray or sequencing to reveal all the interacting pairs (many-to-many). However, this approach is unable to cover the whole genome due to the exhaustive primer design.

High throughput chromosome conformation capture – Hi-C

The highest throughput of all couples chromosome conformation capture with next generation sequencing (Hi-C), which allows a global assessment of all the interacting DNA pairs in the genome (all-to-all, Fig. 1.5) (Lieberman-Aiden et al.,

2009). Initial version of that was known as dilution Hi-C, which accomplishes the proximity ligation in solution after nucleus lysis. Due to the spurious contacts and random ligation in solution, the background noise is high and thus requires more depth in sequencing reads (Nagano et al., 2015). A modified approach, referred as *in situ* Hi-C, was later developed to reduce the noise (Rao et al., 2014). Instead of in solution, proximity ligation is conducted in the intact nucleus, which limits random contacts and therefore the resolution of the data is greatly improved. In both approaches the ligated 3C library is randomly sheared into smaller fragments, with a universal DNA adapter subsequently ligated to both ends. Paired-end sequencing is then performed to read out all the interacting DNA regions. The entire genome architecture can then be inferred by determining all the relative interaction frequencies of any two DNA regions, which are ultimately displayed in a contact matrix, or Hi-C map.

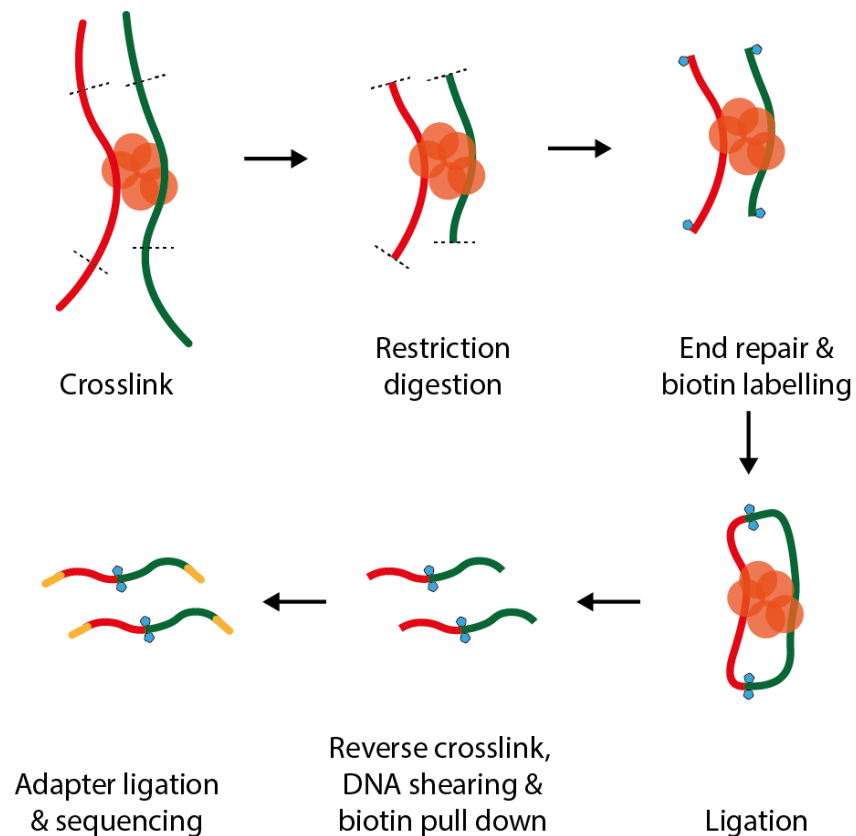


Figure 1.5. Methodology of high throughput chromosome conformation capture (Hi-C).

Cells are crosslinked with formaldehyde, which preserves, or captures the cellular intrinsic DNA-DNA, protein-protein, or protein-DNA interactions. Chromatin is then digested with restriction enzyme to liberate free DNA ends. The sticky ends are then filled in with biotinylated nucleotides, with the resulting blunt ends subjected to proximity ligation. Upon ligation the crosslinks are reversed and then DNA is sheared by sonication. The hybrid, or chimeric DNA fragments, bearing biotin at the junction, are then isolated and purified. After end repair of the sheared DNA ends, sequencing adapters are ligated to the chimeric fragments. Such libraries are then amplified by PCR and subjected to paired-end sequencing.

Manifestation of chromatin structures in Hi-C data

In Hi-C maps, TADs manifest as triangles along the diagonal with intense contact frequency (Fig. 1.6A). Initial description of TADs estimated a median size of 880 kb, which cover 90% of the genome (Dixon et al., 2012; Nora et al., 2012). At a larger scale, compartments manifest as checkerboard patterns scattered across the Hi-C contact matrices (Fig. 1.6A), with alternating A and B compartments (Lieberman-Aiden et al., 2009). From the initial experiments, compartments were described as of mega-base in size and it was then inferred that compartments are composed of multiple interacting TADs with discrete hierarchy. However, with the improvement of resolution, many smaller TADs were uncovered, the median size of TADs was then adjusted to be around 185 kb, ranging from 40 kb to 3 Mb (Rao et al., 2014). Very often several TADs are nested within a TAD and hence designations such as sub-TAD or even sub-sub-TAD emerged (van Steensel and Furlong, 2019; Xu and Corces, 2016). As such, these designations, which are largely based on size, are somewhat arbitrary and confer limited biological meaning – as reflected in a recent mockery where TAD was instead dubbed by Corces (2019) as solely “Triangles At Diagonal”. Also, the assumed hierarchy between TADs and compartments has been brought into question (Nuebler et al., 2018). As mentioned in the earlier section, recent studies have implicated loop extrusion in mediating the formation of TADs, whereas phase separation by various epigenetic states and transcription condensates helps the agglomeration of compartments. Across the genome, depending on cell types, CTCF sites are located at around 75-95% of TAD boundaries (Bonev et al., 2017; Dixon et al., 2012). They are able to engage in persistent loop formation when they are of convergent orientation, which manifests as a sharp dot, or corner dot (Fig. 1.6B), at the tip of the square or triangle on Hi-C maps (Beagan and Phillips-Cremins, 2020; Szabo et al., 2019). This corner dot therefore appears to be a result of loop extrusion. However, around 20% of boundaries are resistant to CTCF removal (Nora et al., 2017), and these boundaries coincide with transcription condensates or demarcation between A and B compartment (Bonev et al., 2017; Dixon et al., 2012; Rao et al., 2014; Rowley et al., 2017). In addition, corner dots are always absent in CTCF-independent domains (Rowley et al., 2017). As suggested before, phase separation, in addition to aggregation, is also able to segregate chromatin into small domains of comparable size to TADs. Altogether, these CTCF-

independent domains are therefore probably of the same nature as compartments. For this purpose, it has been suggested that it would be beneficial to allocate TADs that are mediated by loop extrusion as a “loop domain”, whereas those that are caused by phase separation designated as “compartmental domains” or simply “compartments” (Beagan and Phillips-Cremins, 2020; Rowley et al., 2017). Such a consensus has yet to be reached. Nonetheless, differentiating the interacting domains with respect to the underlying molecular mechanisms may make the further functional perturbation studies of these regions easier.

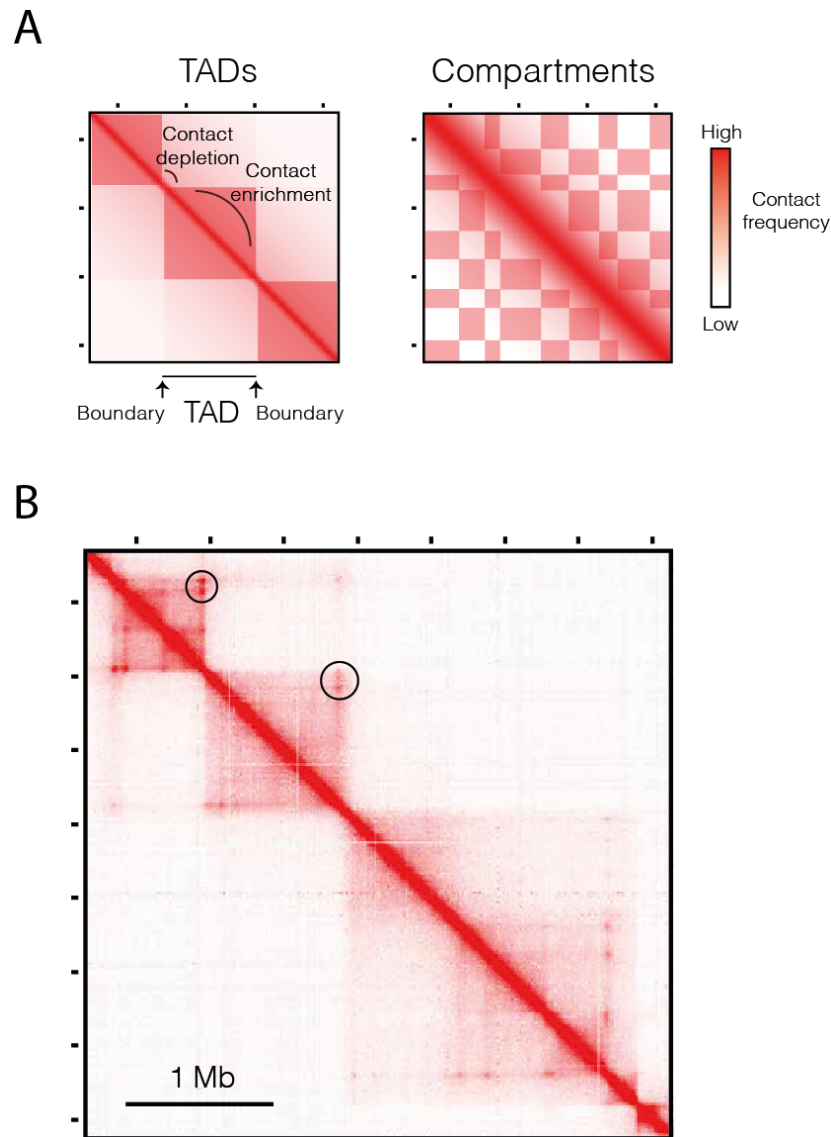


Figure 1.6. Manifestation of different genome structures in Hi-C contact matrices.

A) Schematics of Hi-C matrices at different genomic scales, revealing different levels of 3D genome structures. At the sub-megabase scale, TADs manifest as triangles along the diagonal with intense interaction frequency. At a larger scale, compartments manifest as checkerboard patterns scattered across the contact matrices, representing the A- and B-compartments. B) An example of mammalian Hi-C contact matrices. Featured in some of the TADs are the presence of corner dots at the tip of the triangle (circled in black), indicating an intense interaction frequency and thus the presence of stable, persistent chromatin loops. (Adopted from Szabo et al., 2019)

Improved resolution C-technology – micro-C

While *in situ* Hi-C can capture broad chromatin structures such as TADs and compartments, with a resolution of around 20-50 kb bin size, it fails to describe details of specific local structures, for instance gene loops and other interactions at the nucleosomes level. A Hi-C variant, micro-C, was recently developed to provide additional information below the scale of TADs (Hsieh et al., 2015; Hsieh et al., 2020; Hsieh et al., 2016; Krietenstein et al., 2020). In Hi-C restriction enzymes are used generate fragmented genome with uneven lengths, and proximity ligation is therefore biased towards regions near these restriction sites. To deal with this, micro-C utilises Micrococcal nuclease (MNase) to unleash free DNA ends at almost every nucleosome, with the following proximity ligation step capturing all the interacting nucleosomes. With this, more chromatin loops are detected and 3D genome conformation can be resolved down to 200 bp (Hsieh et al., 2015; Hsieh et al., 2020; Hsieh et al., 2016; Krietenstein et al., 2020).

Ligation independent method – genome architecture mapping

All the C-technologies require ligation of two interacting DNAs therefore any information regarding other potential interacting partners will then be lost upon ligation. As such, three-way or multi-way interactions are inefficiently captured and therefore problematic to quantify. A novel, ligation independent method, genome architecture mapping (GAM), was developed to measure all the contacts between any DNA regions simultaneously in a genome wide manner (Beagrie et al., 2017). GAM involves utilisation of laser microdissection to take nuclear profiles from ultrathin cryosections that have been performed in fixed cells at random 3D orientations, which is followed by DNA extraction and sequencing for each slice of nuclear profile. Through assessing thousands of nuclear profiles, DNA regions of close physical proximity in the 3D nuclear space would then be sampled together in the same nuclear profile more frequently than those afar. Thus, by evaluating the cooccurrence frequency of all genomic regions in ultrathin slices of thousands of nuclei, genome-wide chromatin interaction profile and hence the 3D conformation can then be constructed unbiasedly. In addition, GAM can provide extra information about the spatial organisation of chromatin which is lacking in Hi-C data. Most importantly, GAM requires only few thousands

of cells, which is much less than other C-technologies and therefore favourable in studying rare cell types.

Multi-way assessment of DNA interactions – SPRITE

Another novel technology, SPRITE (split-pool recognition of interactions by tag extension), exploits DNA barcoding to simultaneously assess multi-way and genome wide DNA interactions (Quinodoz et al., 2018). Instead of proximity ligation, the cross-linked and fragmented DNA-protein complexes are subjected to splitting of complexes into individual wells, followed by tagging of mixture in each well with specific DNA tags. The content across wells are then pooled together and successive rounds of such splitting, tagging and pooling are performed to create a unique DNA barcode in each individual DNA complex. Because interacting DNAs are cross-linked in the same complex, they will be dispatched together in every round of splitting and thus acquire an identical barcode. In contrast, DNAs in separate complexes will be sorted independently in every round of splitting, they will therefore depart at some point and receive different barcodes at the end. After this the genomic DNA and barcodes are then sequenced and processed to determine DNA interactions and hence chromatin conformation. Apart from increasing the detection of multiway interactions, SPRITE can identify inter-chromosomal contacts in nuclear bodies in which Hi-C fails to identify. As the efficiency of proximity ligation is inversely proportional to the physical distance, DNA regions that are complexed around large nuclear bodies are located too far apart to be captured by Hi-C. By comparison, without a ligation steps, SPRITE can efficiently detect the inter-chromosomal interactions in large nuclear bodies such as nucleolus and nuclear speckles.

1.3. Methods to manipulate the genome

Genome editing

Deciphering the functions of protein-coding genes has been made possible by traditional transgenesis approach through the introduction of foreign DNA. However, functional studies of cis-elements are technically more challenging due to their non-coding and degenerate nature – transcription factors usually bind to a range of DNA sequences with varying affinities (Berg and von Hippel, 1987; Bulyk, 2003; Dermitzakis and Clark, 2002; Kotelnikova et al., 2005; Weirauch et al., 2013). Manipulation of genome thus becomes an important tool in interrogating such elements by precise insertions, deletions and replacement of nucleotides at targeted genetic locations. Early attempts utilised zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (Gaj et al., 2013). ZFNs and TALENs are composed of a customised DNA binding domain fused with a DNA cleavage domain of the restriction enzyme FokI. Due to the protein nature of the DNA binding domain, separate and extensive plasmid design and construction are needed for each targeted region, which renders these techniques time-consuming and labour-intensive. The discovery and adoption of clustered regularly interspaced short palindromic repeats (CRISPR) in recent years has quickly revolutionised the genome editing approach.

CRISPRs are arrays of short repetitive sequences interspersed with unique DNA spacers in most archaeal and many bacterial genomes (Barrangou et al., 2007; Garneau et al., 2010). Together with the CRISPR-associated proteins (Cas) they constitute the adaptive immune system of these prokaryotes. The unique DNA spacers in the CRISPR arrays are sequences acquired from bacteriophages or plasmids that have once invaded the host, thereby serving as the memory in this adaptive immune system. During the encounter of the same foreign genetic materials, these specific sequences can be transcribed into RNA and guide the associated Cas proteins to target DNA or RNA by Watson-Crick base pairing and results in cleavage (Fig. 1.7A). While a robust mechanism, an additional prerequisite of this cleavage is it requires the presence of a very short DNA motif – protospacer adjacent motif (PAM) – next to the target site (Marraffini and Sontheimer, 2010). PAM is thought to provide self versus nonself discrimination

so as to protect the CRISPR array from self-cleavage. Nonetheless, the RNA nature and the short PAM renders CRISPR easily programmable and capable of targeting majority of the mammalian genome, it is therefore swiftly harnessed for genome editing purposes.

Among the two classes and six types of CRISPR/Cas systems (Makarova et al., 2020), class 2 embodies a single endonuclease protein in the system, which confers a higher simplicity over the multi-protein complexes in class 1. In the class 2 system, Cas9 of type II and Cas12a (also known as Cpf1 – CRISPR from *Prevotella* and *Francisella* 1) of type V are thereby widely adopted to introduce double-strand breaks (DSBs) in the genome. CRISPR/Cpf1 possesses one RNA component – the crRNA (CRISPR RNA) – to recruit nuclease and confer target specificity, whereas in CRISPR/Cas9 an additional tracrRNA (trans-activating crRNA) serves as a structural scaffold in bridging Cas9 and crRNA (Gasiunas et al., 2012; Swarts and Jinek, 2018). The two RNAs in CRISPR/Cas9 were later engineered into one single-guide RNA (sgRNA), which leads to a more convenient genome editing platform (Jinek et al., 2012).

Upon the induction of DSB in genome by Cas9 or Cpf1, the cellular intrinsic DNA repair mechanisms step in to resolve the breakage (Ran et al., 2013). DSBs are repaired either by non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ involves the direct ligation of the break ends and is therefore error prone, which often leads to random insertions or deletions (indels). On the contrary, HDR employs a homologous template for precise repair, which is usually provided by the sister chromatid during and after DNA replication, or exogenously as a donor template for precise editing. While protein-coding gene can be easily knocked out by disruption of reading frame through indels, the degenerate cis-elements are harder to interfere. A large deletion removing the entire region by introducing two DSBs simultaneously is often required.

Engineering the epigenome

Alternatively, the epigenetic landscape surrounding the cis-elements can be altered by CRISPR via a modified Cas9. With only two mutations, D10A and H840A, Cas9 is rendered catalytically inactive or dead (dCas9) while the capability of complexing with sgRNA is unaffected (Gasiunas et al., 2012; Jinek et al., 2012). By tethering various protein effectors directly or indirectly to dCas9, the system can be repurposed to modulate gene expression and epigenetic marks at specific DNA regions, giving rise to a plethora of different possibilities (Gilbert et al., 2013; Pulecio et al., 2017; Qi et al., 2013; Xu and Qi, 2019). For instance, fusion of a repressive factor, KRAB (Krüppel associated box) domain (Fig. 1.7B), to dCas9 enables gene silencing by recruiting the enzymes responsible for laying down H3K9me3 and thereby inducing heterochromatinization (Thakore et al., 2015). Likewise, mSin3 interaction domain (SID) can be attached in multiple copies to dCas9 to repress genes by deacetylating histones (Konermann et al., 2013). Conversely, dCas9 linking to p300 (histone acetyltransferase p300) can activate genes via deposition of H3K27ac (Fig. 1.7B) (Hilton et al., 2015). In another approach four copies of trans-activator VP16 from herpes simplex virus (VP64) were concatenated to dCas9 to induce transcription (Gilbert et al., 2013; Perez-Pinera et al., 2013).

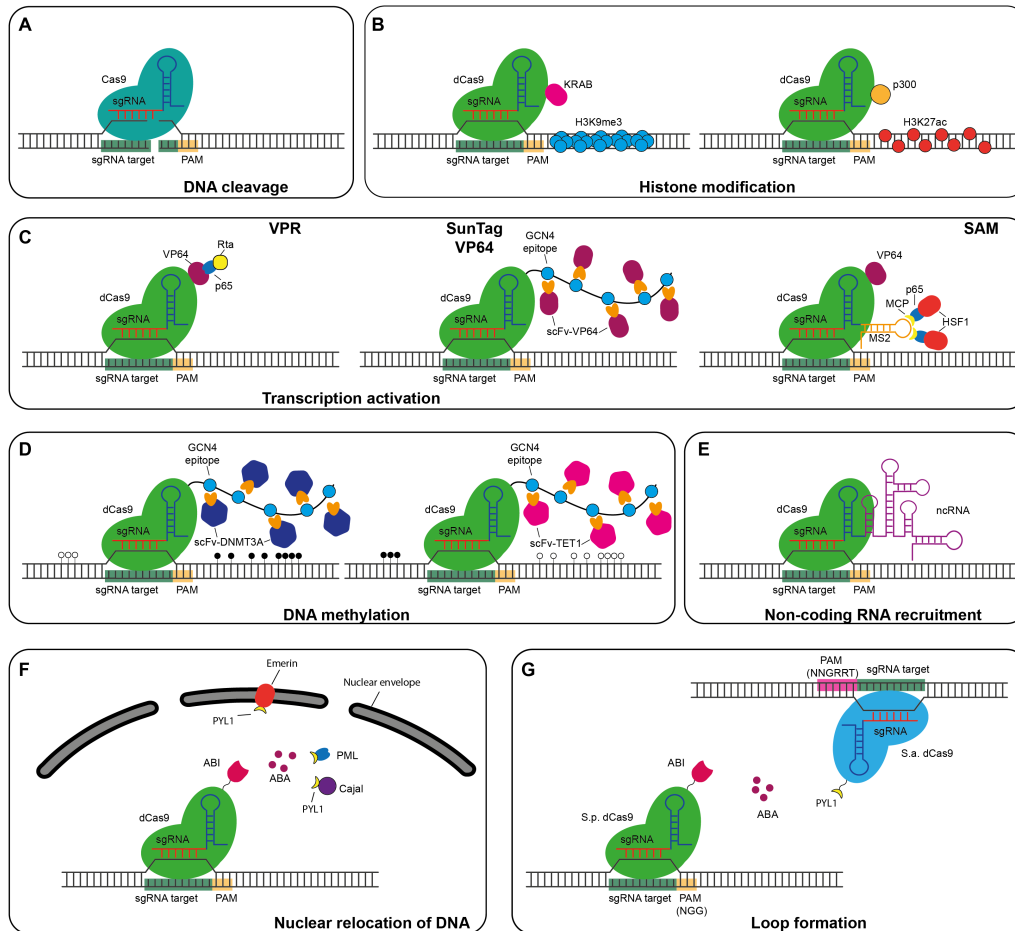


Figure 1.7. The expanding CRISPR toolbox in manipulating genome.

A) CRISPR/Cas9 generates DSBs. B) Catalytically dead Cas9 (dCas9) attached to KRAB domain is able to lay down repressive H3K9me3 histone marks, whereas dCas9-p300 can induce H3K27ac. C) Transcription activation can be mediated by direct fusing dCas9 with VP64, p65 and Rta (VPR); or via the SunTag system – dCas9 is fused to multiple GCN4 epitopes, which in turn recruit multiple VP64. SAM involves direct fusion of dCas9 with VP64, and a unique sgRNA bearing RNA aptamer MS2, which can further recruit p65 and hsf1 via the MS2 coat protein (MCP). D) DNA can be methylated via the SunTag system recruiting DNMT3A, or demethylated using the same system recruiting TET1. E) Non-coding RNA such as eRNA or lncRNA can be appended to sgRNA and relocated to the target region. F) Dimerization partners ABI and PYL1 can be fused to dCas9 and nuclear bodies, respectively. Upon administration of abscisic acid (ABA), DNA can then be relocated to other nuclear compartments. G) Orthologous dCas9 species, bearing different PAM, are separately attached to ABI and PYL1, and chromatin loop can be induced in the presence of ABA.

Among the above approaches involving direct fusion of a single effector domain, KRAB appears to be sufficiently effective in targeted gene repression, whereas trans-activators exhibit modest effects by acting alone (Gilbert et al., 2013; Perez-Pinera et al., 2013; Thakore et al., 2015). Various strategies were then devised to improve the efficiency of CRISPR-mediated activation (CRISPRa). Of note, one approach involves linking additional effectors p65 (NF- κ B p65 subunit) and Rta (Epstein-Barr virus R trans-activator) to dCas9-VP64 (Fig. 1.7C), which is referred as VPR (Chavez et al., 2015). Due to the large size of dCas9, direct fusion of more effector domains becomes problematic. To tackle this, a SunTag system was developed to recruit 10-24 copies of VP64 (Fig. 1.7C) (Tanenbaum et al., 2014). Instead of direct tethering, dCas9 is linked with a tandem array of short peptide fragments GCN4, with each peptide recruiting a GCN4-specific single-chain variable fragment (scFv) that is joined to VP64. A third notable system achieves the indirect recruitment via sgRNA instead (Konermann et al., 2015). Designated as synergistic activation mediator (SAM, Fig. 1.7C), the system modifies the structure of sgRNA so that it harbours two RNA scaffold aptamers for bacteriophage protein MS2, with each aptamer binding to two MS2 coat proteins that are individually linked to p65 and HSF1 (heat shock factor 1). Together with a VP64 that is already tethered to dCas9, SAM can thus recruit one VP64 and four copies of both p65 and HSF1 to facilitate gene activation.

The indirect recruitment strategy can be extended to other effectors to improve the efficacy and flexibility. Through SunTag, multiple DNMT3A (DNA methyltransferase 3A) can be recruited to achieve more specific DNA methylation with less off-target effects (Fig. 1.7D) (Pflueger et al., 2018), whereas multiple TET1 (Ten-eleven translocation methylcytosine dioxygenase 1) catalytic domains allow a more efficient DNA demethylation process (Fig. 1.7D) (Morita et al., 2016). On the other hand, RNA scaffolding enables multi-gene regulation, ascribed to the existence of a variety of RNA aptamers, such as MS2, PP7, boxB, COM and PUF motifs (Cheng et al., 2016; Ma et al., 2016; Zalatan et al., 2015). Together with their specific RNA binding protein partners, targeted gene activation and suppression along with DNA methylation and others can be simultaneously performed in the same cells.

Apart from short RNA aptamer, sgRNA can also be engineered to incorporate other RNA sequences and even natural lncRNAs (Fig. 1.7E) (Shechner et al., 2015). It was demonstrated that lncRNA *HOTTIP* with size of around 4.8 kb can be successfully appended to sgRNA without abrogating its property in directing dCas9 to the target site. As mentioned previously, *HOTTIP* can lay down H3K4me3 and thereby gene activation (Wang et al., 2011). This strategy, termed as CRISPR-display, can thereby relocate lncRNAs to any other regions and help modulate gene expression (Shechner et al., 2015).

Constructing the 3D genome

Inducible heterodimerisation between protein domains can be deployed in CRISPR to precisely regulate gene expression and, in other novel approach, relocate chromatin to a different nuclear space. The heterodimerisation can either be induced by chemicals or light and a number of such dimerising partners have been used in the CRISPR platform. The former group includes the ABI-PYL1 dimers which is induced by abscisic acid (ABA), GID1-GAI by gibberellin (GA) and rapamycin-inducible FKBP-FRB (Bao et al., 2017; Chen et al., 2017a; Gao et al., 2016). Whereas examples of the light inducible dimers are blue light inducible CRY2-CIB1 and red light inducible PhyB-PIF (Levskaya et al., 2009; Nihongaki et al., 2017; Nihongaki et al., 2015; Polstein and Gersbach, 2015). Thus, with one domain attached to dCas9 while the other to any effector proteins, gene transcription or the chromatin landscape can be modulated in a spatiotemporal manner. When the PYL1 domain is fused to emerin (Fig. 1.7F), a component at the inner nuclear membrane, specific chromatin region can then be directed to nuclear periphery via dCas9-ABI in the presence of ABA, leading to gene repression (Wang et al., 2018). This system, designated as CRISPR-GO (CRISPR-genome organiser), is also able to reposition chromatin to other membrane-less nuclear bodies such as Cajal bodies and promyelocytic leukaemia (PML) bodies, making it a versatile platform to manipulate the spatial organisation of the genome.

Other than nuclear repositioning, *de novo* chromatin loops can also be programmed by dCas9 and dimerising partners. One such system, LADL (light

activated dynamic looping), utilises light inducible CRY2-CIB1 dimers to mediate loop formation upon stimulation of blue light (Kim et al., 2019a). Contrarily, the other two systems combine inducible heterodimerisation with orthologous Cas9 variants to establish *de novo* chromatin loop (Fig. 1.7G) (Hao et al., 2017; Morgan et al., 2017). Orthologues of Cas9 from different bacteria differ from one another in their nuclease sizes and structures, their associated sgRNA (or tracrRNA) scaffold sequences and, most importantly, their PAM usages (Esvelt et al., 2013). For example, the conventional and most popular Cas9 is from *Streptococcus pyogenes* with the PAM site NGG, whereas Cas9 from *Staphylococcus aureus* utilises NNGRRT as PAM (Jinek et al., 2012; Ran et al., 2015). Together with their distinct corresponding sgRNAs, the two Cas9 orthologues are capable of localising in different regions simultaneously without any cross-reactivity. Therefore, multigene targeting and regulation can be accomplished by employing the dCas9 version of each orthologue. In one of the systems, referred as CLOuD9 (Chromatin loop reorganisation using dCas9), dCas9 from the above two species are separately fused to ABI and PYL1 and directed to two different DNA regions (Morgan et al., 2017). Upon the administration of ABA, a *de novo* chromatin loop is therefore created consequential to the heterodimerisation between ABI and PYL1. Overall, along with CRISPR-GO, LADL and CLOuD9 thus provide another means to manipulate the genome in 3D.

1.4. Genome reorganisation during B cell development and differentiation

The development and differentiation of B and T lymphocytes is central to the adaptive immune system in that it confers a protective and specific response against an invading pathogen (Chan et al., 2020b). The developmental process involves extensive morphological and functional alterations. A property unique to lymphocytes is the process of programmed DNA rearrangement, which involves random shuffling of gene segments to generate a huge repertoire of receptors, with each specifically recognising a particular foreign antigen. This is enabled by the three-dimensional genome organisation and its intricate interplay with a myriad of other cellular processes such as gene transcription, DNA repair, replication and mitosis.

Early B cell development

B lymphocytes are derived from the common lymphoid progenitor (CLP) in the bone marrow (Fig. 1.8). A synergistic network of transcription factors, such as PU.1, Ikaros, E2A, EBF1 and PAX5, are critical in the specification and commitment to the B cell lineage (Fig. 1.8) (Medina et al., 2004). For example, EBF1 is expressed in progenitor B (pro-B) cells and is involved in the transcription of B cell signature genes and also the suppression of the gene programme for other lymphoid lineages (Zandi et al., 2008). Early FISH experiments comparing pre-B and naïve B cells revealed different spatial locations of CD2 and $\lambda 5$ genes, with the position correlated with the gene expression status, suggesting chromatin repositioning during early B cell differentiation is functionally important (Brown et al., 1997). Similarly, immunoglobulin heavy (*Igh*) and κ (*Igk*) light chains were shown to move away from nuclear periphery to become more centrally located in committed pro-B cells (Kosak et al., 2002). Recent Hi-C experiment further suggested relocation as a global phenomenon, in which an extensive A/B compartment switch takes place during the developmental transition from pre-pro-B cells into pro-B cells (Lin et al., 2012). Gene loci specifically transcribed in B cells such as EBF1, Ig heavy and light chains relocate from B to A compartment, concurrent with their transcriptional activation. However, not all genes that experienced the switch from B to A compartment are accompanied by gene

upregulation, but instead were repressed and associated with the histone modification H3K27me3. This is in agreement with a recent finding mentioned in the previous section, where a B to A compartment switch does not necessarily associate with gene activation, but only provides a permissive environment favourable for gene activation (Miura et al., 2019). As such, the relocation of H3K27me3 repressed gene to A compartment is believed to pave the way for activation in future developmental transitions. Besides CTCF and cohesin, some of the transcription factors are implicated in restructuring chromatin topology, and that is independent of transcription as alterations in compartments and TADs precede transcriptional changes (Stadhouders et al., 2018). The transcription factor PAX5 was demonstrated to play such a role in establishing chromatin topology specific to B cells (Johanson et al., 2018b). Pro-B cells devoid of PAX5 suffer from a loss of thousands of chromatin interactions, some of which were restored upon the re-introduction of PAX5. The process is entirely independent of transcription as the chromatin interactions were rescued even in the presence of α -amanitin, a Pol II elongation inhibitor. With no intrinsic chromatin remodelling capability, PAX5 is speculated to enforce the architectural role by interacting with other protein partners, perhaps CTCF, cohesin, or other ATP-dependent chromatin modifiers.

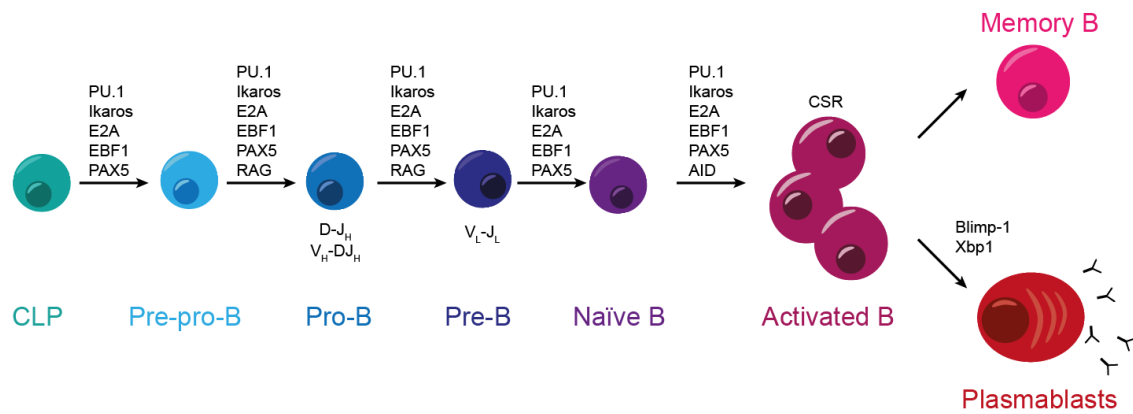


Figure 1.8. B lymphocyte development.

During early B cell development in the bone marrow the common lymphoid progenitor (CLP) differentiates sequentially into pre-pro-B, pro-B and pre-B cells, eventually becoming mature, resting naïve B cells, which then exit bone marrow and enter circulation. The developmental process is dependent on transcription factors PU.1, Ikaros, E2A, EBF1 and PAX5. V(D)J recombination, mediated by RAG recombinase complex, occurs in pro-B and pre-B stages with D-J_H recombination precedes V_H-D_H recombination in pro-B cells, followed by recombination in light chain (V_L-J_L) in pre-B cells. Upon pathogen encounter in periphery, naïve B cells become activated and rapidly proliferate. This clonal expansion is accompanied by SHM as well as CSR, mediated by AID, to alter antibody specificity and isotype, after which the activated B cells upregulate Blimp-1 and Xbp1 to differentiate into antibody-secreting plasmablasts or plasma cells. In a different developmental trajectory, the activated B cells can instead differentiate into memory B cells.

V(D)J recombination

Concomitant with the compartmental switch and chromatin relocation, of particular uniqueness in early B cell development is the assembly of a functional antigen receptor (Fig. 1.8). The genes for immunoglobulin heavy chain (*Igh*), κ (*Igk*) and λ (*Igl*) light chains exist as non-functional loci dispersed with two or three arrays of gene segments, which ultimately encodes for the antigen binding domain of Ig (Hozumi and Tonegawa, 1976; Tonegawa, 1983). These arrays are the variable (V), diversity (D) and joining (J) gene segments (Fig. 1.9C), where arrays of D segments are only present in the *Igh* locus. While there are 4 J_H and 10-15 D segments in *Igh* locus, the V_H region has over a hundred gene segments scattered throughout (Jhunjhunwala et al., 2009). Altogether the three arrays span a huge genetic distance of megabase level. A functional Ig locus is assembled through a programmed and remarkable DNA rearrangement process – V(D)J recombination – where one gene segment from each of the V, D and J regions are randomly selected and recombined (Alt et al., 1984). While only one recombination event – $V_{\kappa/\lambda}$ to $J_{\kappa/\lambda}$ – occurs in either *Igk* or *Igl* loci, the rearrangement in *Igh* is a step-wise process in which D to J_H recombination precedes the final V_H to DJ_H recombination. Each recombination event involves the generation of DSBs at the corresponding gene segments, mediated by the recombination activating gene (RAG) complex (Schatz and Ji, 2011). RAG complex recognises specific DNA motifs – the recombination signal sequences (RSS) – flanking each of the V, D and J segments. RSSs have two variants which differs in the length of the middle spacer region, referred as the 12RSS and 23RSS (Fig. 1.9C). Only a pair of complementary RSSs – 12RSS/23RSS – can bind and stabilise with the RAG complex (Fig. 1.9D). During the recombination process, a pair of compatible RSSs, for example from a D and a J_H segment, would eventually complex with the proteins upon which DSBs are induced (Helmink and Sleckman, 2012). The lesions would then be resolved by NHEJ with the intervening sequences deleted. Due to the vast span of genetic distance by the segments, 3D chromatin topology is therefore essential in tightly regulating the recombination process and also ensuring roughly equal chance of usage of every gene segment.

Along the V(D)J regions, RAG recombinase has a strong binding preference to a region designated as recombination centre (RC) (Fig. 1.9C), which covers all the J_H segments and a downstream intronic enhancer E_μ (Ji et al., 2010; Schatz and Ji, 2011). Many V_H segments are therefore deemed too distant from the RC. However, Ig loci contract in large scale upon relocation away from nuclear lamina, as illustrated by FISH experiments and also the increase of long-range interactions within the loci (Kosak et al., 2002; Lin et al., 2012; Stadhouders et al., 2014). The contraction thus brings distal V_H regions into close vicinity to RC (Fig. 1.9E), thereby rendering relatively unbiased utilisation of V_H segments. The B cell factor PAX5 is critical in mediating this process – pro-B cells deficient in PAX5 fail to undergo *Igh* locus contraction (Fuxa et al., 2004). Consequently, such mutant pro-B cells exhibit a complete bias towards proximal V_H segments (Fuxa et al., 2004; Nutt et al., 1997). It was later discovered that the distal V_H regions are dispersed with multiple distinct PAX5 binding sites known as PAIR (PAX5-activated intergenic repeat), which is implicated in juxtaposing distal V_H segments to the RC (Ebert et al., 2011). Several of these elements were shown by 3C experiment to interact with E_μ, thus forming a loop between distal V_H segment and RC (Verma-Gaur et al., 2012). Other than PAX5, a ubiquitous transcription factor Yin Yang 1 (YY1) was also demonstrated to mediate locus contraction (Guo et al., 2011a; Medvedovic et al., 2013). CTCF and cohesin are, rather unsurprisingly, involved in locus contraction (Degner et al., 2011), but more importantly, they also facilitate other aspects during V(D)J recombination.

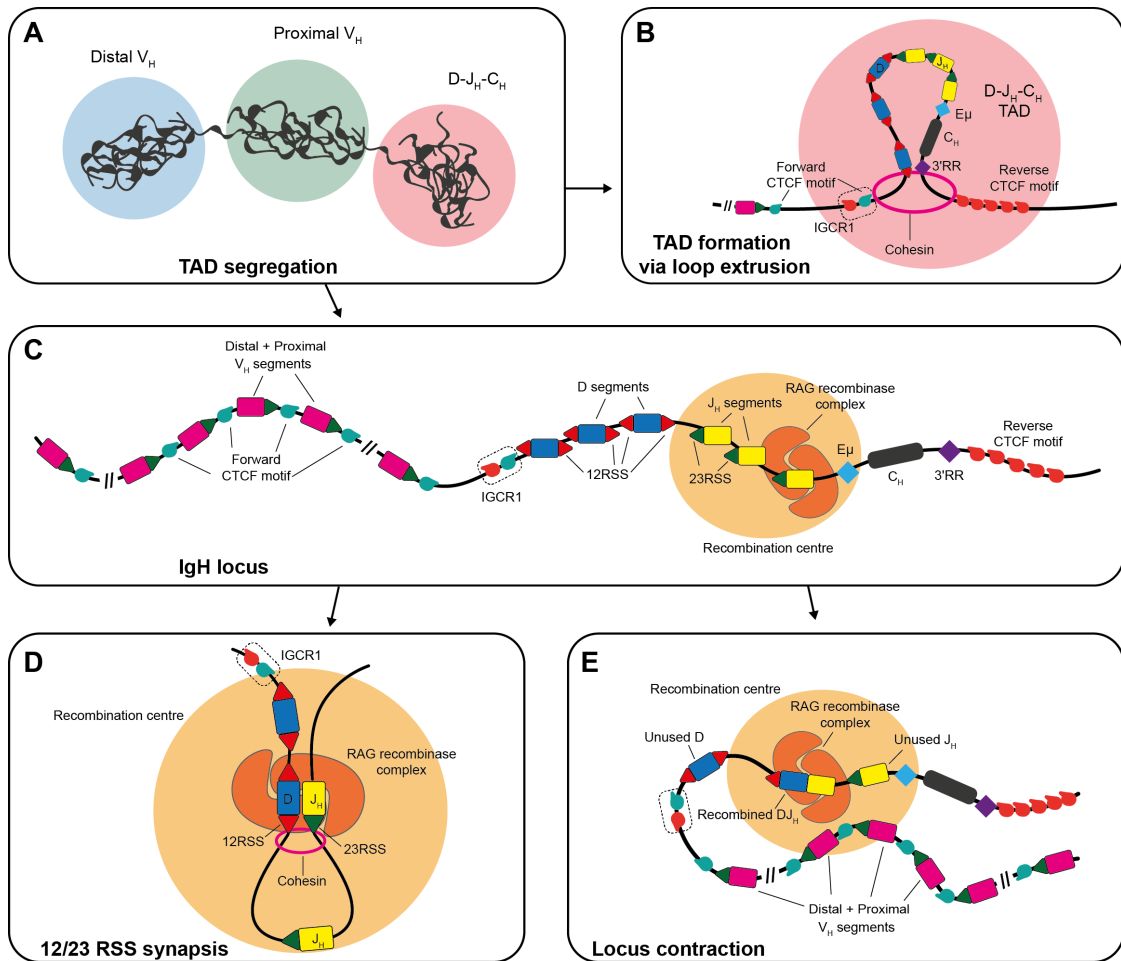


Figure 1.9. TADs and chromatin loops in IgH locus during V(D)J recombination.

A) IgH locus is segregated into three TADs. B) The TAD comprising D, J_H and C_H regions is presumably formed via loop extrusion, with cohesin retained between the forward CTCF motif at the intergenic control region 1 (IGCR1) and reverse CTCF motifs downstream of 3' regulatory region (3'RR). C) The linear IgH locus indicating the gene fragments as well as cis-regulatory elements. D) The pairing between 12RSS and 23RSS of D and J_H segments, respectively, during D-J_H recombination. Such synapsis is presumably facilitated through the linear scanning of cohesin, and further stabilised by RAG recombinase. E) Locus contraction of IgH locus upon successful D-J_H recombination, bringing the distal V_H gene segments to close physical proximity to the recombination centre.

Scattered along the immunoglobulin loci are a wealth of CTCF sites – at the 5' end there are over a hundred CTCF sites interspersed within the V_H region, all of which are in a forward orientation and facing a single reverse CTCF motif at the start of the D region (Fig. 1.9C) (Aiden and Casellas, 2015). Immediate downstream of this reverse CTCF motif rests a forward CTCF site. Together this divergent CTCF pair comprises the intergenic control region 1 (IGCR1, Fig. 1.9B, C), which are critical in segregating the entire *Igh* locus into well-insulated TADs (Guo et al., 2011b; Jhunjhunwala et al., 2008; Montefiori et al., 2016). One such TAD harbours the D and J_H regions, as well as the downstream constant region (C_H) and 3' regulatory region (3'RR, Fig. 1.9B). This is presumably established through loop extrusion by cohesin, anchoring between the forward CTCF site in IGCR1 and a tandem array of reverse CTCF motifs located downstream of 3'RR. On the other hand, the reverse CTCF motif in IGCR1 probably assists in stratifying the entire V_H region into two separate TADs (Fig. 1.9A). Recombination between D and J_H segments initiates from the RC with RAG recombinase associates with one of the J_H segments (Jain et al., 2018; Zhang et al., 2019b). Through the extrusion action of cohesin, RAG can linearly scan through the upstream D segments and pair up complementary RSSs for the subsequent recombination (Fig. 1.9D). Due to the action of linear extrusion, alignment between compatible RSSs is mostly in a convergent orientation, which highly facilitates deletional joining at the expense of the inversional one (Zhang et al., 2019b). Additionally, the scanning is impeded by IGCR1 so that V_H segments cannot be accessed by RAG (Fig. 1.9B, D). This leaves only D and J_H segments to pair up and recombine, which prevents premature V_H -D joining and thus ensures a step-wise recombination process.

In addition to IGCR1 and TAD insulation, there are other mechanisms at force to enact an ordered rearrangement process. Upon repositioning of *Igh* locus away from nuclear periphery, associated histones start to gain acetylation, which is associated with loosening of the chromatin in a sequential manner (Chowdhury and Sen, 2001). D, J_H and C_H regions are first hyperacetylated, making the regions accessible to RAG recombinase. Histones along V_H segments are not acetylated until later in response to a successful D- J_H joining, whereupon the region becomes acetylated and permissive to final V_H -D J_H recombination. Also

concomitant with histone acetylation is the initiation of non-coding, or germline, transcription, which also manifests in a stage-specific fashion (Corcoran, 2010). The first non-coding transcript is generated from the E μ in the RC before D-J $_H$ joining (Bolland et al., 2007; Lennon and Perry, 1985), and after which non-coding transcription occurs in both DNA strands over the V $_H$ region (Corcoran et al., 1998; Yancopoulos and Alt, 1985). The process of transcription is thought to be important in maintaining the accessibility of the locus (Bolland et al., 2004), as the progression by Pol II is implicated in altering chromatin state by recruitment of histone acetyltransferases, methyltransferases and other ATP-dependent chromatin remodelling complexes (Ng et al., 2003; Wilson et al., 1996). Indeed, deletion of enhancer E μ abrogates transcription and compromises D-J $_H$ recombination (Afshar et al., 2006; Perlot et al., 2005). Ablation of non-coding transcription in other studies resulted in loss of histone acetylation and H3K4me3, which then led to underuse of the affected gene segments, overall indicating the Pol II elongation activity is the key factor in maintenance of accessibility (Abarregui and Krangel, 2006, 2007). Furthermore, non-coding transcription is also implicated in restructuring chromatin in three dimensions. As mentioned earlier, PAX5 and YY1 are important factors in locus contraction, yet intriguingly they also concurrently activate non-coding transcription (Ebert et al., 2011; Verma-Gaur et al., 2012). Loss of YY1 leads to drastic reduction of non-coding transcription and impaired long-range PAIR-E μ interaction. However, loss of PAX5, which leads to defects in locus contraction, reduces transcripts initiated from PAIR but does not affect other non-coding transcription (Ebert et al., 2011; Hesslein et al., 2003). Thus, the causal relationship between looping and transcription is still unclear. In addition, active enhancers have been speculated to recruit DNA into transcription hubs or condensates, and thereby act as a focal point for further coalescence of other protein factors and distal DNA regions (Stubbington and Corcoran, 2013). In support of this, E μ was shown to drive non-coding transcription almost continuously and locate at a transcription factory (Bolland et al., 2004; Osborne et al., 2007).

Each B cell expresses only one antigen receptor through a process known as allelic exclusion, where a successful V(D)J recombination at one allele would subsequently render the other homologue epigenetically silenced (Brady et al.,

2010). While D to J_H recombination occurs exceptionally in both *Igh* alleles almost simultaneously, the other V_{K/λ}-J_{K/λ} or V_H-DJ_H recombination is carried out in one allele at a time, with *Igh* recombination precedes *Igk* or *Igl* recombination (Fig. 1.8). One interesting phenomenon involving pairing of homologous alleles appears to ensure one allele cleavage at a time (Hewitt et al., 2008; Hewitt et al., 2009). The pairing, hence allelic exclusion, is regulated by RAG recombinase and the associated DNA repair protein ATM (ataxia-telangiectasia mutated), where the generation of DSB at one locus would lead to the repositioning of the other allele, mediated by ATM, to pericentromeric heterochromatin. Subsequently this is accompanied by the de-contraction of the *Igh* loci to prevent further recombination (Roldan et al., 2005). Around the same time, one of the *Igk* alleles would also reposition to pericentromeric heterochromatin and apparently direct the recruitment of the non-functional *Igh* allele, presumably through an inter-chromosomal interaction (Goldmit et al., 2005; Hewitt et al., 2008; Roldan et al., 2005). Collectively, these events prevent further recombination in the non-functional *Igh* locus while ensuring only one *Igk* allele is undergoing a next round of rearrangement.

B cell activation

Shortly after antigen stimulation, naïve quiescent B cells experience a tremendous change in both morphology and function (Fig. 1.8) (Chan et al., 2020b). Early studies documented an exit of G₀ phase and entry into cell cycle, accompanied by increases in synthesis of RNA and protein, histone acetylation, and an increase in cell size so dramatic that it is easily observable under a conventional microscope (Chan et al., 2020b; Jaehning et al., 1975; Pogo et al., 1966, 1967; Rawlings et al., 2011). Intriguingly, the first cell division does not occur until around 30 hours post-activation, relative to the subsequent rapid cell division every 6-10 hours thereafter (Hawkins et al., 2009; Turner et al., 2008). Several recent studies have further shown during this early activation chromatin decondenses, and this requires the transcription factor Myc and energy input in the form of ATP, indicating it is a highly energy-dependent process (Chan et al., 2020b; Kieffer-Kwon et al., 2017; Kouzine et al., 2013; Nie et al., 2012). The chromatin de-condensation is accompanied by the loss of long-range

chromosomal contacts and the formation of thousands of DNA loops (Bunting et al., 2016; Kieffer-Kwon et al., 2017). These short-range contacts mainly consist of CTCF-mediated loops, 5'-3' gene loops and other enhancer-promoter contacts. It was demonstrated that the promoters gain interactions with multiple non-coding regions during this period and gene expression level is positively correlated with interaction frequency (Chaudhri et al., 2020; Javierre et al., 2016; Kieffer-Kwon et al., 2017). Thus, it appears that the predominant function of these loops is to drive and augment gene expression. As mentioned previously, enhancer-promoter interactions can either be instructive or permissive in modulating gene expression (de Laat and Duboule, 2013). The positive correlation between loop formation and gene expression during B cell activation therefore suggests that these *de novo* loops are presumably largely instructive. Additionally, the genome restructuring was demonstrated to associate with minimal A/B compartment switch (Bunting et al., 2016; Javierre et al., 2016; Johanson et al., 2018b; Kieffer-Kwon et al., 2017; Kieffer-Kwon et al., 2013). Thus the majority of genome reorganisation involves DNA loops and is largely restricted in their initial native compartment type. This is in striking contrast to stem cells and pro-B cells, where there is extensive A/B compartmental switch during lineage specification (Dixon et al., 2015; Lin et al., 2012). Presumably this is a slower process as genes in the inactive compartment will need to flip to the permissive environment before any transcription could happen (Chan et al., 2020b). Contrarily, the minimal compartmental switch during B cell activation can therefore ensure a more rapid chromatin restructuring and hence a more rapid immune response.

Class switch recombination

Once the activated B cells undergo the first cell division, they rapidly divide and expand for many days until terminal differentiation (Fig. 1.8) (Chan et al., 2020b; Hasbold et al., 2004). The rapid expansion is accompanied by some critical local architectural changes. Class switch recombination is the molecular hallmark of B cell activation, where deletional DNA recombination elicits the replacement of the expressed immunoglobulin (Ig) μ and δ heavy chain constant regions (C_H) with the downstream γ , ϵ or α C_H regions (Stavnezer and Schrader, 2014; Xu et al., 2012). This is a tightly regulated process involving the interplay of non-coding

transcription, double strand breaks (DSBs), DNA repair and 3D chromatin looping. In a resting B cell, a DNA loop is formed at the *Igh* locus, between the 3' regulatory region (3'RR) and the intronic enhancer (E_{μ}) just upstream of the constant regions (Fig. 1.10, left), which are around 200 kb apart (Chatterjee et al., 2011; Kenter et al., 2012; Wuerffel et al., 2007). Downstream of 3'RR are multiple CTCF binding sites and recent studies have suggested that the DNA-DNA interaction is mediated by CTCF via loop-extrusion (Zhang et al., 2019a). Additionally, the interaction co-occurs with a non-coding germline transcription (GLT) around the μ switch (S) region, which plays an important role in unwinding the S region DNA for activation-induced cytidine deaminase (AID), a ssDNA mutator, to induce DSBs (Matthews et al., 2014; Muramatsu et al., 2000). Once the B cell is activated, depending on the stimulation and cytokines it received, for example LPS with IL-4 will result in a switch to IgG1, the corresponding S region (Sy1 herein) will come into close proximity to the pre-existing loop, creating a three-point DNA synapsis (Fig. 1.10, right) (Wuerffel et al., 2007). In the meantime, the stimuli also trigger GLT at the acceptor switch region (Sy1) followed by recruitment of AID to the interacting foci. DSBs are then induced at the S regions and subsequently resolved by non-homologous end joining (NHEJ) with the intervening DNA deleted. How such activation-induced chromosomal looping is established and its relationship with germline transcription remains unclear, despite several attempts to address this question (Braikia et al., 2017; Feldman et al., 2015; Kenter et al., 2012; Marina-Zarate et al., 2017; Thomas-Claudepierre et al., 2016; Yewdell and Chaudhuri, 2017). As the induced germline transcription and 3D looping is antigen and cytokine dependent, it is conceivable that the downstream factors which elicit the GLT are also contributing to the juxtaposition of the acceptor switching region to the E_{μ} -3'RR loop (Fig. 1.10, middle path) (Feldman et al., 2015). However, given the emerging evidence about the role of transcription in shaping genome architecture (Mele and Rinn, 2016; Thomas-Claudepierre et al., 2016; van Steensel and Furlong, 2019), the germline transcript itself or the process of transcription could plausibly assist the formation of the DNA synapsis (Fig. 1.10, top path). Also, conversely, it is possible that the looping itself allows the 3'RR to gain access and activate the GLT, herein the downstream protein factors induced by cytokine signals play an additional

architectural role (Chan et al., 2020b; Feldman et al., 2015; Thomas-Claudepierre et al., 2016) (Fig. 1.10, bottom path).

Other than CTCF, recent studies have unravelled the architectural roles of some other proteins in mediating CSR. The DSB damage response protein p53-binding protein 1 (53BP1) was once thought to be critical in resolving the DSB (Manis et al., 2004), however, several studies have demonstrated an additional role of 53BP1 in mediating the chromosomal loop in *Igh* locus (Feldman et al., 2017; Rocha et al., 2016). In the absence of 53BP1, the interactions between 3'RR and E μ are greatly diminished. Furthermore, Mediator, an evolutionarily conserved protein complex that is essential for Pol II transcription, has been shown to facilitate long-range DNA-DNA interactions between the switch regions and the 3'RR, and also the transcriptional activation of GLT at those switch regions (Thomas-Claudepierre et al., 2016). Lastly, a ubiquitous transcription factor YY1 was also revealed to mediate the looping between E μ and 3'RR (Mehra et al., 2016). This loop is still present when YY1 lacks its transcriptional activation domain, suggesting YY1 has a direct architectural role in establishing the long-distance loop (Chan et al., 2020b).

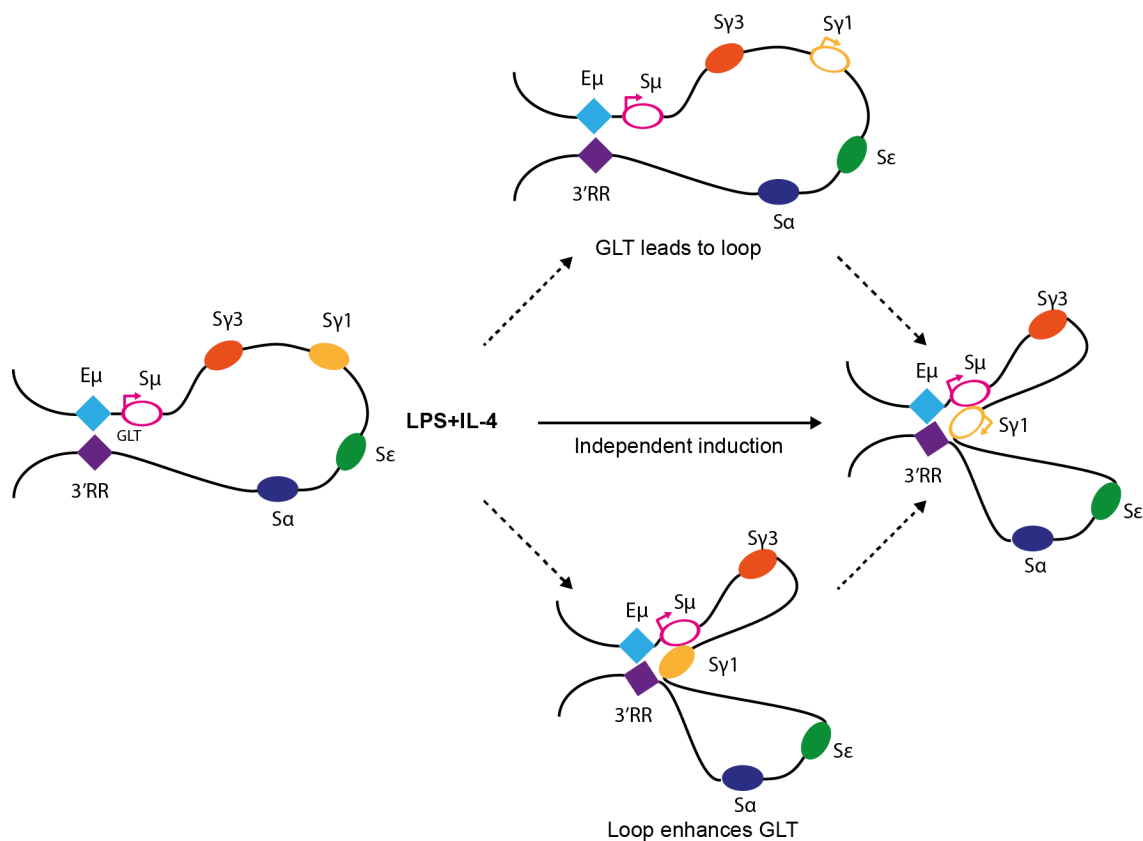


Figure 1.10. Chromatin loops during class switch recombination.

Shown on the left, in naïve mature B cells, a long-range interaction is formed between intronic $E\mu$ enhancer and $3'$ regulatory region ($3'$ RR) in IgH locus. GLT is constitutively active across $S\mu$ region, indicated as an open pink oval with arrow. Upon LPS and IL-4 stimulation, the switch region for $\gamma 1$ ($Sy1$) starts GLT and comes into close proximity to $E\mu$ (and therefore $S\mu$) and $3'$ RR, recombination will then occur between the two S regions with the intervening DNA deleted (depicted on right). In the establishment of CSR, the relationship between transcription and 3D looping remains unclear. GLT and looping can be independently induced by the cytokine signal (middle path). Another plausible scenario is the induction of GLT across acceptor switch region ($Sy1$) helps the juxtaposition (top path). A third model suggests the 3D looping brought enhancers ($3'$ RR) into close proximity to $\gamma 1$ and enhance GLT (bottom path) (Chan et al., 2020b).

Terminal differentiation into antibody-secreting cells

The prominent endoplasmic reticulum, extensive Golgi apparatus and eccentric position of the nucleus with a wheel-like chromatin configuration are all defining features of plasma cells (Fig. 1.8, 1.11), thought to enhance their massive antibody production (Albarracin and Fonseca, 2011; Chan et al., 2020b; Martinez et al., 2005). In contrast, the plasma cell genome organisation during differentiation and its functional significance to massive antibody synthesis and secretion is poorly understood. Similar to other terminally differentiating cells, heterochromatinization is thought to contribute to locking in the plasma cell transcriptional programme, while shutting down other unnecessary genes, for instance the regular B cell signature genes (Chan et al., 2020b; Dillon and Festenstein, 2002; Shaffer et al., 2002). Deposition of H3K27me3 by EZH2 and the removal of H3K4me1 by LSD1 are implicated in this process (Guo et al., 2018; Haines et al., 2018; Scharer et al., 2018). Consistent with this, Hi-C experiments documented extensive chromatin reconfiguration during this final stage of differentiation, which involves enhancer-promoter loops (Bortnick et al., 2020; Chan et al., 2020b; Johanson et al., 2018b). For example, the *Prdm1* locus, which codes for the master regulator for plasma cell differentiation Blimp-1, exhibits the strongest enhancer-promoter interaction at the plasmablast stage. Whereas the DNA contacts for Bcl6, an important transcription factor during B cell activation, are essentially lost. A considerable amount of compartmental switch was also observed, of notable examples Ebf1 and Bcl6, both of which are repositioned to a more repressive environment (Bortnick et al., 2020).

Three-dimensional DNA modelling from conformation capture assays revealed that the chromosomes of plasma cells adopt a unique elongated, α -helical structure, resulting in the characteristic cartwheel appearance (Bortnick et al., 2018; Bortnick et al., 2020; Chan et al., 2020b) (Fig. 1.11). This is in striking contrast to the spherical-like structure exhibited by the chromosomes in naïve B cells. The cartwheel, elongated configurations are mainly contributed by the larger chromosomes (chr1-9) where the long-range intra-chromosomal interactions are severely diminished. Instead, intriguingly, there is an enrichment of inter-chromosomal contacts between chromosome ends (Bortnick et al., 2018).

Using 3D imaging along with chromosome conformation capture, another study has shown that in plasma cells the active immunoglobulin genes (*Igκ*, *Igh* and *IgJ*), from three different chromosomes, are colocalised forming an apparent transcription factory (Chan et al., 2020b; Park et al., 2014) (Fig. 1.11). Even more strikingly, this active cluster is located at the nuclear periphery, a region generally associated with gene silencing. More than 30 years ago, Blobel (1985) proposed a gene gating hypothesis in which active genes are recruited near the nuclear pores for efficient transcript export. As formerly described, nuclear pores appear to be an exception at the nuclear periphery where gene activation is allowed (Blobel, 2010; Buchwalter et al., 2019; Chan et al., 2020b). In plasma cells, the colocalised *Ig* genes often have their transcripts share the same trafficking channel to the cytoplasm (Park et al., 2014). Therefore, it appears *Ig* genes colocalise near the nuclear pore for maximal transcript export, which further facilitates efficient and massive antibody production. In addition, to accommodate massive protein synthesis and secretion, plasma cells elicit the unfolded protein response (UPR) to cope with the ER stress (Gass et al., 2002). It is found that in plasma cells the genes involved in the UPR pathway, for instance *Xbp1*, *Atf4* and *Bax*, also aggregate from different chromosomes and form transcription hubs (Bortnick et al., 2020) (Fig. 1.11). Collectively, the inter-chromosomal interactions of *Ig* and UPR genes in plasma cells enable a coordinated and dedicated production of antibody (Chan et al., 2020b).

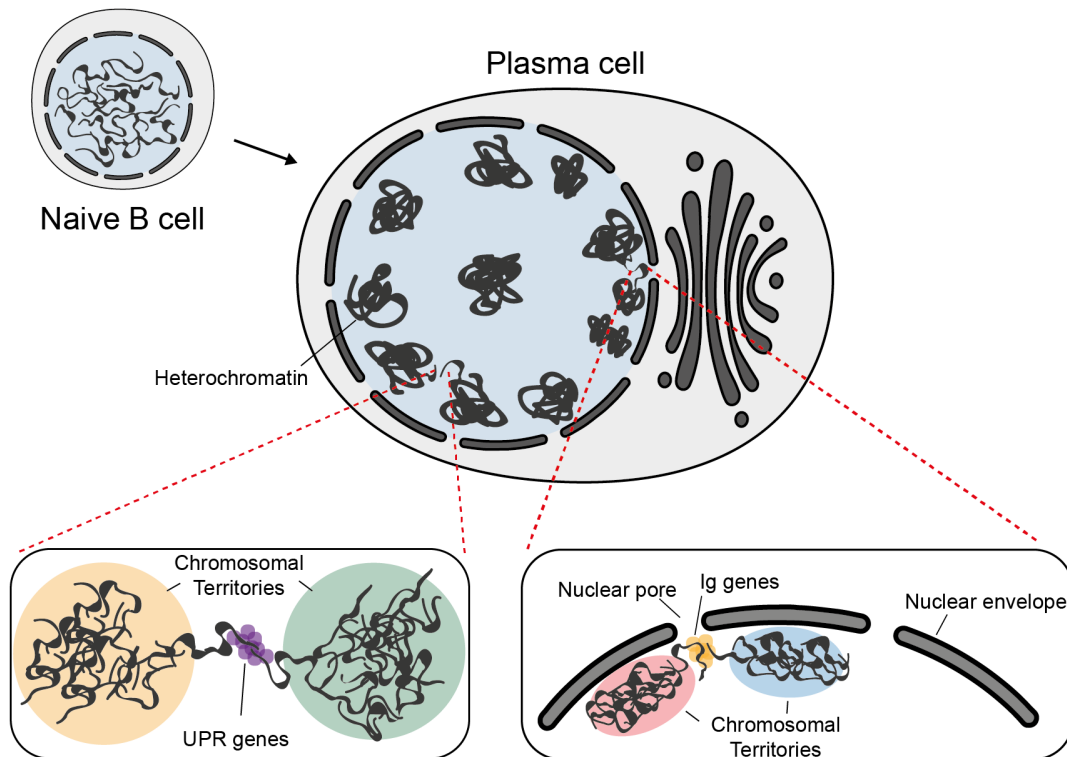


Figure 1.11. Genome reorganisation during terminal plasma cell differentiation.

Extensive heterochromatinization leads to a cartwheel configuration of chromatin in nucleus. Inter-chromosomal transcription hubs for UPR genes and Ig genes are independently established, with Ig transcription hub located near the nuclear pore for maximal transcript export (Chan et al., 2020b).

1.5. Genome reorganisation during T cell development and differentiation

Early T cell development

In contrast to B cells, T cells complete the maturation process in thymus. Notch signalling and the transcription factors TCF-1, Gata3 and Bcl11b are crucial players, in which Bcl11b enforces the lineage specificity and suppresses alternate programmes (Georgescu et al., 2008; Yui and Rothenberg, 2014). Long-range enhancer-promoter interactions have been documented in controlling proper expression of Gata3 (Hosoya-Ohmura et al., 2011; Ohmura et al., 2016) and Bcl11b (Li et al., 2013a) during this entire developmental process. Besides chromatin loops, relocation of genes, akin to B cells, was also concomitantly observed in developing T cells (Brown et al., 1999; Isoda et al., 2017), with Bcl11b being a notable example (Isoda et al., 2017). Before commitment to the T cell lineage, the Bcl11b locus was found to reside at the nuclear lamina, after which it moves to nuclear interior in the DN2 (double negative, CD4-CD8-) stage. As described previously, the repositioning is triggered by the transcription of non-coding RNA *ThymoD*, which was further shown to elicit demethylation of CTCF sites. Having being demethylated, the CTCF sites are then able to facilitate loop extrusion and the establishment of loop between Bcl11b promoter and its distant enhancer, thereby activating Bcl11b expression. Thus, it appears the repositioning of the locus is concomitant with gene upregulation, and further analyses suggest it does not only occur at a limited number of loci (Hu et al., 2018; Isoda et al., 2017). Again, similar to B cells, the repositioning of genes occurs at a genome wide level, with substantial A/B compartment switch detected at various transitions during development. In an attempt to disentangle the timing between transcription and the flipping of compartments, it was observed that the A to B compartment transition generally occurs after gene silencing, whereas transcription could precede after a compartment B to A switch (Hu et al., 2018). Although described from a different perspective, this is still in consistent with the finding from Miura et al. (2019), where A to B switch almost always linked to silencing while B to A transition does not necessarily associate with gene activation. Additionally, the observation is also in line with the transcription of *ThymoD* at the repressive nuclear periphery

(Isoda et al., 2017). Among the different developmental transitions, the most extensive flipping of compartments appears interestingly to take place following Bcl11b upregulation, during the DN2-DN3 transition (Hu et al., 2018). Activation of Bcl11b is also accompanied by an increase in intra-TAD connectivity, which is subsequently followed by the corresponding gene activation. The slight lag of transcription suggests that the chromatin reconfiguration is largely instructive in mediating gene activation. Furthermore, the co-occurrence of Bcl11b upregulation and the massive chromatin restructuring implied an architectural role of this master regulator (Hu et al., 2018). Indeed, similar to the role of PAX5 in B cells, deletion of Bcl11b in naïve CD4⁺ T cell results in a loss of many chromatin interactions, upon which the associated gene expression also drops. Thus, overall this indicates that Bcl11b is critical in maintaining a T cell specific genome architecture, which further supports a lineage specific transcription programme.

Analogous to B cells, during early development T cells also undergo V(D)J recombination to assemble the T cell receptor (TCR). The DNA rearrangement is performed on the TCR α and TCR β loci in the conventional $\alpha\beta$ T cells, or on the TCR γ and TCR δ loci in $\gamma\delta$ T cells. Details and mechanisms regarding genome wiring are largely parallel with that in B cells, with slight differences and tweaks outlined elsewhere (Bassing et al., 2002; Jhunjunwala et al., 2009; Outters et al., 2015). While assembling the TCR, T cells concomitantly transit through several DN stages and reach DP (double positive, CD4⁺CD8⁺) stage. Finally, these cells become mature naïve T cells being single positive for either CD4 or CD8 surface molecule. Mature T cells then exit the thymus and enter circulation.

T cell activation and terminal differentiation

Activation of naïve T cells follows a similar early kinetics as in B cells – entry into cell cycle, increase synthesis of RNA and protein, histone acetylation, increase of cell size and chromatin de-condensation which is dependent on ATP and Myc (Jaehning et al., 1975; Kouzine et al., 2013; Nie et al., 2012; Pogo et al., 1966, 1967; Rawlings et al., 2011). Again, activation is accompanied by reshuffling of enhancer-promoter interaction, which was shown to correlate positively with gene

expression (Burren et al., 2017; Javierre et al., 2016; Lai et al., 2018). However, CD4⁺ T cells contrast with B cells in the way that they possess a potential to differentiate into different subsets of cells with distinct functions (Zhu et al., 2010). Instead of undergoing a second round of DNA rearrangement at a particular locus as in B cells to modulate effector functions, CD4⁺ T cells alter and diverge their transcription profile drastically to express and secrete different cytokines. Upon antigen stimulation, depending on the cytokines that are present in the environment, naïve CD4⁺ T cells can differentiate into a number of T_H cell subsets, namely T_H1, T_H2 or T_H17 cells. While T_H1 cells express T-bet and secrete interferon- γ (IFN- γ), T_H2 cells upregulate Gata3 and release interleukin-4 (IL-4), IL-5 and IL-13, whereas T_H17 cells express ROR γ and produce IL-17 (Zhu et al., 2010).

Intriguingly, in preparation for this divergence of transcription programme, naïve CD4⁺ T cells apparently rely on inter-chromosomal interaction to coordinate the fate choice (Fig. 1.12A). The murine T_H2 responsive region encompassing IL-4, IL-5 and IL-13 genes on chromosome 11 was separately demonstrated to physically associate with the IFN- γ promoter on chromosome 10 (Spilianakis et al., 2005), and the IL-17 locus on chromosome 1 (Kim et al., 2014). Several cis-elements along the T_H2 loci, referred as the T_H2 locus control region (LCR), were shown to be crucial for this interaction, as the deletion of them leads to disintegration of the inter-chromosomal hub and also a skew towards T_H17 differentiation. Besides that, the co-localisation was shown to occupy in a permissive environment of the nucleus, and was demonstrated to be poised for a rapid transcriptional response. Following antigen stimulation and lineage specification, this inter-chromosomal interaction is reduced in the favour of intra-chromosomal interaction during the terminal differentiation process. Thus, collectively, this suggests that specific inter-chromosomal interactions may allow a coordination of choices of cytokine secretion and differentiation path. While interesting, the discoveries, which were based mainly on microscopy and 3C, have recently been called into doubt as a different approach using Hi-C has failed to identify such regulatory hubs (Johanson et al., 2018a). As has been pointed out earlier, the proximity ligation in Hi-C appears to be limited in recapitulating well-known inter-chromosomal structures such as nucleolus and nuclear

speckles (Quinodoz et al., 2018). As such, the disputed existence and functional importance of the inter-chromosomal interactions in CD4⁺ T cells remain to be verified perhaps using a ligation-independent approach such as GAM and SPRITE (Beagrie et al., 2017; Quinodoz et al., 2018).

Focusing on an individual locus, 3C experiments revealed that IL-4, IL-5 and IL-13 promoters, forming part of the T_H2 locus, are preconfigured to cluster with each other within close spatial proximity in all lymphoid cells including natural killer (NK) cells and B cells (Spilianakis and Flavell, 2004). These promoter clusters further interact with the cis-regulatory elements in the T_H2 locus control region (LCR) in CD4⁺ T cells and NK cells (Fig. 1.12B). This interaction with LCR was further demonstrated to be maintained by the transcription factor Gata3. As such, the enhancer-promoter interactions at the T_H2 locus are apparently in a permissive configuration, poised for activation. In contrast, the IFN- γ enhancer-promoter interaction is T_H1 specific (Fig. 1.12C) (Sekimata et al., 2009). The formation of such a loop requires the transcription factor T-bet and relies on CTCF as well as cohesin (Hadjur et al., 2009; Sekimata et al., 2009), presumably through the loop extrusion mechanism. Herein the interaction thus plays an instructive role in activating IFN- γ transcription, contrary to the permissive configuration in T_H2 loci.

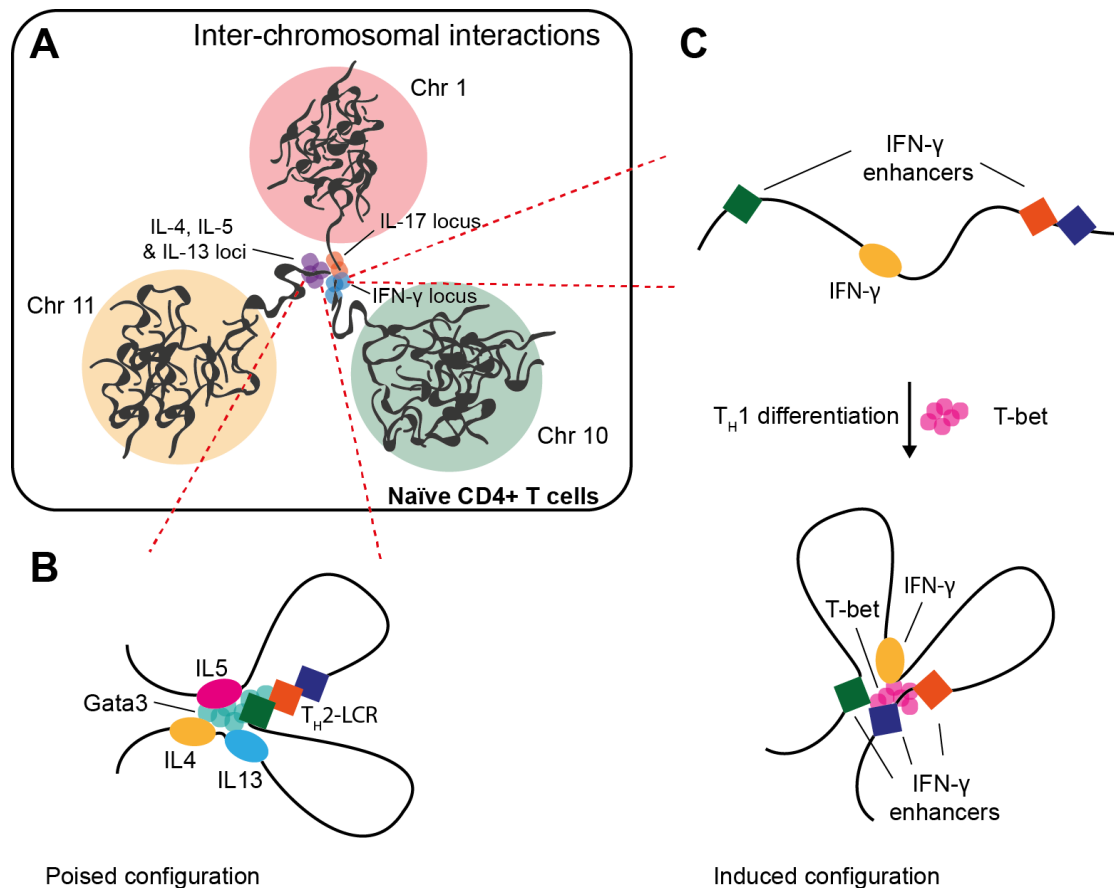


Figure 1.12. Genome organisation in naïve CD4 $^{+}$ T cells.

A) The T_H2 locus – IL-4, IL-5 and IL-13 loci, on chromosome 11 interacts with the IFN- γ promoter on chromosome 10, as well as the IL-17 locus on chromosome 1. B) In naïve CD4 $^{+}$ T cells, within the T_H2 locus, the cis-regulatory elements in the locus control region (LCR) have pre-established interactions with the T_H2 cytokine promoters. Gata3 appears to be important in forming and maintaining such poised configuration. C) The IFN- γ enhancer-promoter interactions are absent in naïve CD4 $^{+}$ T cells, but are induced upon T_H1 differentiation in the presence of T-bet.

The helper subset differentiation and lineage commitment are also accompanied by repositioning of gene loci. T_H1 specification was associated with the repositioning of loci encoding Gata3 and c-maf to heterochromatin, both of which are essential in regulating T_H2 differentiation, yet interestingly the IL-4 locus does not entirely undergo such relocation (Grogan et al., 2001; Hewitt et al., 2004). This suggests other epigenetic mechanisms are already in force to silence IL-4 expression. More intriguingly, the IFN- γ locus appears to constitutively associate with nuclear periphery, even in T_H1 cells. The location could be a reflection of the aforementioned gene gating hypothesis (Blobel, 1985), where the peripheral location enables a highly efficient transcript export of IFN- γ , thereby facilitates the synthesis and secretion of IFN- γ . In another study, the IFN- γ gene was also found to reside in a close proximity to the locus encoding furin from a separate chromosome (Hakim et al., 2013). Furin is an enzyme that is able to enhance IFN- γ secretion (Pesu et al., 2006), and the co-localisation of two genes could again reflect a coordination for an effective IFN- γ production. In addition, the study also conducted 4C assays and unveiled a distinct chromatin interaction network around IFN- γ and IL-4 loci, that are specific to the differentiating T_H1 and T_H2 cells respectively (Hakim et al., 2013). The interaction appears to rely on STAT signalling pathway, which is elicited upon the cytokine stimulation. The interaction around IFN- γ locus in T_H1 cells depends on STAT4, whereas the IL-4 specific wiring in T_H2 cells is induced by STAT6. Similarly, by comparing the interactome between human naïve CD4⁺ T cells and differentiated T_H17 cells, a genome-wide study has recently revealed that T_H17 cells also acquired lineage-specific enhancer-promoter interactions (Mumbach et al., 2017). Overall, it suggests that CD4⁺ T cells would orchestrate a divergence of genome reorganisation during the terminal differentiation into T_H1 , T_H2 or T_H17 cells, as to secure a different transcription programme.

1.6. Research aims

As outlined above, the genome is regulated in a three-dimensional manner, with various chromatin loops and compartmental domains being partitioned and harboured in the interphase nucleus. Loop extrusion, phase separation, transcriptional elongation, the associating non-coding RNAs as well as other architectural proteins all contribute to shaping this 3D genome. Such three-dimensional genome is extensively restructured during the differentiation processes of various immune cell types, with thousands of DNA interactions emerge or disappear. Among these interactions are presumably critical cis-regulatory elements that modulate gene transcription.

Based on the existing Hi-C data generated from B and T lymphocytes in the lab (Johanson et al., 2018b), the current research aims to – by adopting various CRISPR/Cas9 methodologies described above – identify, characterise and functionally dissect any novel and important cis-regulatory elements.

Extending from the current knowledge regarding genome reconfiguration during terminal B cell differentiation, this research as well aims to characterise and stratify the genome reorganisation events during B cell activation, clonal expansion and terminal differentiation, in relation to transcription, cell division and cell cycle.

2. Materials and methods

2.1. Mice

Experiments were performed on male C57BL/6 mice of 6-12 weeks old unless otherwise specified. FUCCI transgenic mice were described previously (Sakaue-Sawano and Miyawaki, 2012; Tomura et al., 2013). All mice were maintained at The Walter and Eliza Hall Institute Animal Facility under specific-pathogen-free conditions and were randomly chosen for experiment.

2.2. Cell culture

All immune cells were cultured in RPMI 1640 with 2 mM GlutaMAX, 50 μ M β -mercaptoethanol and 10% heat-inactivated foetal calf serum (FCS) unless specified otherwise. HEK 293T cells were cultured in DMEM with 2 mM GlutaMAX and 10% heat-inactivated FCS without antibiotics.

2.3. B cells activation and differentiation

Splenocytes were obtained from male C57BL/6 mice with red cell lysis and naïve resting B cells were isolated via negative selection (Miltenyi) as per manufacturer's protocol. Isolated naïve B cells were then labelled with 10 μ M CellTrace Violet (CTV, Invitrogen), then cultured at a density of 1×10^6 cells/ml and stimulated with 25 μ g/ml LPS (*Salmonella typhosa* origin, Sigma). At corresponding time point post-activation, activated but undivided cells were sorted as TCR β - CD19+ CD69+ CTV^{highest} population on BD FACS Aria III (Table 2.1). FUCCI naïve B cells were processed as above, but sorted as CD19+ CD69+ CTV^{highest} mAG- mKO2^{hi} population. For long-term activation, negatively isolated naïve B cells were cultured at a density of 1×10^5 cells/ml and stimulated with 25 μ g/ml LPS. Expanded activated B cells (CD22+ CD138-) and plasmablasts (CD22- CD138+) were harvested on day 4 post-stimulation. Control naïve resting B cells were sorted straight from splenocytes after red cell lysis as TCR β - CD19+ B220+ IgM+ IgD^{hi} population. Experiment were repeated twice for subsequent *in situ* Hi-C and RNA-seq experiment.

Table 2.1 Antibodies used for flow cytometry

Target	Fluorochrome	Clone	Source
TCR β	PE	H57-597	BD
CD19	PE/Cy7	ID3	BD
CD19	PacB	ID3	In house
B220	PacB	RA3-6B2	In house
IgM	FITC	X-54	Miltenyi
IgD	APC	11-26c.2a	Miltenyi
CD69	APC	H1.2F3	Miltenyi
CD138	PE/Cy7	281-2	Biolegend
CD22	PE	OX-97	Biolegend
CD22	APC	OX-97	Biolegend
CD44	FITC	IM7	In house
CD8	APC	53.6.7	In house
CD4	PE	GK1.5	In house
CD25	PE/Cy7	PC61.5	eBioscience
Gata3	PE	TWAJ	eBioscience
T-bet	PE/Cy7	4B10	eBioscience

2.4. Isolation of thymic T cells

Thymic cells were harvested from male C57BL/6 mice into single-cell suspension with red cell lysis and stained with CD19, CD8, CD4, CD25 and CD44 antibodies (Table 2.1). Thymic CD4⁺ cells were obtained as CD19⁻ CD4⁺ CD8⁻, whereas thymic CD8⁺ cells as CD19⁻ CD4⁻ CD8⁺. DP cells were sorted as CD19⁻ CD4⁺ CD8⁺. DN1 population was obtained from CD19⁻ CD4⁻ CD8⁻ CD44⁺ CD25⁻ population. DN2 was sorted as CD19⁻ CD4⁻ CD8⁻ CD44⁺ CD25⁺, DN3 as CD19⁻ CD4⁻ CD8⁻ CD44⁻ CD25⁺. Finally, DN4 was acquired from CD19⁻ CD4⁻ CD8⁻ CD44⁻ CD25⁻ population (Chan et al., 2020a).

2.5. T_H1 and T_H2 cell polarisation

A flat-bottom 96-well plate was coated with 10 μ g/ml anti-CD3 (clone 145-2C11, BD) and 5 μ g/ml anti-CD28 (clone 37.51, BD) in PBS overnight at 4°C. Splenic cells were harvested from male C57BL/6 mice into single-cell suspension with red cell lysis. Naïve CD4⁺ T lymphocytes were then purified via negative selection (Miltenyi). Cells were then resuspended with cell media with corresponding polarising cytokines at a density of 1x10⁶ cells/ml and seeded into the antibody-coated plate at 100,000 cells/well. T_H1 culture medium consisted of

5 ng/ml murine IL-12 (Peprotech) and 10 µg/ml anti-IL-4 (clone 11B.11, eBioscience). T_H2 culture medium contained 50 ng/ml murine IL-4 (Peprotech), 10 µg/ml anti-IL-12 (clone C17.8, eBioscience) and 10 µg/ml anti-IFN-γ (clone XMG1.2, eBioscience). Cells were removed from anti-CD3 and anti-CD28 stimulation 48 hours later and replated in the same T_H1 or T_H2 conditions with addition of human recombinant IL-2 (Peprotech) at 30 U/ml (Chan et al., 2020a).

2.6. In situ Hi-C

In situ Hi-C was performed as previously described (Rao et al., 2014). In brief, 7x10⁵-3x10⁶ cells were resuspended with culture media at 1x10⁶ cells/ml and fixed with 1% v/v formaldehyde. Cell membrane of crosslinked cells were lysed with Hi-C lysis buffer (10mM Tris-HCl pH8.0, 10mM NaCl, 0.2% Igepal CA630, protease inhibitors), leaving nuclei intact. Pelleted nuclei were then digested with 100U of MboI (NEB) overnight and subsequently biotin-labelled with Klenow fragment (NEB). Filled-in restriction ends were ligated with T4 DNA ligase (NEB) and then subjected to sonication to shear DNA into 300-500 bp fragments. The corresponding DNA fragments were biotin-pulled down and end-repaired with T4 PNK (NEB), T4 DNA polymerase (NEB) and large fragment (Klenow) of DNA polymerase I (NEB), follows by A-tailing with Klenow fragment (3'-5' exo-; NEB) and adaptor ligation using Quick ligase (NEB). The resulting Hi-C libraries were amplified with Phusion Polymerase (Thermo) and size-selected and purified with AMPure XP magnetic beads (Beckman). Purified libraries were sequenced on an Illumina NextSeq 500 to produce 81-bp paired-end reads. Approximately 200 million read pairs were generated per sample.

2.7. RNA-seq

RNA was extracted with NucleoSpin RNA Plus (Macherey-Nagel) as per manufacturers' protocol and subsequently quantified in a TapeStation 2200 with RNA ScreenTape (Agilent). Libraries were prepared with a TruSeq RNA sample preparation kit (Illumina) from 500 ng RNA as per the manufacturers' instructions. Libraries were then amplified with KAPA HiFi HotStart ReadyMix (Kapa

Biosystems) and 200-400 bp products were size-selected and cleaned with AMPure XP magnetic beads (Beckman). Final libraries were quantified with TapeStation 2200 using D1000 ScreenTape for sequencing on Illumina NextSeq 500 platform to produce 81 bp paired-end reads. Around 25 million read pairs were generated per sample (Chan et al., 2019).

2.8. Plasmids construction

dCas9-KRAB

pHR-SFFV-dCas9-BFP-KRAB (#46911) was obtained from Addgene. pLH-U6-SpAUGRNA-mKO2 was constructed from pLH-sgRNA1-2XPP7 (Addgene# 75390). In brief, a modified version of sgRNA scaffold, with an A-U flip and hairpin extension as to confer a better stability (Chen et al., 2013), was ordered from IDT as gBlocks gene fragment and amplified with a forward primer containing an AgeI site and a BbsI cloning cassette (Table 2.2), and a reverse primer containing EcoRI site, PCR product was subsequently cloned into the corresponding site in pLH-sgRNA1-2XPP7. The hygromycin resistance gene was replaced with an mKO2 gene that has the BbsI site removed by site-directed mutagenesis (Chan et al., 2020a).

SunTag-VP64

Plasmids pHRdSV40-dCas9-10XGCN4-P2A-BFP (#60903) and pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS (#60904) were purchased from Addgene, sgRNA in this system is conferred by pLH-U6-SpAUGRNA-mKO2.

SAM

pLH-dCas9-VP64-T2A-mCherry was modified from Addgene plasmid dCas9-VP64-GFP (#61422) by exploiting NheI and EcoRI sites to replace the GFP with an mCherry gene. pLH-MS2-p65-Hsf1-P2A-BFP was adopted from Addgene plasmid MS2-p65-Hsf1-GFP (#61423) by replacing the GFP with a TagBFP gene. In this system sgRNA bearing 2 MS2 aptamers was utilised. In brief, pLH-gRNA-MS2X2-sfGFP was modified from pLH-sgRNA1-2XMS2 (#75389) by putting a

BbsI cloning cassette to the BbsI site in the plasmid and replacing the hygromycin resistance with a superfolder GFP that has the BbsI site removed by overlap extension PCR.

SunTag-TET1CD with optional activation cassette

pHRdSV40-dCas9-5xGCN4-P2A-BFP was constructed by amplifying a 5xGCN4 cassette with BamHI and NotI sites from pCAG-dCas9-5xPlat2AflD (#82560) and put into the corresponding site in pHRdSV40-dCas9-10XGCN4-P2A-BFP (#60903). pHRdSV40-scFv-GCN4-sfGFP-TET1CD was constructed by moving the sfGFP-TET1CD fragments with BamHI and NotI cuts from pCAG-scFvGCN4sfGFPTET1CD (#82561) to the matching sites in pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS (#60904). pLH-MS2-p65-Hsf1-P2A-mCherry was further modified from pLH-MS2-p65-Hsf1-P2A-BFP by replacing the TagBFP with an mCherry gene. MS2 aptamer bearing sgRNA with TagRFP657 marker was used in this system. pLH-gRNA-MS2X2-TagRFP657 was amended from pLH-gRNA-MS2X2-sfGFP by replacing sfGFP with a TagRFP657 gene from pMSCVpuro-TagRFP657 (#96939).

Lentivirus packaging

Envelope plasmid pCMV-VSVg and second-generation packaging plasmid psPAX2 was kindly given by Herold lab.

PiggyBac transposase

A hyperactive mutant of *piggyBac* transposase has been previously characterised (Yusa et al., 2011) and the plasmid clone pCMV-hyPBase was kindly provided from the Wellcome Sanger Institute.

PiggyBac transposon vectors

Vector with *piggyBac* 5'- and 3'-LTR (long terminal repeat) was obtained from Addgene (#84241). Non-coding RNA Gm13218, driven by CMV promoter, a bGH polyadenylation signal and TagBFP driven by PGK promoter were assembled into *piggyBac* transposon via Gibson assembly (Chan et al., 2020a; Gibson et al., 2009) using NEBuilder HiFi DNA Assembly (NEB). Before the actual assembly, amplified DNA fragments were subject to DpnI (NEB) digestion as to remove the

original template. From the newly assembled test plasmid, control plasmid with only TagBFP was generated by utilising BamHI sites flanking CMV promoter, Gm13218 and poly(A) signal, in which plasmid was cleaved with BamHI followed by self-ligation.

Plasmids were transformed into NEB stable or Top10 (Invitrogen) competent cells. Plasmid extraction was conducted using NucleoBond Xtra Midi kit (Macherey-Nagel).

Table 2.2 Primers for plasmid construction

Primer	Sequence	Purpose
Agel-Spscaffold-F	GAAACACCGGTGTCTTCGAGACGAAGACC CGTTTAAGAGCTATGCTGGAAACAGCATAG	KRAB
EcoRI-Spscaffold-R	ACATGAATTCAAAAAAAGCACCGACTCG	KRAB
BamHI-Kozak-mKO2-F	ATATGGATCCTTAATGCCGCCACCATGGTG AGTGTGATTAAACCAGAGATGAAG	KRAB
mKO2mut-	CGCCTTGTAAGTAGTTTTTCATTTGGCATTG	KRAB
mKO2mut+	CAAATGCCAAATGAAAACACTTACAAGGC G	KRAB
Sall-mKO2-R	AATTGTGCGACGTCGGCATCTACTTTAGGAA TGAGCTACTGCATCTTCTACCT	KRAB
NheI-T2A-F	ATTAGCTAGCGGCAGTGGAGAGGGCAGAG GAAGTCTGCTAACATG	SAM
EcoRI-mCherry-R	ATTAGAATTCTTACTTGTACAGCTCGTCCAT GCCG	SAM
NheI-P2A-F	ATTAGCTAGCGGCAGTGGAGCTACTAACTT CAGCCTGCTGAAGCAG	SAM
EcoRI-TagBFP-R	ATTAGAATTCTTAATTAAGCTTGTGCCCCAG TTTGC	SAM
BbsI R1	ACCGGTGTCTTCGAGGCTTACAGGACGAA GACCC	SAM
BbsI R2	AAACGGGTCTTCGTCCTGTAAGCCTCGAAG ACAC	SAM
BamHI-Kozak-sfGFP	ATATGGATCCTTAATGCCGCCACCATGCGT AAAGGCGAAGAGCTGTTAC	SAM
sfGFP mut-	CAGGATATTGCCGTCCTCTTTAAAGTCAAT G	SAM
sfGFP mut+	CATTGACTTTAAAGAGGACGGCAATATCCT G	SAM
Sall-sfGFP	AATTGTGCGACGTCGGCATCTACTTTATTTGT ACAGTTCATCCATACCATGCGTGATG	SAM
BamHI-5xGCN4-F	AATAGGATCCAACGGTCCGACTGACGC	TET1
NotI-5xGCN4-R	AATAGCGGCCGCGACCCGGACCCTGAGCC CCCAGAC	TET1
SADCAS9-PB-F	CTCGAGCTAGTTAAAAGTTTTGTTACTTTAT AGAAGAAATTTTGAGTTTTTG	PiggyBac-Gm13218
GM13218-CMV-R	TTCAAACCCTTAGTGACAGAGTAGTTAGCC AGAGAGCTCTGCTTATATAG	PiggyBac-Gm13218
CMV-GM13218-F	AAGCAGAGCTCTCTGGCTAACTACTCTGTC ACTAAGGGTTTGAATGTC	PiggyBac-Gm13218
BGH-GM13218-R	GCAACTAGAAGGCACAGTTTATTATATAAA CAATATTTACTAAGAGTATGATCT	PiggyBac-Gm13218
GM13218-BGH-F	TACTCTTAGTAAATATTGTTTATATAATAAAC TGTGCCTTCTAGTTGCCAGC	PiggyBac-Gm13218

PGK-BGH-R	CTAGAATGCAGTGAAAAAATGCGCCATAG AGCCCACCGCATC	PiggyBac- Gm13218
BGH-PGK-F	GCGGTGGGCTCTATGGCGCATTTTTTTCAC TGCATTCTAGTTGTG	PiggyBac- Gm13218
TAGBFP- PGK-R	TCTCCTTAATCAGCTCGCTCATGGTGGCGT CGACAAGCTTGG	PiggyBac- Gm13218
PGK- TAGBFP-F	AAGCTTGTCGACGCCACCATGAGCGAGCT GATTAAGGAG	PiggyBac- Gm13218
PB-TAGBFP-R	AAAGTAACAAAACCTTTAACTAGCTCGAGTT AATTAAGCTTGTGCCCCAGTTTGC	PiggyBac- Gm13218

2.9. gRNA design

Cas9 used herein was variant from *Streptococcus pyogenes* unless stated otherwise. The target sites for Cas9 were designed either by CHOPCHOP (Labun et al., 2019) or through IDT online design tool. Coordinates and designation of the chosen target regions are as follows (Table 2.3).

Table 2.3 Coordinates (mm10) and designation of target genomic regions

Gene	Location	Putative cis-elements (start-end)		Designation
Bcl11b	Chr12	107,158,670	107,160,300	B1
Ebf1	Chr11	45,523,740	45,526,100	E1
		45,539,100	45,540,200	E2
Gata3	Chr2	9,493,100	9,494,800	G1
		9,590,100	9,598,600	G2
		9,681,000	9,682,700	G3
		9,593,800	9,594,400	G2-1
		9,595,360	9,595,830	G2-2

2.10. gRNA cloning

For cloning target sequence into the corresponding gRNA vector, the 20 bp protospacer sequence, ordered as a pair of complementary oligos with additional nucleotides ACCG- and AAAC- at the 5' end of the sense and antisense oligos, respectively (Table 2.4), were annealed by heating at 95°C for 5 minutes and subsequent cooling to room temperature at a rate of -0.1°C/s (Chan et al., 2020a). An over-saturating amount of annealed oligos (60 pmol) were ligated to 60 ng of BbsI (NEB)-digested, column purified gRNA vector at 16°C overnight and then transformed with NEB stable competent cells (NEB). The resultant plasmids were sequence-verified.

Table 2.4 Primers for gRNA cloning

gRNA plasmid	Target	Sequence
Gm13218-K1	Gm13218	ACCGAACATTCCCATGCGATGCTC
		AAACGAGCATCGCATGGGAATGTT
Gm13218-K2		ACCGACATGGACCAACCAGGAGTG
		AAACCACTCCTGGTTGGTCCATGT
Gm13218-K3		ACCGAAAGCTTACATGGACCAACC
		AAACGGTTGGTCCATGTAAGCTTT

2.11. In vitro transcription (IVT) of sgRNA

To make a transcription template for Cas9 sgRNA, a single-strand oligo 5'-AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCC TTATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC-3' was ordered and synthesised by IDT and is acted as PCR template. Another oligo with the 20 bp target/protospacer sequence flanked between a T7 promoter at the 5' end and an invariant gRNA scaffold sequence at the 3' end are made as follows (Table 2.5):

5'-ATCCTAATACGACTCACTATAG -protospacer sequence- GTTTAAGAGCTA-TGCTGG-3'.

A high concentration (2 μ M) of T7 promoter (5'-TATCCTAATACGACTCACTA-TAG-3') and reverse primer (5'-AAAAAAAGCACCGACTCG-3') were added to the above two long oligos (0.6 μ M) and the transcription template was generated by PCR using Q5 high fidelity DNA polymerase (NEB) with 0.4 mM dNTP and an annealing temperature of 58°C. PCR product was then purified by DNA clean & concentrator-25 (Zymo).

In vitro transcription was performed by incubating 5 μ g of transcription template with 1x T7 reaction buffer (Lucigen), 1.7 mM NTP, 0.1U pyrophosphatase (Thermo), 96 U RNase Inhibitor (Lucigen) and 500 U NxGen T7 RNA polymerase (Lucigen) in a volume of 100 μ l at 37 °C for 18 hours. Transcription template was then digested by the addition of 8 U TURBO DNase (Invitrogen) and buffer in a final volume of 120 μ l at 37 °C for 45 minutes. The remaining sgRNA was then

purified by RNA clean & concentrator-25 (Zymo). The sgRNAs were also subject to digestion by alkaline phosphatase (NEB) to remove the 5'-triphosphate.

Table 2.5. Specific primers for generation of in vitro transcription templates

sgRNA	Primer sequence	Target
Bcl11b-1	TATCCTAATACGACTCACTATAGCTTCCACATT AACAAAGGAGGTTTAAGAGCTATGCTGGAAAC	B1
Bcl11b-2	TATCCTAATACGACTCACTATAGAGTGCCACTT TCTCAGCGGGTTTAAGAGCTATGCTGGAAAC	
Ebf1-5	ATCCTAATACGACTCACTATAGCAATATTTAGC TCACCTCATGTTTAAGAGCTATGCTGG	E1
Ebf11-7	ATCCTAATACGACTCACTATAGAACGGAGTGTG AGCCTAACAGTTTAAGAGCTATGCTGG	
Ebf1-9	ATCCTAATACGACTCACTATAGGTCTCCTTGAA ACCTCAACGTTTAAGAGCTATGCTGG	E2
Ebf1-11	ATCCTAATACGACTCACTATAGAAGAGCCACCA ATGTCTACGGTTTAAGAGCTATGCTGG	
Gata3-7	ATCCTAATACGACTCACTATAGTGATAGACTAC AATGGGATTGTTTAAGAGCTATGCTGG	G2
Gata3-8	ATCCTAATACGACTCACTATAGTCTACAAGTGA CATAGCCTGGTTTAAGAGCTATGCTGG	
Gata3-11	ATCCTAATACGACTCACTATAGCTATAACTCAA TGGTGGACGTTTAAGAGCTATGCTGG	G1
Gata3-13	ATCCTAATACGACTCACTATAGCTCAGCACTGG AGGGGAGATGTTTAAGAGCTATGCTGG	
Gata3-14	ATCCTAATACGACTCACTATAGAGTCTTGAATA TGGGCGCGTGTTTAAGAGCTATGCTGG	G3
Gata3-15	ATCCTAATACGACTCACTATAGATATCACAGAT CCTGATAGGTTTAAGAGCTATGCTGG	
Gm13218-18	ATCCTAATACGACTCACTATAGGAGACTAACTC TAACTGTCGTTTAAGAGCTATGCTGG	Gm13218
Gm13218-K2	ATCCTAATACGACTCACTATAGACATGGACCAA CCAGGAGTGGTTTAAGAGCTATGCTGG	G2-1
Gm13218-C2	ATCCTAATACGACTCACTATAGTCTAGTAAGTC CAGTTGCTTGTTTAAGAGCTATGCTGG	
Gm13218-C3	ATCCTAATACGACTCACTATAGACTTCGTTAAG GCATCAAGGGTTTAAGAGCTATGCTGG	G2-2
Gm13218-C5	ATCCTAATACGACTCACTATAGCTGCAGGAAAC CCACTTGTTGTTTAAGAGCTATGCTGG	

2.12. Ribonucleoprotein (RNP) assembly

For Cas9 RNP, 150 pmol of in vitro transcribed sgRNA was incubated with 100 pmol of recombinant Cas9 nuclease (IDT) at room temperature for 15 minutes. In some cases where the gRNA was chemically synthesised as crRNA and tracrRNA (IDT) format (Table 2.6), equimolar of crRNA and tracrRNA were annealed by heating at 95°C for 5 minutes and followed by cooling to room temperature at a rate of -0.1°C/s. RNP was then assembled using 120 pmol of the annealed gRNA as aforementioned.

RNP together with 100 pmol of electroporation enhancer (IDT) were then delivered into cells via electroporation.

Table 2.6. Chemically synthesised crRNA

Interacting gene	Target	Protospacer
Bcl11b	B1 (1.6 kb)	CTTCCACATTAACAAAGGAG
		GAGTGCCACTTTCTCAGCGG
	B1 (0.8 kb)	ACCAGTGGTTATGGTCCAAG
		GGGTCATGTTAGGAAGGGAA
	B1 (2.5 kb)	AGCATATGGAAACGAGAGGG
		ACAATCCATTGACTCTGGCC

2.13. Electroporation

Transfections of immune cells were performed with the 4D-Nucleofector System (Lonza). The electroporation buffers and parameters for different cell types were listed in Table 2.7. In brief, 200,000 cells were resuspended with 20 µl of buffer, materials of interest were then added to the cells and transferred to electro-cuvette strip, followed by pulsation. After electroporation cells were plated in 96-well plate and cultured in the appropriate culture media.

Table 2.7. Electroporation parameters for 4D-Nucleofector

Cell type	Buffer	Pulse code
A20	SF	FF113
EL4	SF	CM120

2.14. Screening and isolation of knockout clones

RNP transfected cells were single-cell sorted into U-bottom 96-well plate via BD FACS Aria III. Cells were then incubated at 37°C until confluent. Through a multi-channel pipette, 20-30 µl of cells were transferred to 100 µl of Direct PCR Lysis reagent (Viagen) with proteinase K (Sigma). Genomic DNA was directly extracted by heating at 55°C for 20 minutes followed by 85°C for 15 minutes to inactivate proteinase K. PCR reaction was prepared on 96-well plate by transferring 2-3 µl of cell lysate to 22 µl of mastermix comprising MyTaq Red Mix (Bioline) and primer sets (forward, reverse and internal, Table 2.8). Screening was then performed by running a thermocycle of 5 min at 95°C, followed by 35 cycles of 15s at 95°C, 15s at 60°C and 30s-60s at 72°C.

Table 2.8. Primers for screening positive clones

Loci	Target	Primers
Bcl11b	B1	CGAGGAGAAGAGAGACCTCTCCTTC
		GCCCCTTCCCTGCTCATTGAC
		CAGCAATCACCTCCCAGCAAG
	B1	TCCCTCCCTAAGCCTTGGGTC
		GCAAGCGAAGTTTCCATCTTTTAAGG
Ebf1	E1	GATGTAGGAAGTCAGATGCCGTGCTTC
		CTTTGTGTATTTACATTGCATCATTCCAG
		CATCTGTGCTGCCAGTCAGTTTACG
	E2	GCAGCCTAGCATCAAAATTGCACAGA
		GGCATGACCTCCTCTGACCTTAGTTTG
		TAGAGTACAGCGCCATCTAGCAGATG
Gata3	G1 (1 st)	AGTCAAGACCCTGAGTCAGGAGG
		CTGAAGCTCATGTTCTAGGATGGCAG
		GCAAGCCATAGAAGCCAGAACTC
	G1 (2 nd)	ACCGTTCTATGGCTTGCTGCAGGG
		ACCGCGTACTCACTCAAGACTGCA
	G2	GCAGACGTGTGTAGAAATTAACAAGAGG
		GGCAGCGAGAGCTCCTTTGC
		GAAGCATGGGACACAGCAAGGAAG
	G3	AGCCTATGTGGTGAAGTTCTAGGTC
		GAAGCATCCGACAATCTAGCTTGC
		CTTCTGTGCTTCCTGTCCAG
	G2-1	ATAATAAAAAGCTAGACTGACAAATCCCAATATAG
		AATGCAAATAAATAAAATATCCAATAGAAAATAATT
		TCATC
	G2-2	CCACAGTTAACAGGGTGCAAAAAG
		GCTAGCCTTGAACACAGTGCAC
		AGAATAAGCAAGGACAACAACTGTTGCA
Gm13218	Poly(A) inserts	ATAATAAAAAGCTAGACTGACAAATCCCAATATAG
		GCTTTATTTGTGAAATTTGTGATGCTATTGC
		AAACGAGCATCGCATGGGAATGTT

2.15. Polyadenylation signal insertion

A modified sgRNA incorporating an S1m aptamer was adopted here (Carlson-Stevermer et al., 2017). A different ssDNA template (5'-AAAAAAGCACCG-ACTCGGTGCCACTTTTTCCGAATACGAGATGCGGCCGCCGACCCGCGACT ATCTTACGCACTTG CATGATTCTGGTCGGCGGCCGCATCTCGTATTTCGAAG TTGATAACGGACTAGCCTTATTTAACTTGCTATGCTGTTTCCAGCATAGCT CTAAAC-3') was ordered through IDT and used to generate the transcription template with a forward primer (5'-ATCCTAATACGACTCACTATAGGAGACTA-ACTCTAACTGTCGTTTAAGAGCTATGCTGG-3') bearing the gRNA sequence, which targets downstream of Gm13218 transcription start site (TSS).

A biotinylated HDR repair template, bearing the SV40 polyadenylation signal, flanked by 50 bp homology and bearing phosphorothioate modifications at both ends was synthesised through IDT, illustrated as followed:

/5Biosg/G*G*CCAGGTGCTTTCTGGAGTCCTTTAAGCAGGAGACTAACTCTA ACTGTGCGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCT TTATTTGTAACCATTATAAGCTGCAATAACAAGTTAGGTAATGAAGGAGAC ATGTTGAATTGGCGAGGGTTTGAATCAAGAAC*A*C, where Biosg indicates biotinylation and * represents phosphorothioate modifications.

The S1m bearing sgRNA was assembled with Cas9. Subsequently streptavidin protein (Invitrogen) was added to the mixture and bound to the S1m aptamer. The biotinylated repair template was then added to link with streptavidin. Eventually, an RNP-streptavidin-repair template complex was created. The repair template-linked RNP was then delivered into EL4 via 4D-nucleofector.

2.16. Lentivirus production

One day prior to transfection, HEK293T cells were seeded at a density of 1.2×10^6 cells/well in a 6-well plate filled with 2 ml Opti-MEM (Invitrogen), 2mM GlutaMAX, 1mM Sodium Pyruvate and 5% FCS. Transfection of HEK293T was performed using Lipofectamine 3000 (Invitrogen) as per the manufacturers' instructions (Chan et al., 2020a). Cells were co-transfected with packaging plasmids (pCMV-VSV-g and psPAX2) at 0.17 pmol each and around 0.20-0.23 pmol transfer construct to make up a final mass of 3.0-3.3 μ g. Fresh packaging Opti-MEM media was replaced 6 hours after transfection. Virus was harvested 24- and 52-hour post-transfection and subject to concentration by using Lenti-X concentrator (Clontech) or Amicon Ultra columns (Millipore). Virus was aliquoted and stored in -80°C .

2.17. Transduction

Transduction was performed in a 12-well plate, with 500,000 cells resuspended in 1 ml viral supernatant supplemented with 8 μ g/ml polybrene (Millipore) (Chan et al., 2020a). Viral supernatant is subject to dilution with Opti-MEM packaging media to adjust the titre. Cells were spun at 2500 rpm at 32°C for 90 minutes. Media was replaced thereafter, or after overnight incubation at 37°C , with regular RPMI media. FACS was performed on the transduced cells after 48 hours to enrich for stable expression.

2.18. Droplet digital PCR (ddPCR)

In validating efficiency of large deletion, primer-probe sets were designed to target both the region of interest (FAM probe) and a region outside (HEX probe) as a reference (Table 2.9). Genomic DNA was extracted from edited cells by mixing cells with 60 µl of QuickExtract DNA extraction solution (Lucigen) and incubating at 65°C for 10 minutes followed by 98°C for 5 minutes. Lysate (2 µl) was mixed with 900 nM of primers, 250 nM of probes and ddPCR Supermix (Bio-Rad). The mixture was emulsified into a 96-well plate with droplet generation oil (Bio-Rad) using an automated droplet generator (Bio-Rad) according to the manufacturer's protocol. PCR was performed as followed: 10 min at 95°C, 40 cycles of 30s at 94°C followed by 1 min at 60°C, finally with 10 min at 98°C. The plate was then transferred to a QX200 droplet reader (Bio-Rad) to quantify the positive droplets and from that the copy number of targets were calculated by the QuantaSoft software (Bio-Rad).

Table 2.9. Droplet digital PCR primers and probes

Region	Target	Primer/Probe
Bcl11b	B1	TCGCTCAGTCACATCTCAAC
		GTCATTCACTAGCCCACAGG
		/56-FAM/AAGACACTT/Zen/CACAGTTCCAGCCTTCC/3IABkFQ/
	Ref.	AGCTTTCAGTTCCCAAGAGG
		CTGTCTAGAATCCAACTCCCC
		/5HEX/AGTGGCCTT/ZEN/CTCCGTGACATGTT/3IABkFQ/
Gata3	G2-1	GCTTTTGAGTGTTCCCATCAC
		TGGCTGAGTCGCTGAAAAG
		/56-FAM/CCGCTCTCC/Zen/TCAAGACCACAATCTAT/3IABkFQ/
	G2-2	GCCTAACAAGTGGGTTTCCTG
		CTCCTTTTGGTTGCCTTTGTAG
		/5HEX/AGCGTTGCC/Zen/CTGACTCTCACC/3IABkFQ/

2.19. T7 endonuclease I (T7EI) assay

To validate the cleavage efficiency of a sgRNA, high fidelity PCR was performed to amplify the targeted region to generate amplicon that was <1kb in size. Amplicon was heated in NEB Buffer 2 to 95°C for 5 minutes and then reannealed back by cooling to 22°C at a rate of -0.1°C/s. Assay was then conducted by adding 1 µl T7EI (NEB) to the reannealed DNA and incubate at 37°C for 30 minutes. Appearance of smaller bands indicates the occurrence of cleavage and the intensity of that corresponds the cleavage efficiency.

2.20. End point reverse transcription PCR (RT-PCR)

RNA was extracted using NucleoSpin RNA Plus (Macherey-Nagel) with gDNA removal, 1000 ng of that was then reverse transcribed with random hexamer and anchored oligo dT primer using SensiFast cDNA synthesis kit (Bioline). Expression of Gm13218 was confirmed by subsequent PCR with primers 5'-GAAGCATGGGACACAGCAAGGAAG-3' and 5'-TAAATATAACAATATTTAC-TAAGAGTATGATCTTCTGAGTTAGGCTG-3'.

2.21. Quantitative reverse transcription PCR (RT-qPCR)

RNA was extracted using NucleoSpin RNA Plus (Macherey-Nagel) with gDNA removal, 1000 ng of that was then reverse transcribed with random hexamer and anchored oligo dT primer using SensiFast cDNA synthesis kit (Bioline). Quantitative PCR was performed on the diluted cDNA using SensiFast probe (Bioline) as per manufacturers' protocols. In other circumstance 20 ng RNA was directly used in one step RT-qPCR by utilising iTaq Universal probe supermix (Bio-Rad). In both cases the β -actin gene would act as a reference (Table 2.10). Gene expression was normalised to the endogenous control gene β -actin and evaluated using $\Delta\Delta CT$ method.

Table 2.10. RT-qPCR primers and probes

Gene	Primers/Probe
Actb	GACTCATCGTACTCCTGCTTG
	GATTACTGCTCTGGCTCCTAG
	/56-FAM/CTGGCCTCA/ZEN/CTGTCCACCTTCC/3IABkFQ/
Bcl11b	ACCCTGGAGAAACACATGAAA
	AAGGTTGGCGATGGTCAC
	/56-FAM/CCTCTCAGC/ZEN/CTGCTCGATTTTGACA/3IABkFQ/
Ebf1	GGAATGACCTTCTGTAACTCTG
	TCGTACAAGTCCAAGCAGTTC
	/56-FAM/CTACACAGC/ZEN/ACTCAATGAACCCACCA/3IABkFQ/
Gata3	GTCCCCATTAGCGTTCCTC
	CCTTATCAAGCCCAAGCGAA
	/56-FAM/TGTCCCTGC/ZEN/TCTCCTTGCTGC/3IABkFQ/
Gm13218	CTTTGCTCTTCACTCTGGACT
	CTGTGTCCCATGCTTCTCTG
	/56-FAM/TCAAAGCCC/Zen/CAGATTCTTCATCTCC/3IABkFQ/

2.22. 5' RACE (Rapid amplification of cDNA ends)

RNA of EL4 cells was extracted using NucleoSpin RNA Plus (Macherey-Nagel) with gDNA removal, 7.5 µg of RNA was subjected to dephosphorylation by treatment with 30U calf intestinal alkaline phosphatase (CIP, NEB) and 40U RNase inhibitor (Lucigen) through an incubation at 37°C for 90 minutes. Dephosphorylated RNA was cleaned up by RNA clean & concentrator-25 (Zymo). Decapping of capped RNA was conducted by adding 25U RNA 5' Pyrophosphohydrolase (RppH, NEB) and 40U RNase inhibitor in ThermoPol buffer (NEB), followed by incubation at 37°C for 2 hours. Decapped RNA was then purified. RACE DNA-RNA chimeric adapter (5'-GTCGCACGGTCCATCGCAGCAGTCACArUrArGrArArA-3') was ligated to the decapped RNA by adding 5 µg of adapter, 40U T4 RNA Ligase 1 (NEB), 0.1 mM ATP and 40U RNase inhibitor, followed by incubation at 17°C for 16 hours, RNA clean- up and quantified.

The adapter-ligated RNA was reverse transcribed by using Superscript IV (Invitrogen) with 50 pmol anchored oligo dT primer, 0.5 mM dNTP, and incubating at 50°C for 20 minutes followed by 80°C for 10 minutes.

A strategy known as step-out PCR was employed in retrieving unknown 5' sequence. In brief, 0.02 pmol of overhang primer specific to RACE adapter (5'-ACGCTGACGCTGAGCCTACC-TGACGTCGCACGGTCCATCGCAGCAGTC-3') was used together with 0.15 pmol of overhang-specific primer (5'-ACGCTGACGCTGAGCCTACCTGAC-3') and 0.15 pmol of gene specific primer (5'-CTAAGAGTATGATCTTCTGAGTTAGGCTGTAGAATTTGAATCT-AG-3') to amplify RNA of interest. PCR was performed using Phusion HS (Thermo) with thermocycle of 98°C for 30s, followed by 35 cycles of 98°C for 10s and 72°C for 1 minute. Enzymatic clean-up of PCR amplicon (5 µl) was performed by adding 10U exonuclease I (Exo I, NEB) and 1U shrimp alkaline phosphatase (rSAP, NEB), followed by incubation at 37°C for 15 minutes and 80°C for another 15 minutes. A nested PCR using a second gene specific primer (5'-CTTGCTGTGTCCCATGCTTCTCTGTAAAATGTGGATGATACCATAGA-AG-3') was performed on the diluted, clean-up PCR product from the first round.

RACE product was cloned and amplified with CloneJET PCR cloning kit (Thermo) and sequence verified.

2.23. 3' RACE

RNA of EL4 cells was extracted using NucleoSpin RNA Plus (Macherey-Nagel) with gDNA removal, 1 µg of RNA was reverse transcribed using Superscript IV (Invitrogen) with an overhang oligo dT primer (5'-GTCGCACGGTCCATCGCAG-CAGTCCATGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3').

Same step-out PCR strategy as in 5' RACE was adopted with the following gene specific primers: 5'-GAAGCATGGGACACAGCAAGGAAG-3' for 1st round, and 5'-CTAGATTCAAAT-TCTACAGCCTAACTCAGAAGATCATACTCTTAG-3' for nested amplification.

RACE product was cloned and amplified with CloneJET PCR cloning kit (Thermo) and sequence verified.

2.24. Antisense oligonucleotide (ASO) mediated knockdown

Antisense oligos were designed to target different exons, together with negative control ASO (Table. 2.11), they were all ordered as LNA (locked nucleic acid) GapmeRs (Qiagen) (Chan et al., 2020a). ASOs were reconstituted in concentration of 50 µM and 4 ASOs targeting Gm13218 were pooled together at a final reaction concentration of 6 µM, whereas negative control ASO was used alone at 6 µM. Electroporation was performed in EL4 via 4D-Nucleofector with two pulsation, in which the second pulsation was conducted 48 hours after. RNA was extracted on day 3.

Table 2.11. GapmeRs antisense oligos

LNA GapmeR	Sequence	Targeted exon
Gm13218-3030	GAGTATGATCTTCTGA	3
Gm13218-3032	AGCGTAGGGTAGAATT	2
Gm13218-3033	GCGACATTCAAACCCT	1
Gm13218-3034	AGGTAATGAAGGAGAC	1

2.25. *PiggyBac* non-coding RNA overexpression

Stable cell line expressing Gm13218 was generating by co-transfecting 1 µg of transposon plasmid together with 1 µg *piggyBac* transposase plasmid into cells using 4D-Nucleofector with the corresponding pulsation parameters. In control group transposon with TagBFP alone was used. Selection was done by sorting BFP positive cells.

2.26. Hi-C pre-processing

The sequence from each sample was aligned to the mouse genome (mm10) with diffHic package (v1.14.0) (Lun and Smyth, 2015) which uses cutadapt v0.9.5 (Martin, 2011) and bowtie2 v2.2.5 (Langmead and Salzberg, 2012) for alignment. The BAM file generated was then sorted by read name. The FixMateInformation command from the Picard suite v1.117 was applied, duplicate reads were marked and then re-sorted by name. Read pairs were determined to be dangling ends and removed if the pairs of inward-facing reads or outward-facing reads on the same chromosome were separated by less than around 1000 bp for inward-facing reads and around 10000 bp for outward-facing reads. Read pairs with fragment sizes above 1000 bp were removed. An estimate of alignment error was obtained by comparing the mapping location of the 3' segment of each chimeric read with that of the 5' segment of its mate. A mapping error was determined to be present if the two segments were not inward-facing and separated by less than 1000 bp, and around 1-3% were estimated to have errors (Chan et al., 2019).

2.27. Detecting differential interactions (DIs)

Differential interactions (DIs) between different libraries were detected using the diffHic package v1.16.0. Read pairs were counted into 50 kbp bin pairs for all autosomes. Bins were discarded if the count is less than 10, or contains blacklisted genomic regions as defined by ENCODE for mm10 (Consortium, 2012) or if they were within a centromeric or telomeric region. Filtering of bin-pairs was performed using the filterDirect function, where bin pairs were only retained if they have average interaction intensities 4-fold higher than the background

ligation frequency. The ligation frequency was estimated from the inter-chromosomal bin pairs from a 1 Mbp bin-pair count matrix. Diagonal bin pairs were also removed. The counts were normalised between libraries using a loess-based approach with bin pairs less than 1.5 Mbp from the diagonal normalised separately from other bin pairs. Tests for DIs were carried out using the quasi-likelihood (QL) framework (Chen et al., 2016; Lund et al., 2012) from the edgeR package v3.26.5 (Robinson et al., 2010). A design matrix was constructed using a one-way layout that specified the cell group to which each library belong. A mean-dependent trend was fitted to the negative binomial dispersions with the estimateDisp function. A generalised linear model (GLM) was fitted to the counts for each bin pair (McCarthy et al., 2012), and the QL dispersion was estimated from the GLM deviance with the glmQLFit function. The QL dispersion were then squeezed toward a second mean-dependent trend, using a robust empirical Bayes strategy (Phipson et al., 2016). A p-value was computed against the null hypothesis for each bin pair using the QL F-test. P-values were adjusted for multiple testing using the Benjamini-Hochberg method. A DI was defined as a bin pair with a false discovery rate (FDR) below 5%. DIs adjacent in the interaction space were aggregated into clusters using the diClusters function to produce clustered DIs. DIs were merged into a cluster if they overlapped in the interaction space, to a maximum cluster size of 500 kbp. The significance threshold for each bin pair was defined such that the cluster-level FDR was controlled at 5%. Cluster statistics were computed using the csaw package v1.18.0 (Lun and Smyth, 2016). Overlaps between unclustered bin pairs and genomic intervals were performed using the InteractionSet package (Lun et al., 2016) (Chan et al., 2019).

2.28. Detecting TAD boundaries

TADs were detected with the TADbit v0.2.0.5 python-based software (Serra et al., 2017). Read pairs were counted into 50 kbp bin pairs (with bin boundaries rounded up to the nearest Mbol restriction site) using the squareCounts function of diffHic with no filter. A count matrix containing a read pair count for each bin pair in each library was resulted and was then converted into a contact matrix for each somatic chromosome with the inflate function of the InteractionSet package v1.12.0. Replicate contact matrices were summed. TADs were detected for each

chromosome with the function `find_tad` from the `TADbit` package. Only TADs boundaries with score higher than 7 were included in the results (Chan et al., 2019).

2.29. Detecting differential TAD boundaries

Differential TAD boundaries between all Hi-C libraries were detected using the `diffHic` package (Lun and Smyth, 2015) using the approach described in Chapter 8 of the `diffHic` User's Guide. This approach adapts the statistical strategy recently described for differential methylation (Chen et al., 2017b) to identify DTBs. The strength of a putative TAD boundary was assessed based on the up-stream vs down-stream intensity contrast at that genomic loci, defined as the ratio of up-stream to down-stream HiC reads anchored at that genomic region. `edgeR` was used to test whether the ratio at each locus significantly increased or decreased in absolute size between B cell stages. This method directly assesses differential boundary strength relative to biological variation without needing to make TADs calls in individual samples. Up-stream and down-stream read counts were determined for 100 kbp genomic regions and for 1 Mbp up- and downstream. Low abundance regions with average log2-counts per million below 1 were removed. Tests for DTBs were performed using the QL framework of `edgeR`. Tests were conducted with `edgeR` `glmTreat` relative to a fold-change threshold of 1.1 in order to prioritize larger fold-changes (McCarthy and Smyth, 2009). Regions with `Treat FDR` below 0.05 were considered to be DTB (Chan et al., 2019).

2.30. Detection of A/B compartments

A or B compartments were identified across all the stages of B cell activation at a resolution of 100 kb with the `HOMER` Hi-C pipeline as described by Lieberman-Aiden et al. (2009). After processing with the `diffHic` pipeline, libraries were converted to the Hi-C summary format in R. Then input-tag directions were created for each library with the `makeTagDirectory` function, with the `mm10` genome and `Mbol` restriction recognition site (`GATC`) specified. Biological-replicate tag directories for each cell type were summed. The `runHiCpca.pl`

function was used on each library with `-res 100000` and the `mm10` genome specified to perform principal component analysis to identify compartments. To identify changes in A/B compartments between libraries, the `getHiCcorrDiff.pl` function was used to directly calculate the difference in correlation profiles (Chan et al., 2019).

2.31. Detecting differentially interacting promoters (DIPs)

Differentially interacting promoters were detected using across all libraries with the `diffHic` package (Lun and Smyth, 2015). Gene promoters were defined with the genes from the `TxDb.Mmusculus.UCSC.mm10.knownGene` package v3.4.7 and by applying the `promoters` function from the `GenomicFeatures` package v1.36.2 with `upstream=5kbp` and `downstream=5kbp`. Interactions between the promoters and the entire genome was counted with the `connectCounts` function from `diffHic` with `filter` set to 0 and the `second.region=10kbp`. Interchromosomal interactions were excluded along with interactions contained in blacklisted regions as defined by ENCODE, centromeres or telomeres. Interactions were filtered by abundance as a function of distance between the anchors. Using the `loessFit` function from the `limma` package v3.40.2 (Ritchie et al., 2015) with `span 0.05`, a loess curve was fitted to the average log counts per million (logCPM) for all libraries (calculated with the `cpmByGroup` function of `EdgeR` with `log=TRUE`) as a function of $\text{distance}^{0.25}$. An interaction was then required to have an abundance larger than the fitted curve plus two times the mean of the absolute values of the residuals from the loess fit. For each unique promoter, interactions were then aggregated to produce a count matrix. Low-abundance promoters were filtered using `edgeR`'s `filterByExpr` function with `min.count=200` and `min.totals=200`. Obsolute Entrez Gene Ids were removed, as were mitochondrial, ribosomal (rRNA/ncRNA), Riken, olfactory and X or Y chromosome genes. The counts were normalized between libraries using a loess-based approach with the `normOffsets` function from the `csaw` package v1.18.0 (Chan et al., 2019).

2.32. Detecting correlation between structures and gene expression

Two different DIPs analyses were performed; with and without the FUCCI B cells HiC libraries. DIPs were detected with the quasi-likelihood (QL) framework (Chen et al., 2016; Lund et al., 2012) of the edgeR package. A design matrix was constructed with a one-way layout that specified the cell type. Using the promoter counts, the estimateDisp function was used to maximise the negative binomial likelihood to estimate the common dispersion across all promoters with trend=none and robust=TRUE (Chen et al., 2014b). A generalized linear model (GLM) was fitted to the counts (McCarthy et al., 2012) and the QL dispersion was estimated from the GLM deviance with the glmQLFit function with robust=TRUE and trend.method=none. The QL dispersions were then squeezed toward a second mean-dependent trend, with a robust empirical Bayes strategy (Phipson et al., 2016) to share information between genes. A P-value was calculated for each promoter using a moderated t-test with glmQLFTest. The Benjamini–Hochberg method was used to control the false discovery rate (FDR) below 5%. Heatmaps of the filtered and normalized logCPM value were plotted with the coolmap function from the limma package. Mean-difference plots (MD plots) were plotted with the plotMD function (Chan et al., 2019).

Looping interactions were detected in the naïve B cell libraries with a method similar to that described by Rao et al. (2014) with a bin size of 20 kbp. Established loops were defined as those with enrichment values above 0.5, with an average count across libraries greater than 5, and that were more than 60 kbp away from the diagonal. Loops were annotated with genes in the anchors with annotatePairs function from the diffHic package. The significance of the enrichment of DE genes between naïve B cells and 3 hours after activation (without logFC threshold) in the loops was performed with a pearson's chi-squared test (df=2). The test was between the portions of genes within/not within loop anchors that are: DE genes increasing, DE genes decreasing or non-DE genes (Chan et al., 2019).

2.33. Visualisation of Hi-C

Multidimensional-scaling plots were constructed with the plotMDS function in the limma package applied to the filtered and normalised logCPM values of each bin pair or promoter for each library. The distance between each pair of samples was the “leading log fold change”, defined as the root-mean-square average of the 500 largest $\log_2$ fold changes between that pair of samples (Chan et al., 2019).

Normalized contact matrices at a 50 kbp resolution were produced with the HOMER HiC pipeline for visualization. With the summed biological replicate tag directories, the analyzeHiC function was used with the matrix balance option which iteratively balances matrices to ensure that the total interactions are the same for each column and row. Plaid plots were constructed using the contact matrices and the plotHic function from the Sushi R package v1.22.0 (Phanstiel et al., 2014). The colour palette was inferno from the viridisLite package v0.3.0 (Garnier, 2018) and the range of colour intensities in each plot was scaled according to the library size of the sample. The plotBedpe function of the Sushi package was used to plot the unclustered DIs as arcs where the z-score shown on the vertical access was calculated as $\log_{10}(\text{p-value})$ (Chan et al., 2019).

2.34. Differential expression analysis

RNA-seq Reads were aligned to the mm10 mouse genome with Rsubread package v1.28.0. Genewise counts were obtained with featureCounts and Rsubread’s inbuilt Entrez gene annotation. Low-abundance genes were filtered out with the filterByExpr functions. Obsolete Entrez Gene IDs were removed, as were mitochondrial, ribosomal, X or Y chromosomal genes and variable immunoglobulin gene segments. Normalisation was performed with the trimmed mean of M-values (TMM) method (Chan et al., 2019; Robinson and Oshlack, 2010).

The differential expression analysis was conducted using the quasi-likelihood (QL) framework (Chen et al., 2016; Lund et al., 2012) of edgeR. The empirical Bayes procedure of glmQLFTest was run in robust mode to increase power and protect

against hypervariable genes (Phipson et al., 2016). Differential expression was evaluated using edgeR's glmTreat function with log-fold-change threshold set to 1.5 (McCarthy and Smyth, 2009). The Benjamini-Hochberg method was used to control the FDR below 5%. Log2 RPKM (reads per kilobase per million) were computed with edgeR's rpkm function. MDS plots, MD plots and heatmaps were created using the plotMDS, plotMD and coolmap functions. Distances on the MDS plots correspond to leading log2-fold-changes from the top 500 differentially expressed genes (Chan et al., 2019).

2.35. Gene set enrichment

Tests for over-representation of gene ontology terms was performed with the goana function from the limma package. Gene set enrichment was tested using limma's fry function and visualised using limma's barcode enrichment plot. Terms containing less than 100 genes were removed. Categories were ranked according the percentage of genes within the category that are differentially expressed. Terms in the CC ontology category were also removed (Chan et al., 2019).

3. Exploring and identifying functional cis-regulatory elements of lymphocytes

It is well established that the transcription of many genes is regulated in a three-dimensional manner, in which a gene promoter can be simultaneously interacting with multiple cis-elements confined in the same TAD, be it a CTCF-dependent loop domain or CTCF-independent compartmental domain (chapter 1.1, 1.2) (Bonev and Cavalli, 2016; Szabo et al., 2019). In contrast, elements across TADs are largely shielded from crosstalk by TAD boundaries (Dixon et al., 2012; Fudenberg et al., 2016; Rao et al., 2014; Sanborn et al., 2015; van Steensel and Furlong, 2019). CTCF sites and housekeeping genes are enriched at these boundaries and hence they are important to enact proper TAD insulation. As such, TADs form a gene regulatory unit to contain appropriate enhancer-promoter interactions. Although TADs appear to be largely conserved between cell types, certain TADs display lineage specificity (Bonev et al., 2017; Johanson et al., 2018b; Pekowska et al., 2018; Rao et al., 2014; Stadhouders et al., 2018). It is therefore conceivable that these lineage-restricted TADs harbour distinct cis-elements that are critical in driving the transcription of the underlying genes, and thereby influence the development and divergence of different cell lineages.

The immune system consists of a large variety of functionally and phenotypically distinct cell types that differentiate from a common progenitor. This functional diversity arises from the divergence of transcriptional programmes, which is tightly linked to the unique chromatin topology possessed by each cell type, including lineage-specific chromatin loops, TADs and compartments. Previous work from our lab and that of others have constructed Hi-C 3D contact maps for various immune cell types, which were revealed to individually exhibit distinct genome organisation (Bunting et al., 2016; Hu et al., 2018; Javierre et al., 2016; Johanson et al., 2018b; Lin et al., 2012). With the utilisation of the diffHic package (Lun and Smyth, 2015), comparison between different immune lineages allowed the identification of thousands of differential interactions (DIs) (Johanson et al., 2018b). Many of these DIs often cluster as TADs of sub-megabase scale and coincide with lineage-specific genes. Therefore, further analyses of these

lineage-restricted TADs could potentially unveil essential and novel cis-regulatory elements.

Initially the cis-regulatory elements that were identified were generally those that display species conservation and DNase I hypersensitivity, with their functions studied either through *in vitro* reporter assays or transgenesis. The elements of interest would generally be cloned, individually or in combination, into plasmids harbouring functional reporter genes that are easy to be assayed, for example luciferase, chloramphenicol acetyltransferase (CAT), β -galactosidase or GFP (Schenborn and Groskreutz, 1999). On the other hand, *in vivo* studies generally involved microinjections of these reporters in plasmids or bacterial artificial chromosomes (BAC) into zygotes. Many important cis-regulatory elements in the immune system were characterised in this way, including multiple enhancers of *CD8* (Ellmeier et al., 1997), silencer of *CD4* (Donda et al., 1996; Sawada et al., 1994; Siu et al., 1994; Taniuchi et al., 2002a; Taniuchi et al., 2002b) and a multicomponent cis-regulatory system regulating *Acida* (Activation-induced cytidine deaminase) (Tran et al., 2010). While robust, these approaches require extensive cloning and generation of mouse lines, and also tend to disregard the influences contributed by local chromatin landscape and topology. With the recent adoption and development of CRISPR (Garneau et al., 2010), it becomes a very convenient method to functionally dissect cis-regulatory elements, simply by deleting them directly in the genome.

In this chapter I focussed on defining the functionally important B- or T-cell specific cis-elements, with a focus on accessible chromatin regions that are exerting functions at a long distance from the gene promoters.

3.1. Results

Immune lineage-restricted TADs coincide with lineage-specific genes

In order to uncover potential cis-regulatory elements from our Hi-C data, we identified and ranked genomic regions bearing the most extensive clustering of DIs (Johanson et al., 2018b) and overlaid that with the top differentially expressed genes (DEGs) retrieved from the Immunological Genome (Immgen) Project (Heng et al., 2008). From the top 100 such DEGs, visualisation of some select DIs around DEGs reveals the existence of elaborate chromatin structures clustered regionally and manifested as TADs of sub-megabase in size. As expected, both B- and T-cell specific TADs were identified. The select B cell specific TADs are located around *Bcl11a*, *Ebf1* and *Mef2c* loci (Fig. 3.1), whereas T cell specific TADs rest upon *Bcl11b*, *Gata3* and *Ppm1h* loci (Fig. 3.2). Despite the large size of these TADs, all but *Ppm1h* comprise no other protein-coding genes.

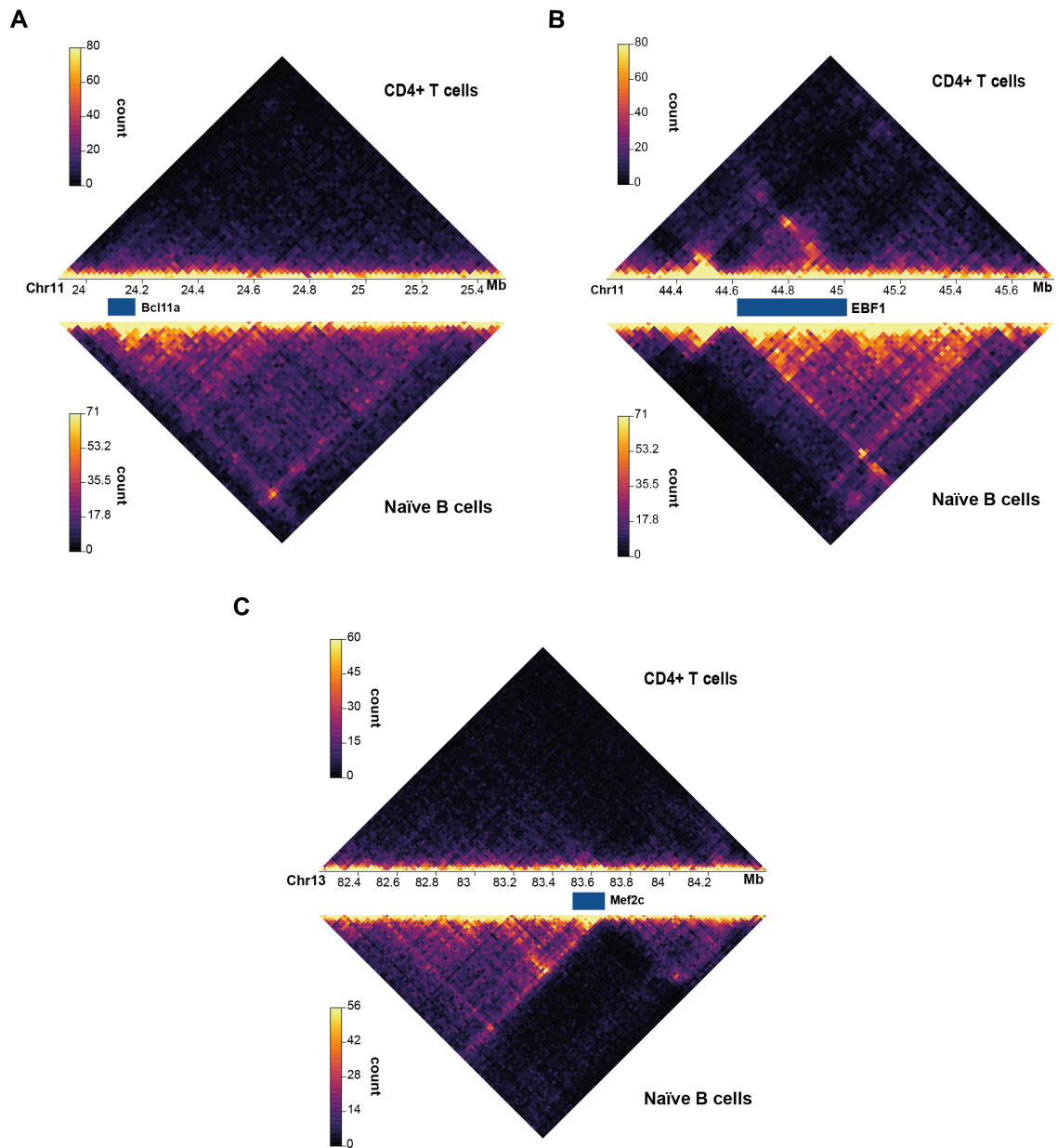


Figure 3.1. B cell specific TADs.

Hi-C contact matrices around A) *Bcl11a*, B) *Ebf1* and C) *Mef2c* loci between murine splenic naïve CD4+ T and B cells. Bar in the middle denotes the locus. Coordinates are in mm10.

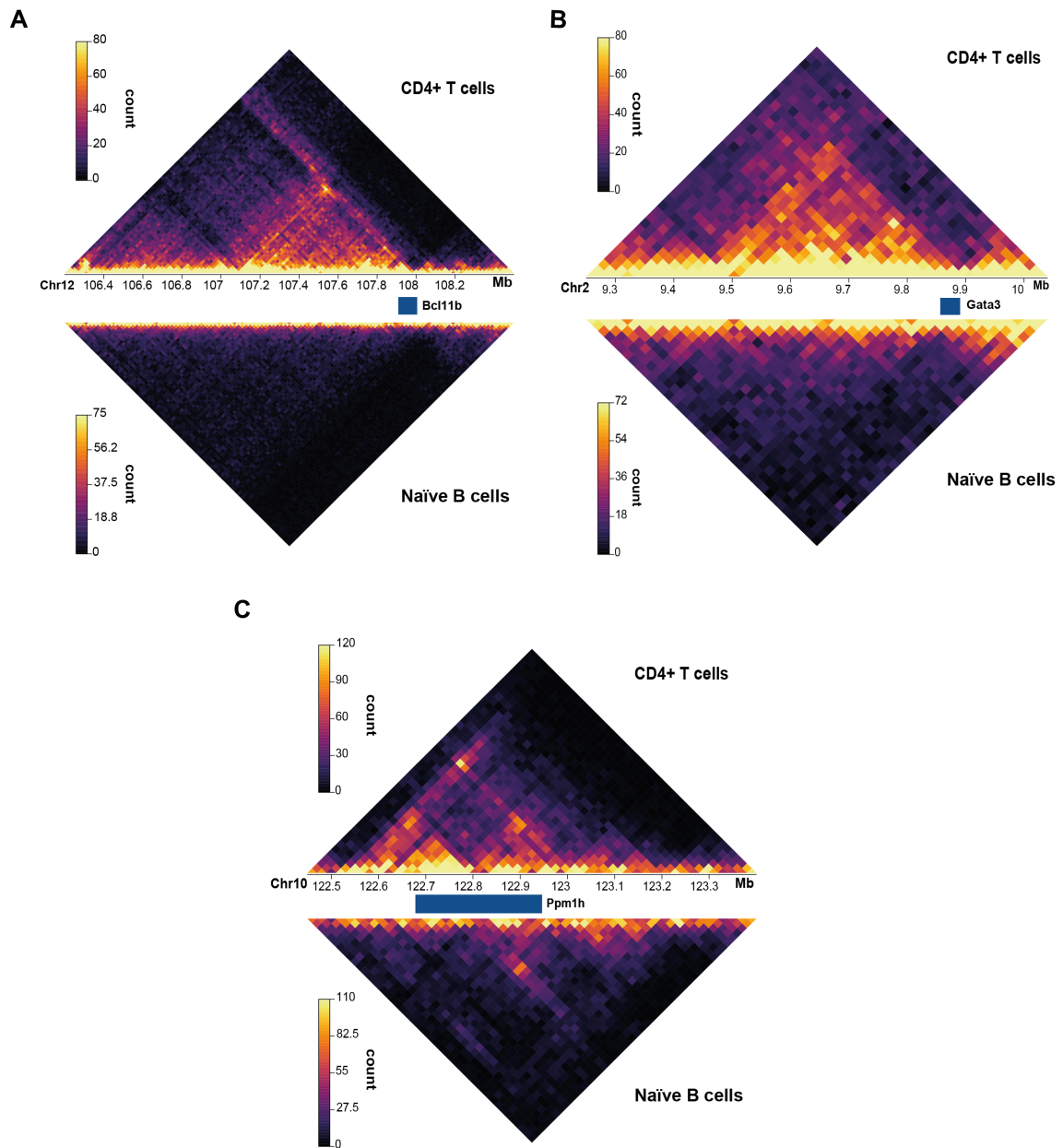


Figure 3.2. T cell specific TADs.

Hi-C contact matrices around A) *Bcl11b*, B) *Gata3* and C) *Ppm1h* loci between murine splenic naïve CD4+ T and B cells. Bar in the middle denotes the locus. Coordinates are in mm10.

Differential chromatin accessibility profiles enable discoveries of novel cis-elements

To assist in more accurately identifying enhancer regions, the Hi-C contact matrices were further overlaid with chromatin accessibility profiles of various immune subsets inferred by ATAC-seq data retrieved from the Immgen Project (Yoshida et al., 2019). These open chromatin regions can be bound by CTCF or other transcription factors, and thereby maintain TAD integrity or modulate gene transcription. In some lineage-specific TADs, there are often multiple accessible putative cis-elements with some coinciding with the corner dots – presumably indicating the presence of a stable, persistent loop – in the contact matrices. For instance, in the *Ebf1* and *Bcl11b* loci (Table 3.1, Fig. 3.3A, B). These loop anchors are therefore depicted as candidates for further functional dissections. In contrast, *Gata3* TAD lacks a prominent loop but instead there are several clusters of cis-elements interacting with *Gata3* TSS (Table 3.1, Fig. 3.3C). Three such regions (G1, G2 and G3) were selected for subsequent studies.

Table 3.1. Lineage-specific TADs and the putative cis-regulatory elements.

Gene	Location	Hi-C map span (mm10)	Putative cis-elements (start-end)	Designation
<i>Bcl11b</i>	Chr12	106,200,000-108,500,000	107,158,670-107,160,300	B1
<i>Ebf1</i>	Chr11	44,150,160-45,749,670	45,523,740 - 45,526,100	E1
			45,539,100 - 45,540,200	E2
<i>Gata3</i>	Chr2	9,250,000-10,300,000	9,493,100 - 9,494,800	G1
			9,590,100 - 9,598,600	G2
			9,681,000 - 9,682,700	G3

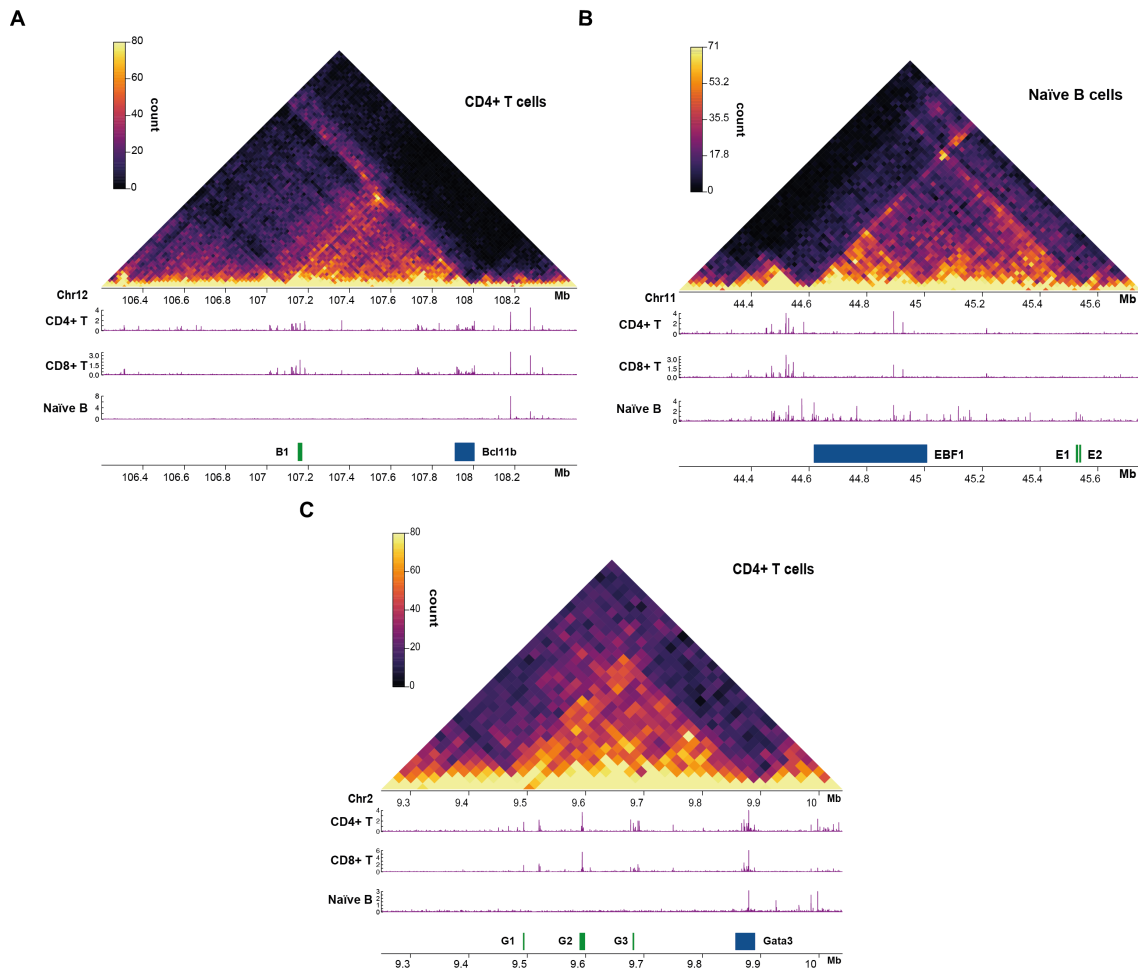


Figure 3.3. Differential accessibility profiles allow identification of potential cis-regulatory elements.

ATAC-seq accessibility profiles of splenic naïve B, CD4+ and CD8+ T cells along A) *Bcl11b*, B) *Ebf1* and C) *Gata3* loci overlaying with Hi-C contact matrices of splenic CD4+ naïve T or B cells. Blue bars at the bottom denote the loci, whereas green bars indicate the potential cis-regulatory elements.

Genome structures are preserved in lymphoma cell line EL4 and A20

It is well known that the reading frame of a protein-coding gene can be easily interfered by small indels, very often introduced after the repair of a single DSB. However, disrupting the aforementioned putative cis-regulatory elements appears to be more problematic than disrupting protein-coding genes. This is due to the transcription factor binding sites of cis-regulatory elements are scattered across a relatively wider genetic span, and more importantly they suffered from degeneracy – transcription factors would bind to a range of similar DNA sequences with equal or varying potencies (Berg and von Hippel, 1987; Bulyk, 2003; Dermitzakis and Clark, 2002; Kotelnikova et al., 2005; Weirauch et al., 2013). Thus, perturbation of such probably requires the removal of the entire element. Such large deletions can still be accomplished, albeit at a much lower efficiency than the introduction of indels, by imposing two DSBs simultaneously (Canver et al., 2014). Due to this apparent lower efficiency, as well as the lack of obvious phenotypic effect at the cell surface, manipulation of such elements is therefore technically challenging to perform in primary lymphocytes. As such, I chose cell lines that would be more favourable for this initial interrogation, with T cell lymphoma line EL4 for studying T cell specific interaction around *Bcl11b* and *Gata3* loci, and B cell line A20 for B cell specific structures at *Ebf1* locus.

Although these cell lines are considered similar to their primary cell counterparts, it is not known if the transformation of the cell lines would have altered the genome architecture around those regions. I therefore conducted Hi-C on EL4 and A20 and compared them to naïve CD4⁺ T and B cells, respectively. In the Hi-C contact matrices, it was revealed that the EL4 possesses similar genome topology to CD4⁺ T cells around *Bcl11b* and *Gata3* loci (Fig. 3.4A, B). Likewise, A20 displays similar architecture to resting B cells around *Ebf1* locus (Fig. 3.4C). The preservation of structures thus renders the usage of cell lines feasible for this initial interrogation.

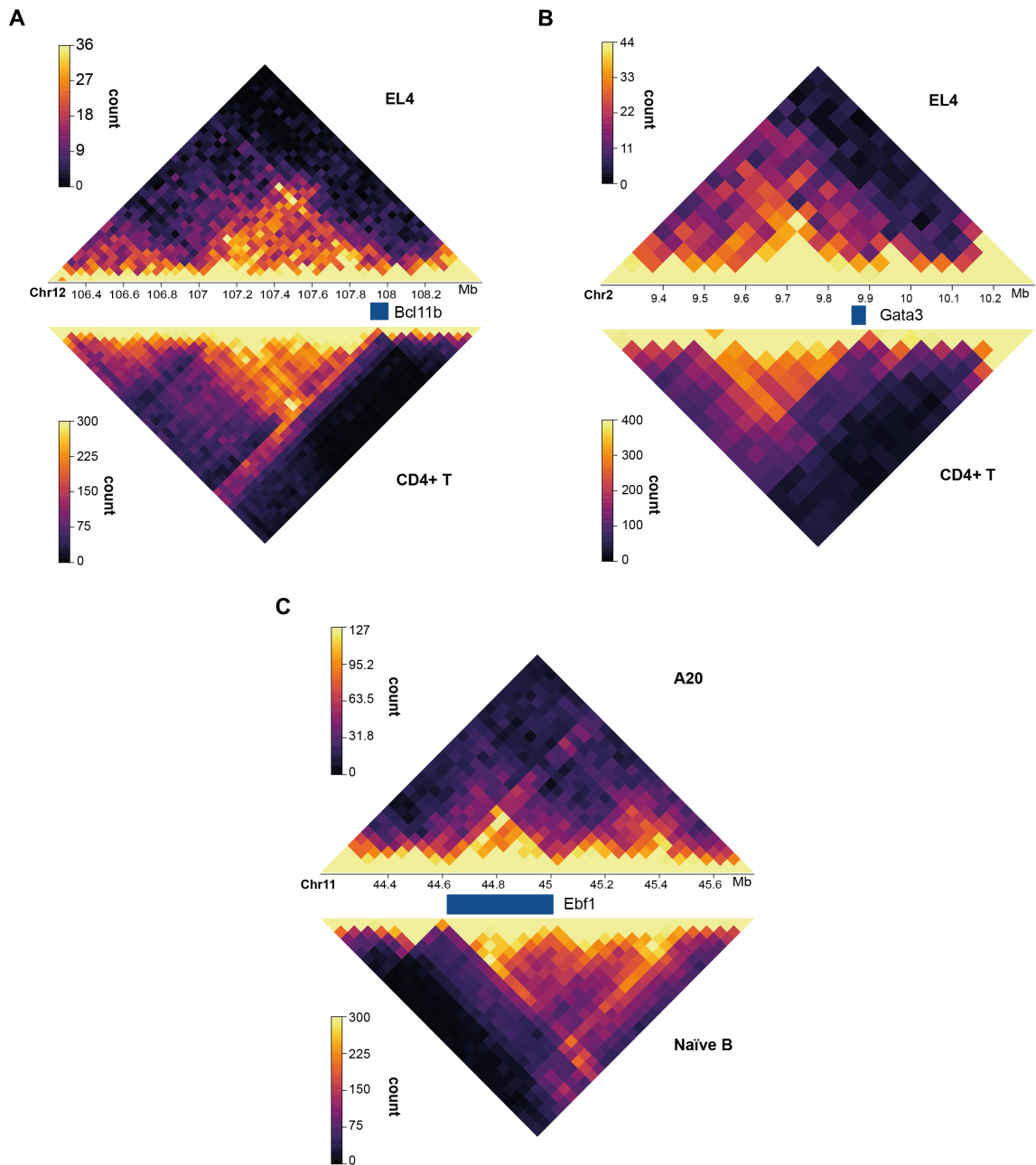


Figure 3.4. Genome structures are preserved in lymphoma cell line EL4 and A20.

A, B) Hi-C contact matrices around *Bcl11b* and *Gata3* loci between EL4 and murine splenic naïve CD4+ T cell. C) Hi-C contact matrix around *Ebf1* locus between A20 and murine splenic naïve B cell. Bar in the middle denotes the locus. Coordinates are in mm10.

CRISPR/Cas9 RNP intracellular concentration does not influence deletion efficiency

To interrogate the regulatory functions of the putative cis-elements, I employed the CRISPR/Cas9 genome editing technique in an RNP format (Liang et al., 2015). Among all the different delivery methods, RNP transfection was shown to be the most efficient one (Liang et al., 2015). As an initial optimisation and proof of principle, cis-element B1 in the *Bcl11b* TAD, which has been characterised in previous study (Li et al., 2013a), was first dissected. A pair of CRISPR RNAs (crRNAs) were designed to target each end of B1, aiming to generate a large deletion of 1.6 kb in length (Table 3.1, Fig. 3.3A). The chemically synthesized crRNAs were then assembled with the fluorescently labelled transactivating crRNA (tracrRNA) and subsequently with Cas9 endonuclease. The resultant complex, the ribonucleoprotein (RNP), was then delivered into EL4 via electroporation. By flow cytometry, it was shown that the transfection efficiency of RNP is approaching 100% (98.1%, Fig. 3.5A). Notably, the cytometry profile suggested the RNP was not uniformly distributed in the transfected EL4 cells and it is therefore important to determine whether the concentration of transfected RNP affects deletion efficiency. Therefore, cells with either a high, medium or low level of intracellular RNP were isolated and the deletion efficiency was determined by droplet digital PCR (ddPCR). It was revealed that the overall deletion efficiency is around 7-8%, with negligible difference between population of varying RNP concentration (Fig. 3.5B). Overall, this data shows that transfection of RNP is a highly efficient process and the RNP concentration does not influence the eventual deletion efficiency. As such, all the subsequent CRISPR/Cas9 experiments were performed via electroporation without any enrichment.

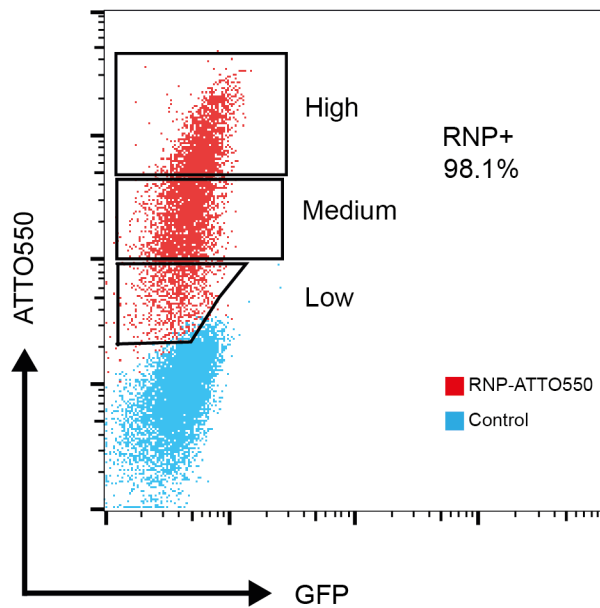
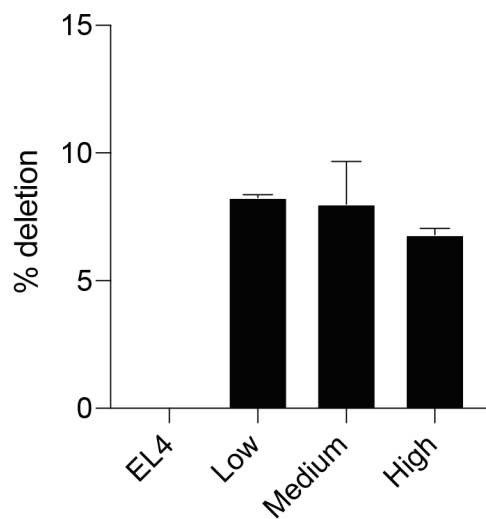
A**B**

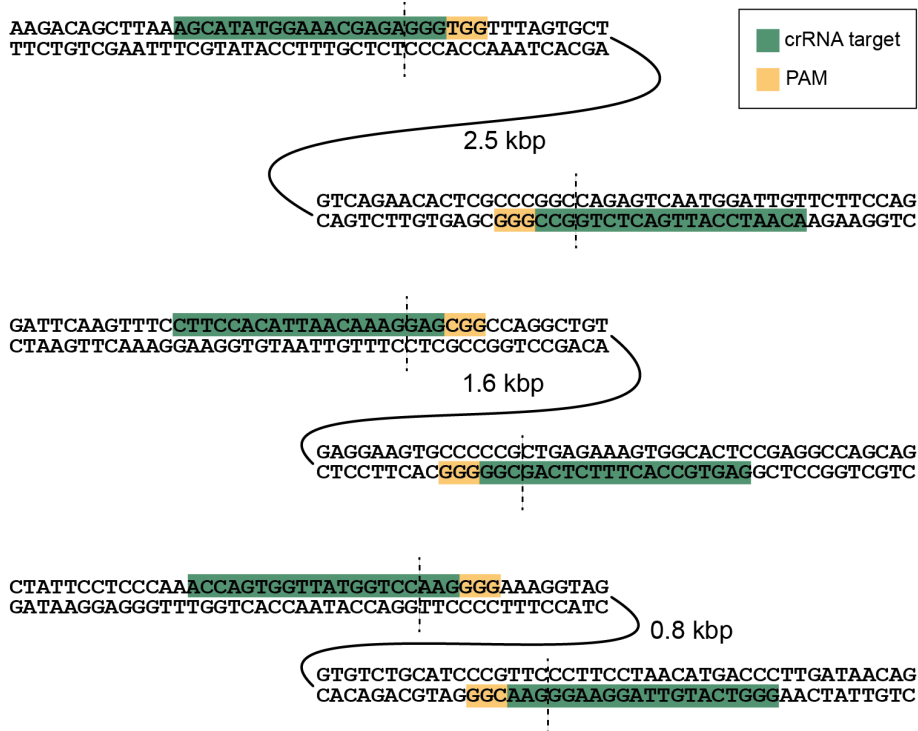
Figure 3.5. CRISPR/Cas9 RNP concentration does not affect deletion efficiency.

A) Flow cytometry data of EL4 transfected with ATTO-550 fluorescently labelled RNP, compared to the non-transfected control. Population with high, medium and low level of RNP was sorted as indicated. B) Efficiencies of the removal of cis-element B1 in EL4 with high, medium and low concentration of intracellular RNP, determined by ddPCR. Shown are mean and SEM of data from two independent experiments.

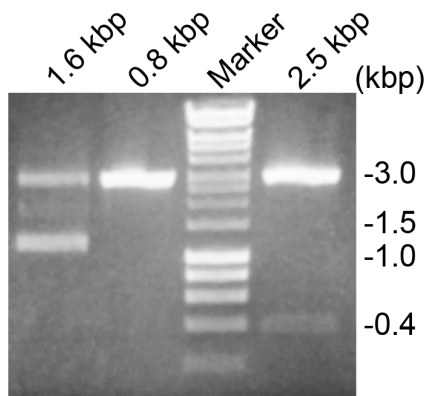
The effects of the removal length to deletion efficiency

Previously it was demonstrated that the efficiency of generating large deletion is inversely correlated with the length of target fragment (Canver et al., 2014). The deletion of larger fragment is therefore more technically challenging. Thus, I attempted to investigate the effects of the removal sizes at the Bcl11b cis-regulatory element and thereby optimise the deletion efficiency. In addition to the pair of crRNAs which aimed to generate a 1.6 kb deletion, I also designed a pair of crRNAs targeting a wider, 2.5 kb, and a narrower, 0.8 kb, genetic span (Fig. 3.6A). While both RNPs deleting 1.6 and 2.5 kb fragment were able to generate deletion visible by regular PCR (Fig. 3.6B), surprisingly the RNP targeting the smaller, 0.8 kb fragment failed to generate deletion. Through ddPCR, the deletion of both 1.6 and 2.5 kb fragments appears to be similar (Fig. 3.6C), although there is a slightly higher percentage of deletion detected when removing shorter than longer fragment (8.7% vs 7%). Overall, from this preliminary trial, it appears that in addition to the deletion fragment size, the deletion efficiency also depends largely on the design of crRNA. As such, based on the existing data, the pair of crRNAs, which generates a 1.6 kb deletion, remains as the best pair for the subsequent study.

A



B



C

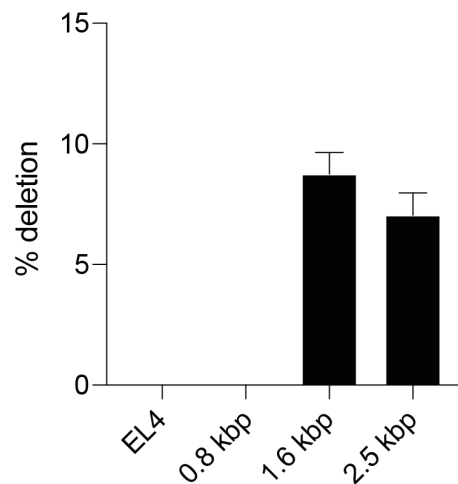


Figure 3.6. The effects of deletional length to deletion efficiency.

A) Designs of crRNAs in removal of cis-element B1, with the deletion of larger (2.5 kb, top), regular (1.6 kb, middle) and smaller (0.8 kb, bottom) fragments. Dashed lines indicated the presumed cleavage site. B) PCR was performed with primers amplifying a 2.8 kb region (WT). Successful deletion of 2.5 kb would result in a mutant allele of roughly 400 bp in length, whereas deletion of 1.6 kb leads to mutant of 1.2 kb in length. C) Deletion efficiency of different removal length, estimated by ddPCR. Shown are mean and SEM of two independent experiments.

In vitro transcribed sgRNA confers similar efficiency to chemically synthesised crRNA system

While the current CRISPR system consists of two chemically synthesised RNA components – crRNA and tracrRNA (Fig. 3.7A), recent advances has engineered the two components into one single-guide RNA (sgRNA), which can be executed in a simpler and more convenient fashion (Jinek et al., 2012). In addition, sgRNA can be easily produced in-house through in vitro transcription (IVT) using T7 RNA polymerase (Fig. 3.7B), which is economically vastly more favourable than the expensive chemically synthesised crRNA and tracrRNA. As such, I optimised the generation of in-house IVT sgRNA. A pair of sgRNAs targeting cis-element B1 was made and compared with the chemically synthesised crRNA. Through ddPCR, the B1 deletion incurred by both IVT sgRNA and crRNA was of roughly equal efficiencies, around 8.5% (Fig. 3.7C). As such, IVT sgRNA were used in all the subsequent studies.

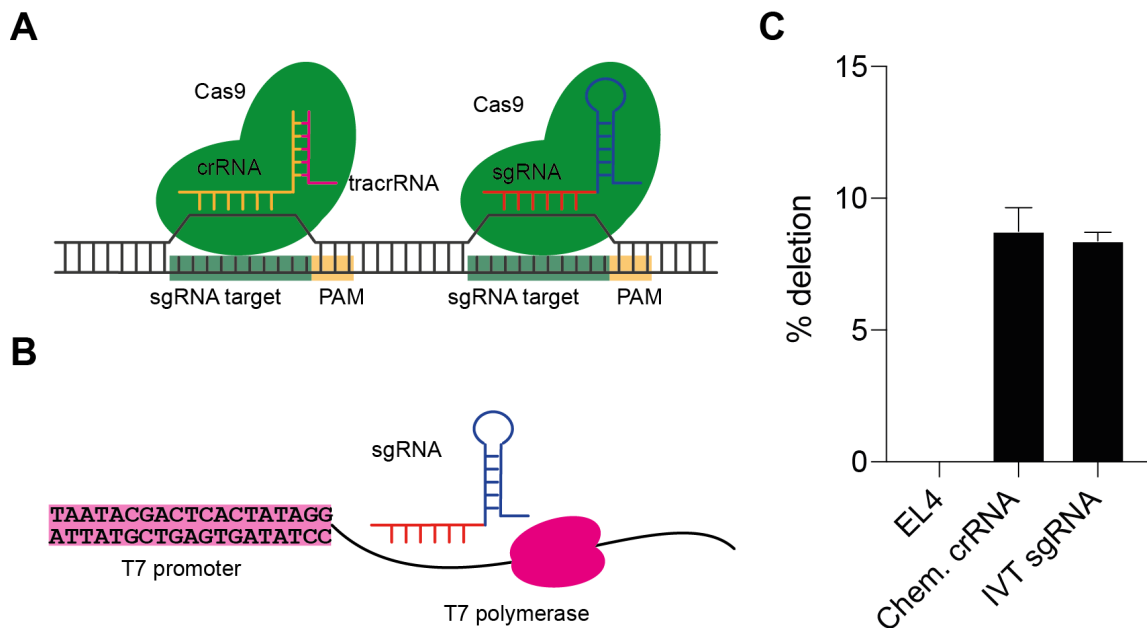


Figure 3.7. IVT sgRNA confers similar deletion efficiency to chemically synthesised crRNA system.

A) The ordinary CRISPR/Cas9 system consists of two RNA components – crRNA and tracrRNA (left), whereas a recent innovation has engineered the two components into one sgRNA (right). B) Schematic of the in vitro transcription process to generate sgRNA. PCR was first conducted to incorporate the T7 promoter upstream to the sgRNA sequence. The resultant PCR amplicon then serves as IVT template for T7 RNA polymerase to massively produce sgRNA. C) Efficiencies of deletion of B1 with using IVT sgRNA and chemically synthesised crRNA system. Shown are mean and SEM of data from two independent experiments.

Removal of cis-element B1 leads to downregulation of Bcl11b

With the above optimisation, the deletion efficiency of B1 in the bulk culture remained at a lowly 8%. Apart from the low efficiency, the diploid nature of eukaryotic cells has further complicated the interrogation – deletion would occur either in a monoallelic or biallelic manner, resulting in a great variability within population. In order to obtain a homogenous population with clear phenotypic consequence, isolation and growth of individual clones was deemed necessary. As such, the edited cell population was subjected to single cell sorting into 96-well plate via FACS. Over hundreds of such clones were isolated and grown for more than a week until confluent. Genotyping PCR was then performed to isolated mono- or bi-allelic knockout (KO) clones (Fig. 3.8A). Out of the 96 clones being genotyped, four monoallelic KO and only one biallelic KO clones were isolated, thus with a 4.17% and 1.05% efficiency, respectively. The overall allelic deletion efficiency is 3.13%, which is apparently lower than that determined by ddPCR before the sorting.

The mono- and bi-allelic KO clones were further validated by ddPCR (Fig. 3.8B), which precisely and accurately demonstrated 50% and 100% deletion for mono- and bi-allelic knockout mutants, respectively. The biallelic KO alleles were then directly sequenced without subcloning, which confirmed the removal of cis-element B1 (Fig. 3.8C). Intriguingly, the lack of sub-peak in the chromatogram suggests that both KO alleles possess the same mutation. Overall, the results illustrated the robustness of the high throughput genotyping and also neatly demonstrated the accuracy of ddPCR. As such, only regular genotyping PCR will be performed for the other subsequent deletional studies unless otherwise specified.

In order to understand whether perturbation of B1 resulted in alterations in Bcl11b gene expression. RT-qPCR was performed in both mono- and bi-allelic KO mutant, with EL4 unexposed to electroporation acting as control. This revealed a 30% down-regulation of Bcl11b in biallelic KO mutant (Fig. 3.7B), whereas the monoallelic KO clones have similar Bcl11b expression to WT EL4. The result here confirms with the previous studies (Isoda et al., 2017; Li et al., 2013a) and

thus demonstrates the effectiveness of this overall approach in identifying functional cis-regulatory elements.

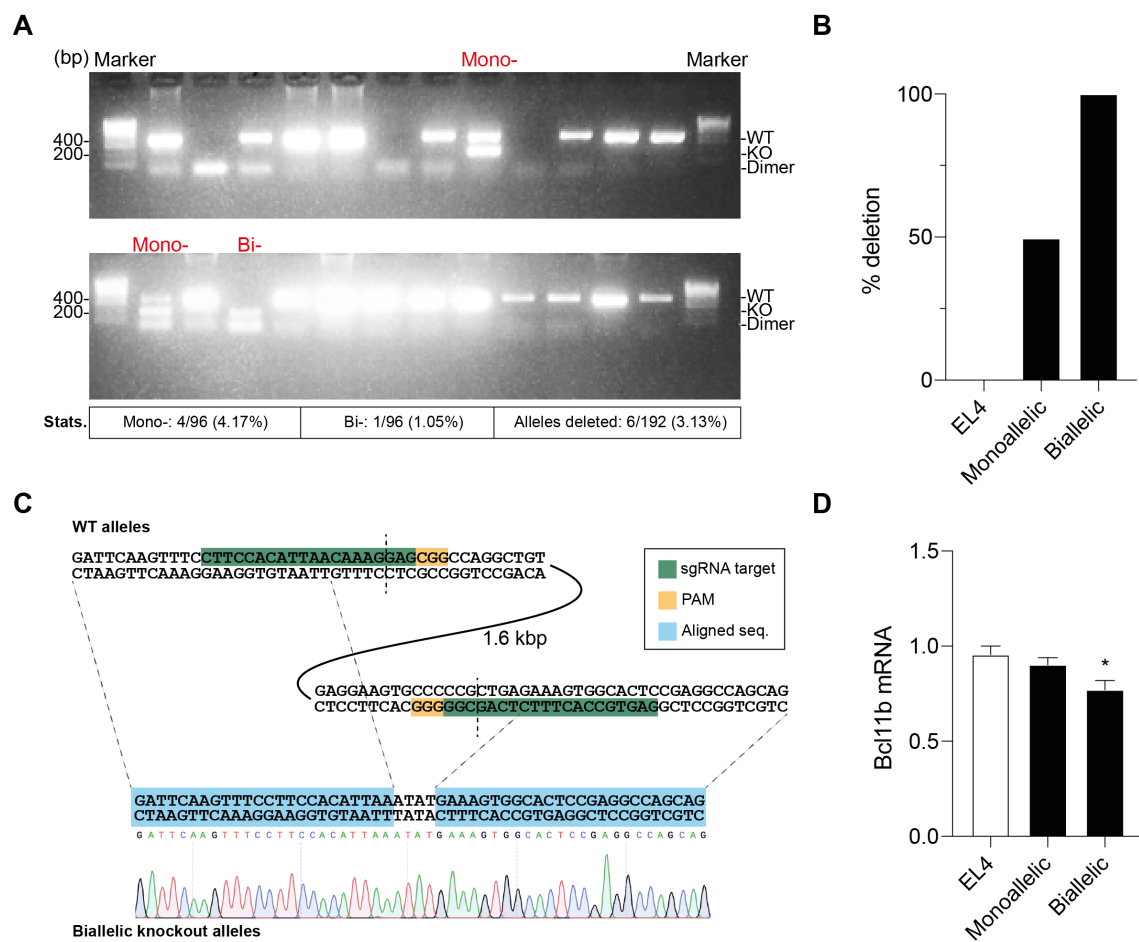


Figure 3.8. Deletion of cis-element B1 leads to downregulation of Bcl11b.

A) Representative gel image of genotyping. Using forward, reverse and internal primers, WT allele will be amplified into band of 400 bp, whereas allele with B1 knockout (KO) will lead to band of 200 bp. Primer dimer does exist in this PCR at the bottom with size smaller than 100 bp. Mono- and bi-allelic clones are indicated. Indicated below the image are the number and % occurrence of mono- and bi-allelic KO clones, as well as the allelic deletion efficiency calculated from this genotyping. B) Verification of deletion in mono- and bi-allelic KO clones by ddPCR. Shown are mean and SEM of data from two independent experiments. C) DNA sequence of biallelic KO clone aligned with WT. PCR product amplifying both mutant alleles was directly sequenced, chromatogram shown below the mutant sequence. Both KO alleles possess the same mutation. D) Transcription level of Bcl11b in mono- and bi-allelic KO mutants compared to untreated control EL4 cell line. Expression levels were normalised to endogenous control Actb. Shown are mean and SEM of data from 4 RNA samples, $*p < 0.05$ from t test comparing to control.

Loop anchors confers no regulatory functions in steady state Ebf1 expression

Similar to *Bcl11b*, in the *Ebf1* TAD there is a clear loop formed between the *Ebf1* promoter and a distantly downstream region (Fig. 3.3B). The loop anchor coincides with several open chromatin regions and thus two of them were designated as E1 and E2 for interrogation (Fig. 3.3B, Table 3.1). Since the genome structures around *Ebf1* locus is preserved in A20 (Fig. 3.4C), functional dissection was carried out using this cell line. Genotyping PCR identified ten E1 and E2 biallelic KO clones (Fig. 3.9A, B). In order to determine if perturbation of E1 or E2 interfere with *Ebf1* gene expression. RT-qPCR was performed in the biallelic KO mutants, with untreated A20 as control. Unfortunately, neither region appeared to exert a significant influence over the expression of *Ebf1* in A20 (Fig. 3.9C), albeit there appears to be a trend of increase of expression in the E1 KO, and a reduction of *Ebf1* expression in the E2 KO.

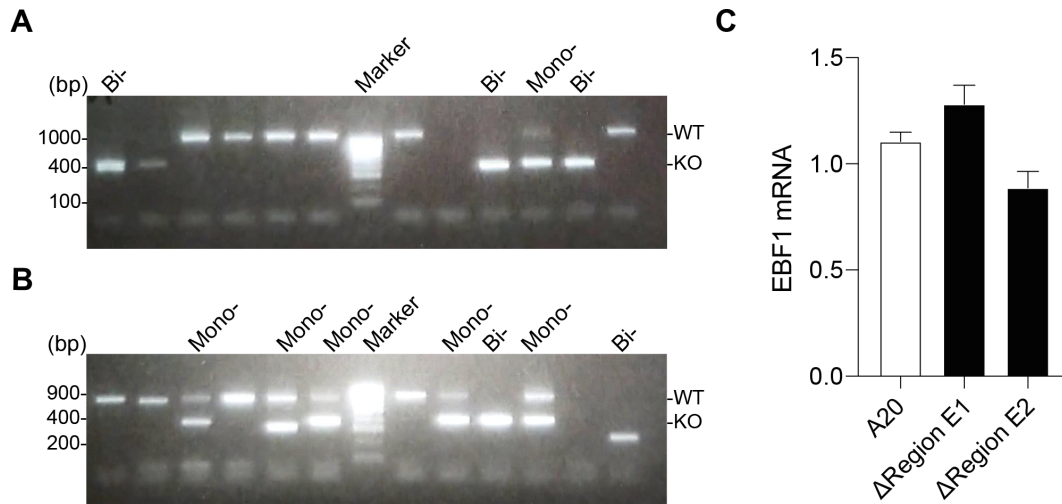


Figure 3.9. Deletion of putative cis-elements does not significantly affect Ebf1 expression in steady state.

A) Genotyping PCR of A20 clones for screening biallelic knockout of region E1. WT allele will be amplified into band of 1100 bp, whereas allele with E1 KO will lead to band of 480 bp. Mono- and bi-allelic clones are indicated. B) Genotyping PCR of A20 for isolating clones with E2 deletion. WT allele will manifest as band of 780 bp, whereas E2 KO allele will be amplified as band of 400 bp or lower. Mono- and bi-allelic clones are indicated. C) Transcription levels of Ebf1 in region E1 and E2 biallelic KO mutants compared to untreated control A20 cell line. Expression levels were normalised to endogenous control Actb. Shown are mean and SEM of data from ten different clones.

Multiple cis-regulatory elements modulate Gata3 transcription

Unlike the Bcl11b and Ebf1 loci, there is no sharp, persistent loop interacting with the Gata3 promoter (Fig. 3.3C). Instead, potential cis-regulatory elements are scattered across vast genetic distance, and three such regions were designated as G1, G2 and G3 (Fig. 3.3C, Table 3.1). Genome structures around *Gata3* locus observed in naïve T cells are also observed in the EL4 T cell line (Fig. 3.4C), allowing functional dissection of the role of these regions using this cell line.

Through genotyping PCR, clones with biallelic KO of the regions were isolated (Fig. 3.10A-C). Of note, region G2 is a very large region with 8.4 kb in length (Fig. 3.10B), and only one biallelic KO clone was obtained upon screening of 400 single-cell clones. This results in an efficiency of 0.25%, and appears to agree with the aforementioned correlation between deletion size and efficiency (Canver et al., 2014).

With RT-qPCR, I found that deletion of region G1 in EL4 cells led to a modest increase in Gata3 expression by around 50%, whereas G2 deletion caused an 80% reduction in transcription (Fig. 3.10D). In contrast, removal of element G3 did not interfere with Gata3 expression. Of note, G2 overlaps an enhancer previously identified by the Engel group (Hosoya-Ohmura et al., 2011; Ohmura et al., 2016), but not previously shown to physically interact with the Gata3 locus.

3.2. Discussion

Over the past decade, chromosome conformation capture techniques have revealed various three-dimensional chromatin structures within mammalian genome (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Nora et al., 2012). Local individual DNA interactions, including chromatin loops formed between enhancers and promoters, are preferentially clustered and harboured within TADs. While the chromatin loops exhibit a very dynamic variation between cell types, initial description regarded TADs to be apparently invariant (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Rao et al., 2014). However, several recent reports, including ours and data herein, revealed that certain TADs display lineage specificity (Bonev et al., 2017; Johanson et al., 2018b; Pekowska et al., 2018; Stadhouders et al., 2018). Cells either gain or lose a number of TAD borders during differentiation or reprogramming, resulting in merging or segregating of TADs (Bonev et al., 2017; Stadhouders et al., 2018). Similarly, during the divergence of immune lineages many loci encoding lineage-specific genes experience the emergence or disappearance of the entire TAD (Johanson et al., 2018b), as identified in this thesis the *Bcl11a*, *Ebf1*, *Mef2c*, *Bcl11b*, *Gata3* and *Ppm1h* loci. These TADs foster the formation of a local interactome which appropriately drives the expression of the underlying transcription factors, and thus serve as valuable models for disentangling the local regulome of a gene.

Several architectural features can be easily spotted by examining the HiC contact matrices. Among some of the TADs, several of them – *Bcl11a*, *Ebf1*, *Bcl11b* and *Ppm1h* – exhibit the sharp corner dots in the contact matrices. This suggests the existence of stable, persistent loops, plausibly established by CTCF and cohesin via the loop extrusion mechanism (Beagan and Phillips-Cremins, 2020; Rowley et al., 2017; Szabo et al., 2019). In contrast, the TAD encompassing the *Gata3* locus lacks such a feature. On the other hand, attributed to the large genetic span of *Ebf1* gene, an obvious loop can be observed linking between the promoter and 3' end of the gene, which is believed to be the gene loop as discussed before. Intriguingly, this gene loop appears in both naïve B and CD4⁺ T cells. Nonetheless, the gene loop of CD4⁺ T cells shows minimal interactions within, in

contrast to that of B cells which harbours intense interactions, possibly indicating a lack of transcription elongation in the former.

In the preliminary interrogation, a CRISPR based approach was established to directly delete the cis-elements of interest in the genome. Certain cis-elements were verified, whereas deletions of some others seemed to confer no functional consequences. Several intrinsic properties of cis-regulatory elements, especially enhancers, have complicated this functional dissection, namely – cis-regulatory element abundance, enhancer redundancy, spatiotemporal patterning and preestablished or remnant configuration. 1) A number of genome-wide approaches have revealed that a single promoter often engages with more than one cis-element (Chepelev et al., 2012; Ernst, 2012; He et al., 2014; Li et al., 2012a; Sanyal et al., 2012). In theory, these cis-elements could have possessed different sets of transcription factor binding sites, and thereby act additively, synergistically or antagonistically to effectuate the final pattern of gene expression. As such, deletion of one of these cis-elements may only result in a very subtle effect. 2) Further adding to this complexity is the redundancy of enhancers. This refers to the overlapping activities of two or more enhancers in driving similar patterns of expression, hence the removal of one does not hamper the overall function of the gene (Cannavo et al., 2016; Hong et al., 2008; Osterwalder et al., 2018; Perry et al., 2011). The redundant enhancers, or sometimes called shadow enhancers (Barolo, 2012; Cannavo et al., 2016; Hong et al., 2008), are believed to support a robust transcription programme that is less prone to environmental and genetic perturbations. While this redundancy appears to be a widespread feature in the mammalian genome (Osterwalder et al., 2018), it may also be a reflection of another level of complexity – 3) the spatiotemporal patterning. A gene may acquire multiple roles at different times and tissue during development (Gibson, 1996). In assist of this, the associated enhancers could solely participate in a particular tissue or developmental window to confer the precise patterning of gene expression along the course of development (Barolo, 2012; Dunipace et al., 2011; Dunipace et al., 2013; Lam et al., 2015). In addition, the cis-elements may only become important under specific environmental challenges such as pathogen invasion or stress. Accordingly, deletion of the candidate cis-element in a certain context would not

lead to disruption of transcription, as it plays a more crucial role in other conditions.

4) In fact, some of these patterning can be attributed to a preestablished three-dimensional chromatin structure. Other than immediately directing transcription upon *de novo* enhancer-promoter contact, it is suggested that enhancers can additionally activate gene expression in a lagged, permissive manner (de Laat and Duboule, 2013). Herein enhancers can engage with promoters and stay poised for subsequent activation, in which transcription is delayed until triggered. As such, deletion of these poised structures would hence confer no apparent effects. On the other hand, albeit still largely unexplored, existing structures could also be a remnant of the past developmental trajectory. Theoretically, during the lineage specification some of the structures may form in preparation for one of the developmental pathways, which could then remain in place even when the cells have diverged down a different trajectory. As illustrated in the *Ebf1* TAD (Fig. 3.2A), where a gene loop was retained in the naïve CD4⁺ T cells, even though they do not express any *Ebf1* protein. Similarly, although being a lab-induced process, incomplete rewiring with the persistence of remnant structures has been documented during reprogramming (Beagan et al., 2016). Thus, it appears that some chromatin structures are present as relics of the past, and deletional interrogation of these would therefore be in vain.

Complexity also lies at the level of TADs. As mentioned before, loop extrusion and phase separation help consolidate the final patterning of TAD segregation, apparently leading to two different nature of TADs – CTCF-dependent loop domains and CTCF-independent compartmental domains (Beagan and Phillips-Cremins, 2020; Nuebler et al., 2018; Rao et al., 2017; Schwarzer et al., 2017). While in some cases deletions of the TAD boundaries or CTCF loop anchors are accompanied by pathological consequences (Flavahan et al., 2016; Lupianez et al., 2015), it appears that the majority of gene expression are unaltered upon the abolishment of loop domains (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017). As such, even though persistent loops, manifested as sharp corner dots in the Hi-C matrices, are present in the immune loci studied herein, the disruption of those loops might not lead to any aberrant effects. Intriguingly, the disappearance of loop domains has unmasked a finer underlying compartmental structure (Rao et al., 2017; Schwarzer et al., 2017). It thus remains in question

whether these compartmental domains which took over the loop domains contribute to the robustness and maintain the gene expression. In addition, it also remains to be elucidated whether the loop domains are only critical in a temporal manner, for example during the establishment of the underlying gene regulatory circuit.

With the extensive screening upon single cell isolation, it continues to be technically challenging for this CRISPR-based approach to be applied in *ex vivo* primary cells. Unlike interrupting a protein-coding reading frame, deletion of cis-elements requires a long fragment of DNA to be removed. Achieving this large deletion is of much lower efficiency compared to disrupting protein expression (Canver et al., 2014). One of the reasons is probably due to the asynchronous generations of DSBs and the early re-joining of either by NHEJ. Furthermore, the efficiency is also inversely correlated with the length of target fragment (Canver et al., 2014). Among all the different delivery methods, RNP transfection was shown to be the most efficient (Liang et al., 2015). Although a plasmid-based approach might superficially sound more effective due to the persistent expression of Cas9, conceivably the early mutations of the PAM or sgRNA target sites by NHEJ after initial targeting will thus render the later presence of Cas9 entirely useless. As such, the efficiencies accomplished in this study are unlikely to be improved by alternative delivery approaches, and therefore appear to be largely constrained by the premature re-joining and the length of the target fragments. With this sub-optimal efficiency, which adds to the aforementioned subtleness and redundancy of the cis-elements, deletional interrogation in primary cells is likely to be an even greater challenge.

Therefore, functional dissection of potential cis-elements, be it in cell lines or primary cells, appears to be a trade-off. The interaction windows in Hi-C matrices are of large size – 20 or 50 kb in length, overlaying with ATAC-seq, or other epigenetic profiles such as H3K4me3, H3K27ac or CTCF binding, allows us to narrow down the putative cis-elements. However, this leads to an increasing number of such candidates, and, apart from the more laborious interrogation that follows, these shorter cis-elements are also more likely to suffer from the subtleness and redundancy mentioned above. On the other hand, removing

several such cis-elements together, especially if they reside far apart, remains to be a technically challenging process, which perhaps requires screening of thousands of clones. In partially dealing with these confounding factors, several CRISPR screening approaches have been developed (Shukla and Huangfu, 2018). One such study involved the design of over tens of thousands of sgRNAs pairing up to generate deletions of about 2 kb over a large genetic distance (Diao et al., 2017). The resulting massive number of mutants were then interrogated with extensive deep sequencing to identify the cis-regulatory elements. On the other hand, other screening approaches adopted the catalytically inactive dCas9 fused with different transcription effectors to activate (Joung et al., 2017; Simeonov et al., 2017) or silence (Fulco et al., 2016; Liu et al., 2017) , and therefore identify, cis-regulatory elements.

4. Identification and characterisation of the lncRNA Gm13218 at enhancer of Gata3

In the previous chapter a downstream region of Gata3 was identified as an enhancer that regulates its steady state expression (Fig. 3.9). As an important transcription factor, Gata3 is expressed throughout T lymphocyte development and has been shown to be essential in various T cell stage transitions, as well as during the final maturation of both CD4⁺ and CD8⁺ T cells (Garcia-Ojeda et al., 2013; Hosoya et al., 2009; Pai et al., 2003; Tai et al., 2013; Wang et al., 2008). Gata3 is also critical for T_H2 differentiation where it acts as the master regulator of the process (Pai et al., 2004; Zhu et al., 2004). During this lineage specification Gata3 establishes and maintains a T_H2 specific chromatin configuration around the IL-4, IL-5 and IL-13 loci, thereby facilitates the synthesis and secretion of these T_H2 cytokines (Spilianakis and Flavell, 2004). In contrast, B cell development does not require Gata3 (Ting et al., 1996). The appropriate level of Gata3 expression appears to be critical during T cell development with any deviation from normal expression levels shown to be detrimental (Anderson et al., 2002; Ano et al., 2013; Chen and Zhang, 2001; KleinJan et al., 2014; Nawijn et al., 2001a; Nawijn et al., 2001b; Taghon et al., 2001; Taghon et al., 2007). Apart from thymic developmental defects, deregulation of Gata3 can lead to lymphoma (Manso et al., 2016; van Hamburg et al., 2008; Zhang et al., 2016) and also increase the susceptibility of allergic asthma (Ano et al., 2013; KleinJan et al., 2014).

Intriguingly, Gata3 is also expressed in other non-haematopoietic tissues of neural, urogenital and cardiac origins, and it appears that different distinct cis-regulatory elements are utilised (Lakshmanan et al., 1999). For example, an inner ear specific enhancer is located at 571 kb downstream of the Gata3 transcription start site (TSS) (Moriguchi et al., 2018), whereas a kidney specific Gata3 enhancer rests at 113 kb upstream of TSS (Hasegawa et al., 2007). Contrarily, using transgenesis as well as CRISPR deletion mice, expression of Gata3 in T cells was shown to be regulated by an enhancer located 280 kb 3' of TSS (Hosoya-Ohmura et al., 2011; Ohmura et al., 2016). While this enhancer coincides with the region G2 that was identified in the previous chapter, upon a

closer inspection of this region I discovered that the existence of a transcription unit of a poorly annotated non-coding RNA (known as Gm13218). Given the importance of non-coding RNA in modulating chromatin topology and gene expression as mentioned in earlier sections, the current study embarked to characterise this Gm13218 and attempt to elucidate if and how it regulates the expression of Gata3.

4.1. Results

A long non-coding RNA Gm13218 is located in an enhancer of Gata3

Inspection of the G2 region that interacts with the Gata3 promoter revealed the existence of a poorly annotated expression unit denoted as Gm13218. To further examine its expression, end point RT-PCR was subsequently performed on whole thymi and also EL4 cells (Fig. 4.1A). Gm13218 was indeed highly expressed in both the thymus and EL4, whereas the EL4 cells with region G2 deleted showed no expression. These results suggest that Gm13218 is expressed coincident with Gata3. Next, to determine the full-length sequence and transcription start site (TSS) of Gm13218, I conducted 5' and 3' RACE (Fig. 4.1B). Resultant DNA bands were harvested and sequence analysed. Since the 5' RACE was performed via ligation of an RNA adaptor to capped RNA species after a cap removal process, this suggests Gm13218 bears 5' cap. On the other hand, 3' RACE conducted via an oligo dT primer suggests that Gm13218 is polyadenylated. The sequencing revealed that Gm13218 consists of three exons, with alternative splicing caused by a variable intron boundary at the first intron-exon junction, resulting in variable lengths of the ultimate transcripts, ranging from 630 to 690 nt. As the size is over 200 nt, Gm13218 fits into the category of lncRNAs. In addition, the promoter or TSS of Gm13218 coincides with the most prominent T cell specific accessible region in region G2 (Fig. 4.1C).

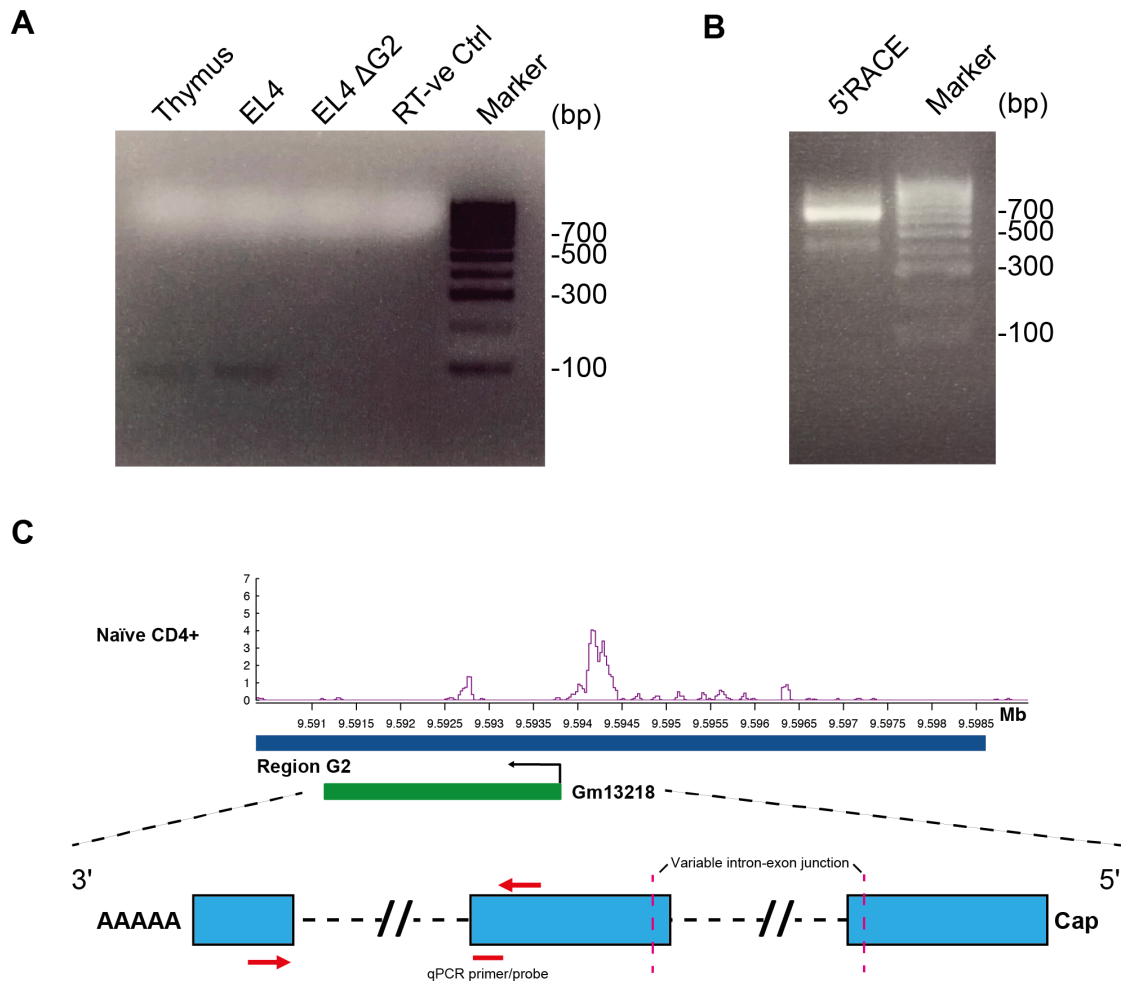


Figure 4.1. Expression of long non-coding RNA Gm13218 at Gata3 enhancer.

A) End point RT-PCR performed in cDNAs from thymus, EL4, mutant EL4 with region G2 deleted and no RT control. Expected size of amplicon is roughly 100 bp. B) Gel image of Gm13218 5' sequences from 5' RACE. C) Schematics of Gm13218 locus and RNA transcript structure. Chromatin accessibility profile of splenic naïve CD4⁺ T cells around region G2 shown and aligned with the GM13218 TSS. Sky blue boxes indicate exons, whereas red arrows and bar represent primers and probe used for RT-qPCR.

Gm13218 locus DNA is differentially methylated and differentially accessible

Further analysis of the chromatin accessibility landscape (Yoshida et al., 2019) in Gm13218 locus reveals that the promoter region, designated here as G2-1, becomes accessible as early as DN1 stage (Fig. 4.2), and it remains open in naïve CD4⁺ and CD8⁺ T cells. Conversely, this promoter is not accessible in B cells. Although less prominent, there is another differentially accessible region, G2-2, with similar opening pattern located further upstream of Gm13218 TSS. Other than chromatin accessibility, DNA methylation, as discussed previously, can modulate gene expression with methylated DNA at promoter predominantly correlating with reduced transcription (Deaton and Bird, 2011). To look into the DNA methylation pattern, the publicly available whole genome bisulphite sequencing data from naïve CD4⁺ T and resting B cells were retrieved and aligned along the Gm13218 locus (Kazachenka et al., 2018). The TSS of Gm13218 and a large portion of gene body are differentially methylated, with very little CpG dinucleotide methylated in naïve CD4⁺ T cells, contrasting with the widespread DNA methylation of this region in B cells (Fig. 4.2). Surprisingly, the accessible region G2-1, or the apparent promoter of Gm13218, appeared to have low levels of methylation in both B and CD4⁺ T cells. However, the upstream accessible G2-2 displayed a higher degree of DNA methylation in B cells, compared to the complete absence of methylated CpG in CD4⁺ T cells.

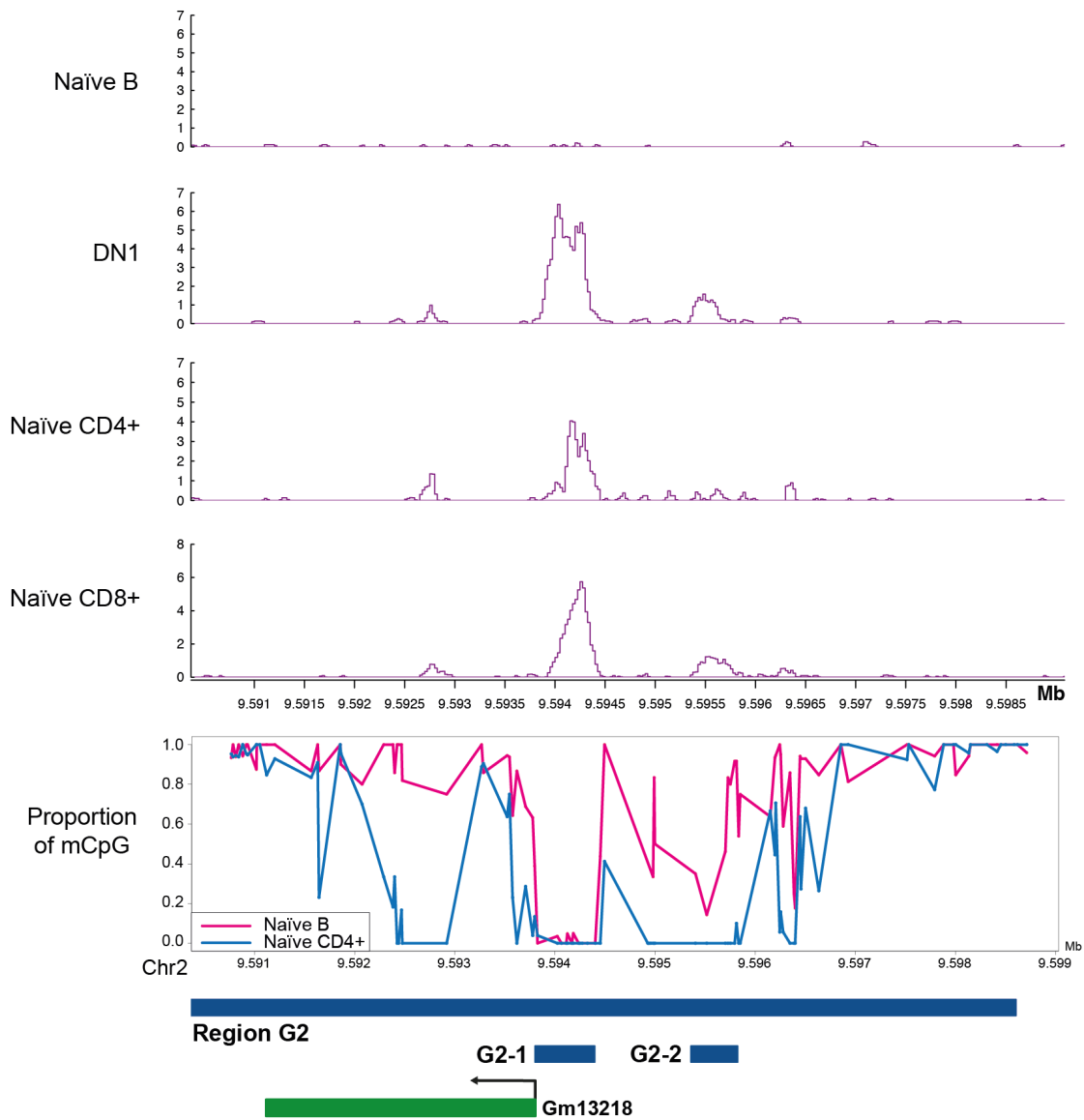


Figure 4.2. Chromatin accessibility and DNA methylation profiles at the Gm13218 gene.

ATAC-seq chromatin accessibility landscapes of DN1 thymocytes, splenic naïve B cells, CD4+ and CD8+ T cells at the Gm13218 locus. Aligned below are the DNA methylation profiles of naïve B and CD4+ T cells, plotted as population proportion of methylated cytosine in each CpG dinucleotide motif. The long blue bar indicates region G2 or Gata3 enhancers identified in previous chapters, whereas short blue bars indicates T cell specific accessible regions G2-1 and G2-2. Green bar denotes Gm13218 locus with arrow illustrating transcription directionality and TSS determined from 5' RACE.

Gm13218 is co-regulated with Gata3 in thymus

Next, to investigate the expression pattern of Gm13218 in relation with that of Gata3, I performed RT-qPCR across different tissues and thymocytes subsets (Fig. 4.3B, C). From the different tissues analysed, Gata3 transcripts are abundant in the thymus, and, albeit at a lower level, present in lymph node, spleen, kidney, and barely present in bone marrow. Similarly, Gm13218 is mostly expressed in thymus with limited expression in the lymph node, but virtually undetectable in spleen, bone marrow and kidney. Among thymocytes of different T cell developmental transitions (Fig. 4.3A), both transcripts of Gata3 and Gm13218 peak at the DN2 and DN3 stages (Fig. 4.3B, C). Once transited to DN4, expression of both decrease and remain at low level in the later DP, CD4+ and CD8+ stages. Significantly, the expression patterns of Gata3 and Gm13218 in different thymocyte subsets were found to be positively correlated (Fig. 4.3D).

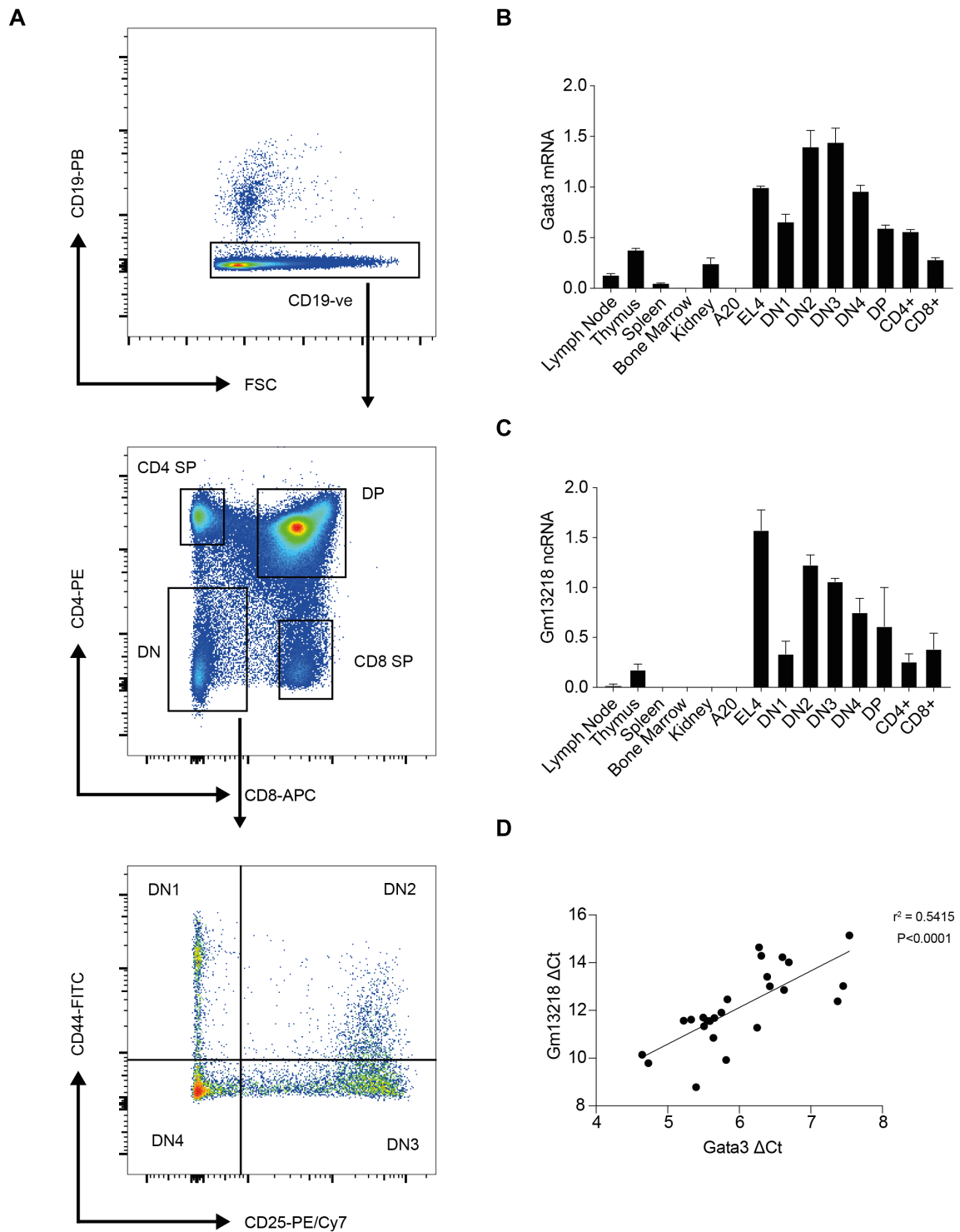


Figure 4.3. Expression pattern of Gata3 and Gm13218 across different tissues.

A) Gating strategies for sorting different thymocyte subsets. B, C) Gata3 and Gm13218 expression levels across tissues and various thymic T cell subsets with Actb as endogenous control. Show are mean and SEM from three different RNA samples. D) Correlation of expression between Gata3 and Gm13218 across different thymic T cell subsets, plotted as ΔCt of each gene to Actb.

Gm13218 is co-regulated with Gata3 during T_H2 differentiation

Gata3 is well known for its lineage-specifying role during T_H2 differentiation, deregulation of which has been implicated in the development of allergic asthma (Ano et al., 2013; KleinJan et al., 2014; Pai et al., 2004). I therefore examined the expression levels of Gata3 and Gm13218 during CD4⁺ T cell differentiation. I stimulated naïve CD4⁺ T cells under non-biased, T_H1 (IL12 + anti-IL4) or T_H2 (IL4 + anti-IL12 + anti-IFN γ) conditions for seven days (Fig. 4.4A). While expression of both transcripts remained low in unbiasedly activated CD4⁺ T cells and throughout the 7 days of T_H1 polarisation (Fig. 4.4B, C), Gata3 and Gm13218 transcripts were significantly upregulated in the first 24 hours under the T_H2 condition to 3-fold and 24-fold respectively and remained high throughout the 7-day culture. Interestingly, Gata3 transcription increased and stayed at the peak level (5-fold) until the experimental end point, whereas Gm13218 transcripts peaked at day 2-4, reaching around 70-fold, and declined to 28-fold on day 7. Nonetheless, the expression patterns across the polarisation period are highly positively correlated (Fig. 4.4D). This data suggests that Gata3 and Gm13218 are both responsive to IL-4 signalling. Overall, by sharing the same TAD, Gata3 and Gm13218 are co-regulated throughout thymic development and respond in concert to IL-4 during T_H2 polarisation.

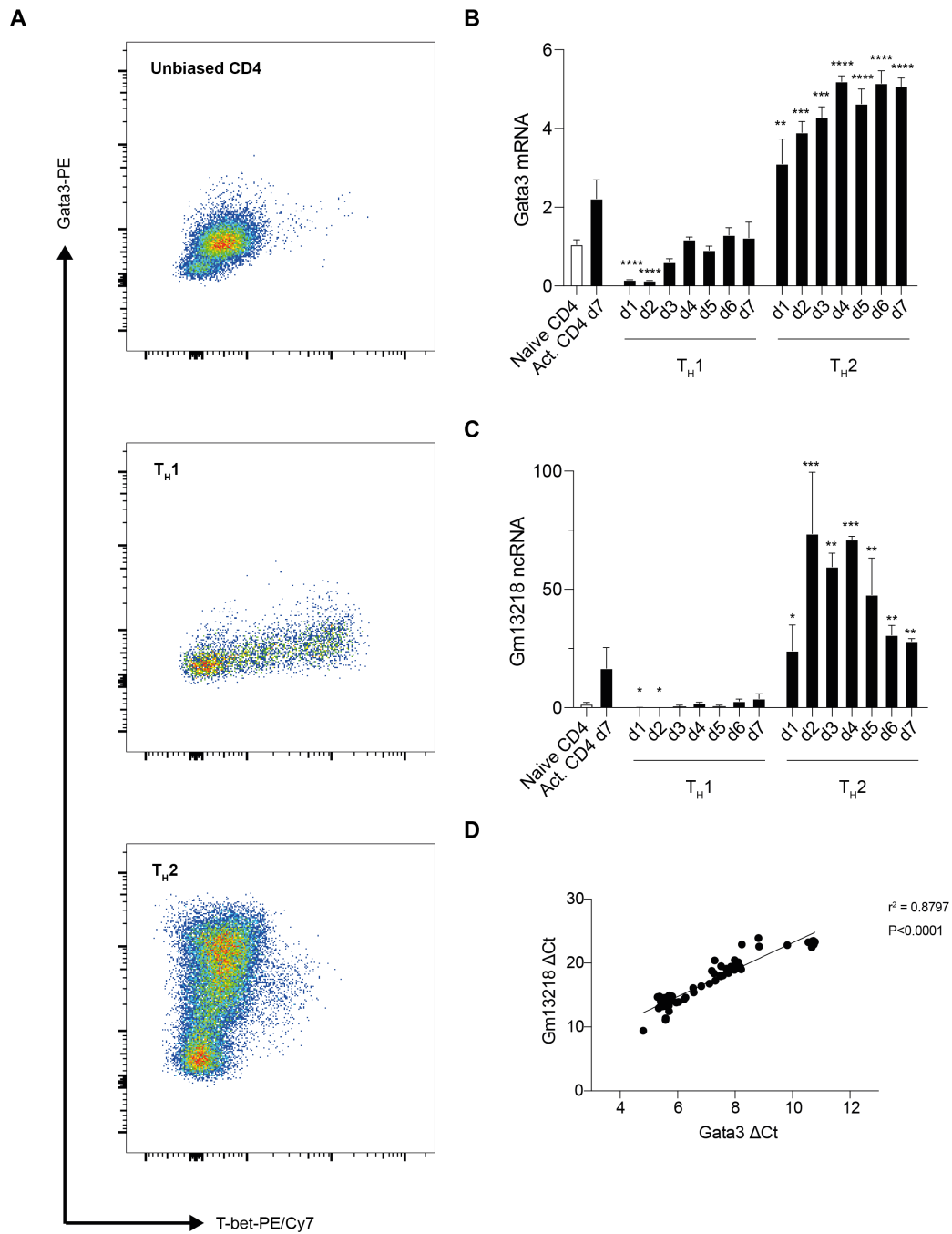


Figure 4.4. Expression of Gata3 and Gm13218 during TH1 and TH2 differentiation.

A) Intracellular staining of Gata3 and T-bet in activated unbiased CD4⁺, T_H1 and T_H2 cells after 7 days of culture. B, C) Gata3 and Gm13218 transcription level across 7 days of unbiased, T_H1 and T_H2 specification with Actb as endogenous control. Show are mean with SEM from 3 independent experiments, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ from ANOVA with Dunnett test comparing to naïve CD4 after logarithmic transformation of data. D) Correlation of Gata3 and Gm13218 expression during CD4⁺ T cell differentiation, plotted as ΔCt to Actb.

Gm13218 is co-regulated with Gata3 during early T cell lineage commitment

As shown in Figure 4.3, Gm13218 is already expressed as early as during DN1 stage in thymus. In order to understand if Gm13218 was expressed at an even earlier developmental stage, for example during the lineage commitment from haematopoietic progenitor cells (HPC), I explored the FANTOM5 CAGE (Cap analysis of gene expression) database. I analysed one line of samples generated from a six-day stimulation of EBF1-knockout HSC to induce T cell differentiation (Arner et al., 2015; Noguchi et al., 2017), through co-culturing with TSt-4 and DLL1 stromal cells (Ikawa et al., 2010). Intriguingly, as early as one day after stimulation towards the T cell lineage, Gata3 and Gm13218 expression levels were upregulated and both continued to gradually increase with time over the six-day course of stimulation (Fig. 4.5A, B). Again, the expression patterns across the induction period are positively correlated (Fig. 4.5C).

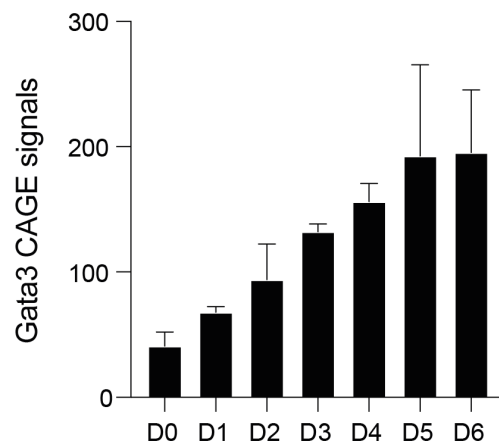
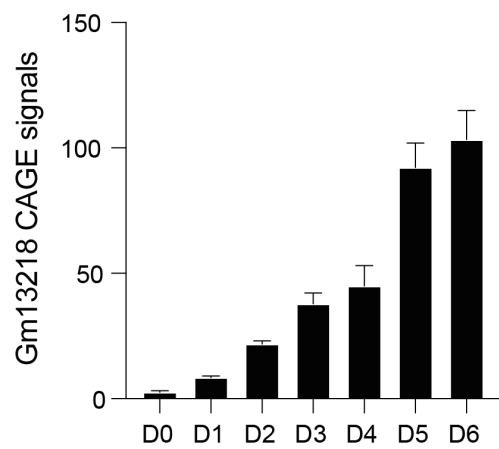
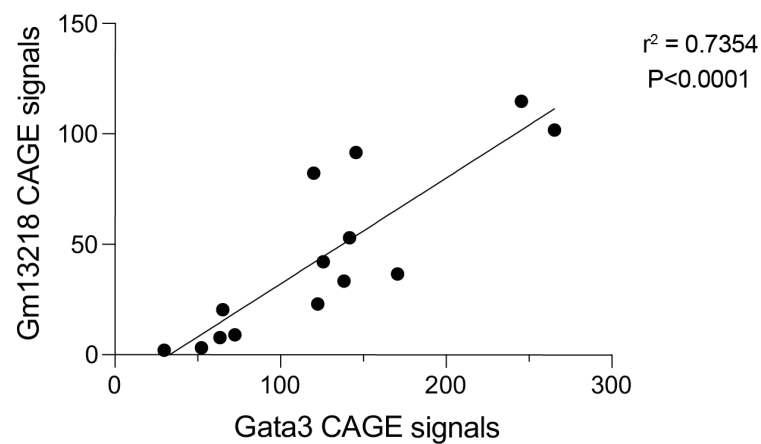
A**B****C**

Figure 4.5. Gm13218 is co-regulated with Gata3 during early T cell lineage commitment.

A, B) Gata3 and Gm13218 expression levels, measured as CAGE signals, retrieved from FANTOM5 across 6 days of T cell differentiation from EBF1-knockout HPC. Shown are mean and SEM from three samples. C) Correlation of CAGE signals between Gata3 and Gm13218 across the stimulation period.

T cell specific TAD and non-coding transcription are conserved between mouse and human

Upon inspection, I noticed that the sequence of accessible regions within region G2 is conserved with a non-coding DNA region near GATA3 in human chromosome 10 (Fig. 4.6A). Apart from nucleotide sequence of the cis-elements, the synteny and genetic distance between are also conserved as well – the cis-element is located roughly 280 kb downstream of GATA3 TSS. Strikingly, Hi-C previously performed in the lab in human naïve CD4⁺ T and B cells (Johanson et al., 2018a) unveiled a TAD comprising both GATA3 and the conserved cis-element with intense interactions between (Fig. 4.6B). Such interactions and TAD structure are completely absent in naïve B cell (Fig. 4.6B). I then exploited the FANTOM database (Hon et al., 2017) to investigate the occurrence of non-coding transcription along the GATA3 TAD. Intriguingly, bidirectional transcription appears to initiate from the conserved cis-element, generating two lncRNAs: CAT00000105356.1 and CAT00000117261.1 (Fig. 4.6C). More importantly, these two lncRNAs show lineage specificity in T cell (Fig. 4.6D, E), expression of which are enriched in almost all T cell samples but absent in other cell types. Since the lncRNAs are transcribed outwards from the conserved region, it is unlikely that the RNA transcripts *per se* show any sequence conservation with murine Gm13218 and indeed this is the case (Fig. 4.6A). Thus, overall, the T cell specific TAD, the synteny and interaction between GATA3 and expression of downstream lncRNA initiating cis-elements, are well conserved between mouse and human, implicating that the two lncRNAs may be important for the regulation of human GATA3.

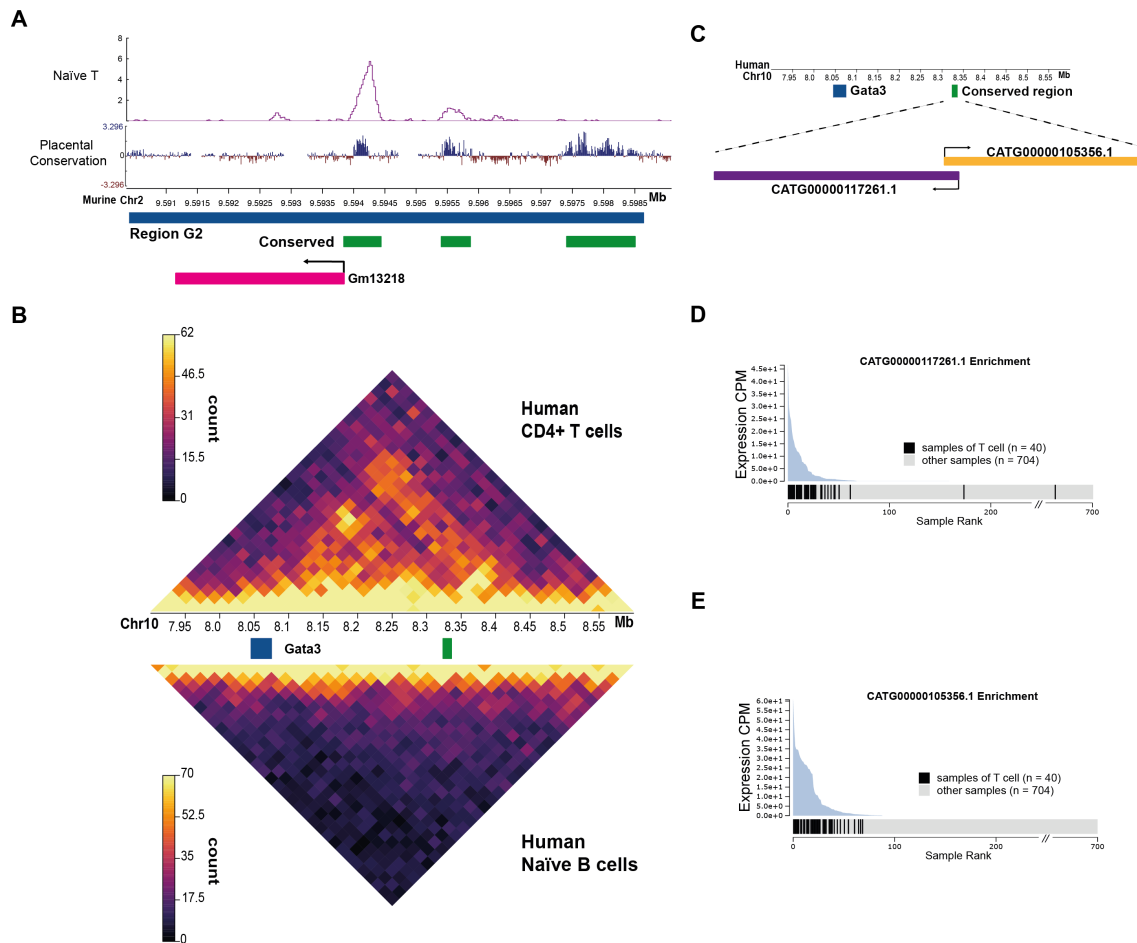


Figure 4.6. Conservation of GATA3 TAD, lncRNA expression and synteny in human.

A) Placental species sequence conservation along Gata3 enhancer, retrieved (Hubisz et al., 2011) and aligned with accessibility profile of murine T cells. Green bars designate conserved regions. Pink bar illustrates the Gm13218 locus with arrow indicates TSS and transcription directionality. B) Hi-C contact matrices around GATA3 locus between human naïve B and CD4+ T cells. Blue bar indicates GATA3 locus, whereas green bar represents the sequence conserved cis-elements. C) lncRNA species transcribed from the conserved cis-elements, identified by FANTOM database. Length of bars for lncRNAs not in scale. D, E) Expression enrichment profiles of lncRNAs CAT00000105356.1 and CAT00000117261.1 across different cell types. Data retrieved from FANTOM. Black vertical lines indicate the T cell samples, with expression ranked in a descending order from left to right.

Antisense oligonucleotide knockdown of Gm13218 does not affect steady state Gata3

Next, I embarked on a series of experiments to assess the function of the lncRNA Gm13218. First, I attempted to knockdown the endogenous Gm13218 by deploying antisense oligonucleotides (ASO, Fig. 4.7A), where introduced ASOs would base-pair with targets, leading to subsequent cleavage by RNase H (Gleave and Monia, 2005). The RNase H cleavage elicited by ASO occurs in both cytoplasm and nucleus (Liang et al., 2017), compared to silencing via small-interfering RNA (siRNA) which is mostly restricted in cytoplasm (Zeng and Cullen, 2002). Thus, ASO appears as a more favourable strategy since the spatial distribution of Gm13218 is still unknown. To achieve significant knockdown, four ASOs targeting all exons of Gm13218 were pooled together and delivered twice, with the second dose delivered 48 hours following the first administration, to the EL4 cell line via electroporation (Fig. 4.7B). In this way, Gm13218 was downregulated to 20% of the original in EL4 (Fig. 4.7C). However, Gata3 transcription was surprisingly unaffected.

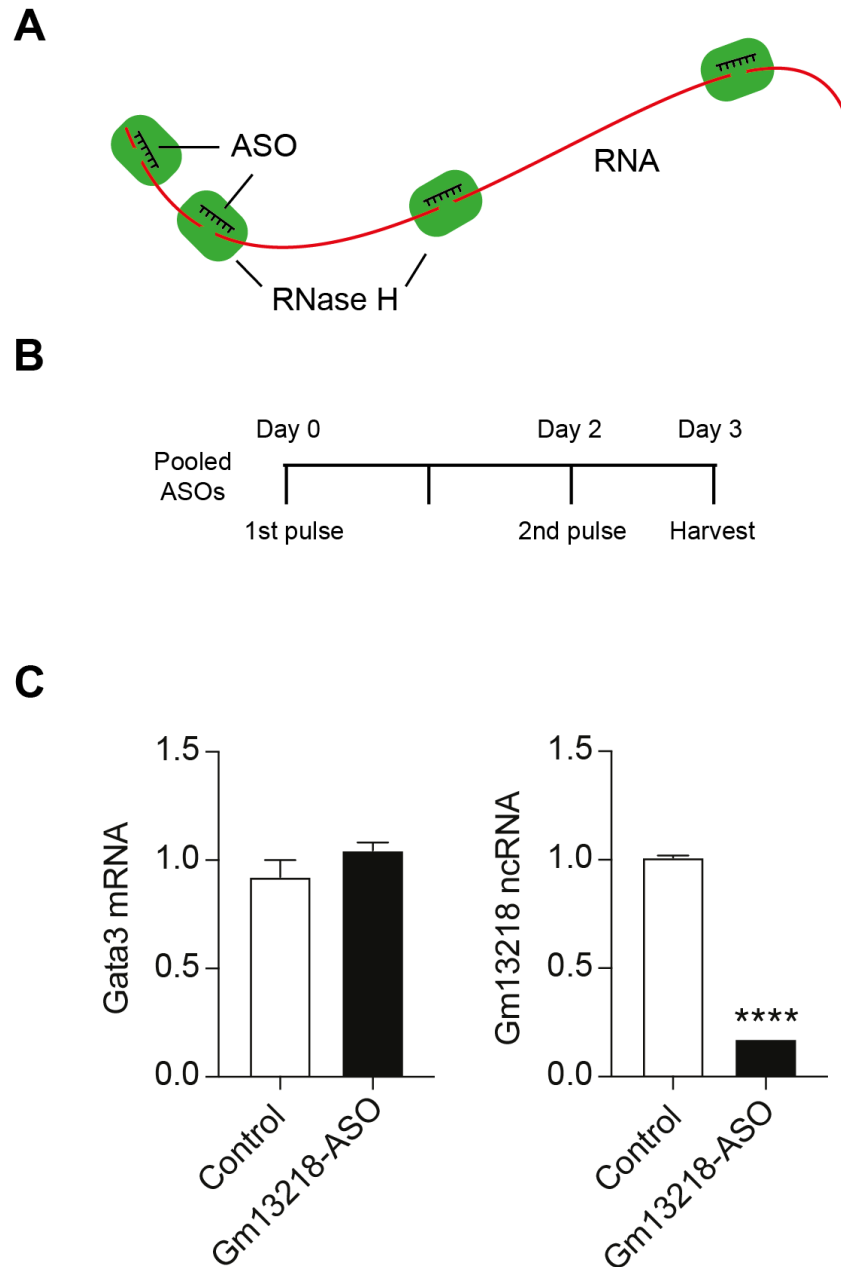


Figure 4.7. Gm13218 knockdown via ASO confers no influence to Gata3 expression.

A) Mechanism of RNA degradation by ASOs, in which base pairing between ASOs and RNA would elicit subsequent cleavage by RNase H. B) Schematic of ASO delivery. Four ASOs targeting all exons of Gm13218 were pooled together and delivered on day 0 and day 2, with RNA harvest on day 3. C) Gata3 and Gm13218 expression levels upon ASOs administration with Actb as endogenous control. Shown are mean and SEM from two independent experiments with **** $p < 0.0001$ from t -test comparing to control.

CRISPRi via dCas9-KRAB silence Gm13218 and prolonged targeting downregulate Gata3

Although ASOs degrade most of the lncRNA, there was still a significant 20% of transcripts left in the cells. In an attempt to silence Gm13218 via a different and perhaps more potent approach, CRISPR interference (CRISPRi) was employed using dCas9 fusing with a KRAB domain (Thakore et al., 2015) to target Gm13218 TSS (Fig. 4.8A, B). Transfectants of EL4 harvested 7 days after transduction showed the transcription of Gm13218 was greatly suppressed to only 3% of the original level (Fig. 4.8C). However, Gata3 expression was, similar to what was observed during ASO-mediated knockdown, unaltered. While transgenes incorporated might still be unstable, or the modulatory effects of the lncRNA might require a longer time to establish, transduced EL4 was thus allowed to grow for another week. Stable transfectants of EL4 harvested 14 days after showed equally potent, 97% reduction of Gm13218. Interestingly, Gata3 transcription was shown to be downregulated to 68% of the original level (Fig 4.8D).

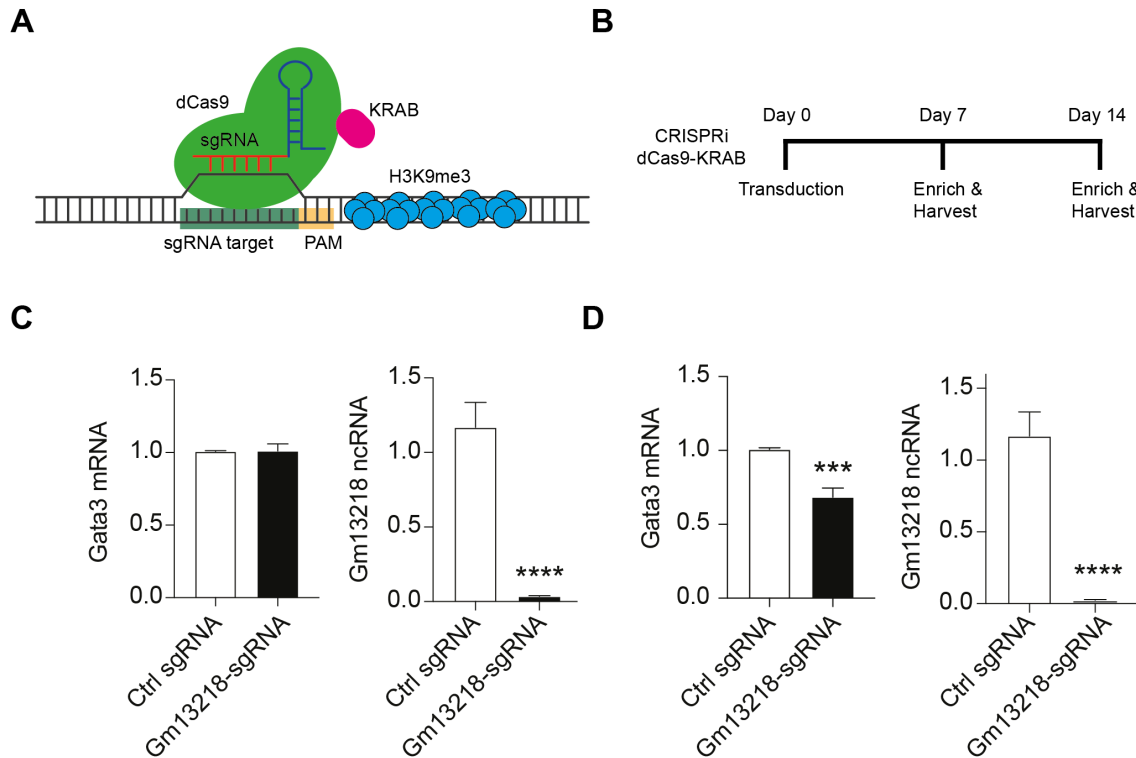


Figure 4.8. CRISPR interference via KRAB domain silences Gm13218 and prolonged targeting leads to downregulation of Gata3.

A) Schematic of mechanism of CRISPRi via dCas9-KRAB, in which KRAB would lead to laying down of H3K9me3. B) Experimental outline for CRISPRi, stable transfectants of EL4 with dCas9-KRAB targeting to Gm13218 TSS were generated by lentiviral transduction, with cells enriched and harvested on day 7 and day 14. C) Gata3 and Gm13218 expression levels on day 7. D) Gata3 and Gm13218 expression levels on day 14. Expression levels were normalised to Actb. Shown are mean and SEM from three independent experiments with *** $p < 0.001$, **** $p < 0.0001$ from t -test comparing to control.

Introduction of polyadenylation signal near Gm13218 TSS reduces Gm13218 but not Gata3

LncRNAs can exert their function via the non-coding transcript *per se*, for example when used as a scaffold to recruit histone modifiers (chapter 1.1, Fig. 1.4), or through the act of transcription itself, as to alter the local chromatin landscape or interfere with loop formation (chapter 1.1, Fig. 1.4). While silencing induced by KRAB domain indeed is able to halt transcription of Gm13218, the histone modification H3K9me3 that was induced will also mask the transcription binding sites. Additionally, the KRAB-induced histone modification has been documented to spread to nearby regions, both genetically and three-dimensionally (Thakore et al., 2015). It therefore remains unclear whether the effect observed in CRISPRi is due to masking of transcription binding sites, spreading of heterochromatin, halt of transcription elongation or absence of RNA transcripts.

To understand the importance of the act of transcription of non-coding elements, a number of studies have exploited the transcription termination signal, whereupon the insertion of which would prevent downstream transcription elongation (Engreitz et al., 2016; Ho et al., 2006; Isoda et al., 2017; Kornienko et al., 2013; Latos et al., 2012; Ling et al., 2004). I therefore tried to introduce a SV40 polyadenylation signal, by CRISPR-mediated homology directed repair (HDR), into EL4 downstream near Gm13218 TSS. To achieve this, I adopted a strategy of incorporating into sgRNA a streptavidin-binding RNA aptamer S1m (Carlson-Stevermer et al., 2017), and using a biotinylated single-strand oligodeoxynucleotide (ssODN) repair template (Fig. 4.9A). The sgRNA bearing the S1m aptamer can be bound by streptavidin added exogenously, which can then recruit the biotinylated ssODN HDR template, thereby increasing the spatial concentration of ssODN in the vicinity of cleavage site. As such, this strategy is believed, and has been shown, to increase the rate of HDR. I designed this S1m bearing sgRNA to target 144 bp downstream of Gm13218 TSS, whereas the ssODN template is of 182 nt long, with a SV40 polyadenylation signal, 82 nt in length, flanked by 50 nt homology arm on each side (Fig. 4.9B). After in-house synthesis of the S1m bearing sgRNA through IVT, I assembled it with Cas9,

streptavidin and biotinylated ssODN template. The resultant RNP-ssODN complex was then introduced to EL4 via nucleofection. Subsequent genotypic screening by PCR has isolated, out of around 800 clones screened, only one bi-allelic knock-in clone, #B9 (Fig. 4.9C). Unexpectedly, this clone possesses unequal length of alleles, and both exceed the predicted size if precise HDR was undergone.

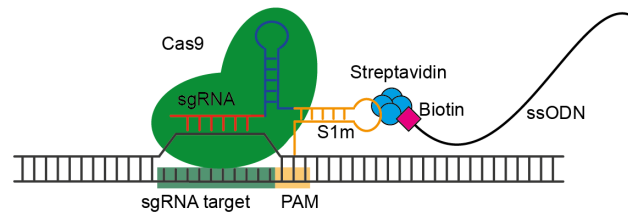
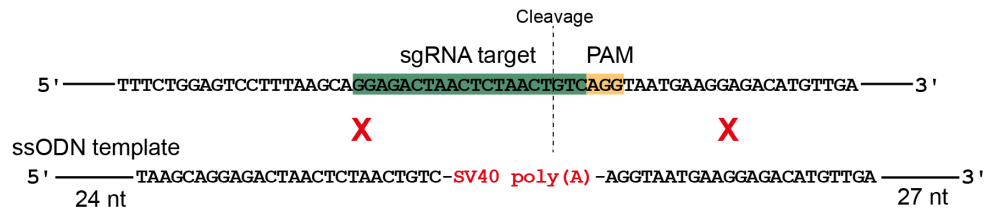
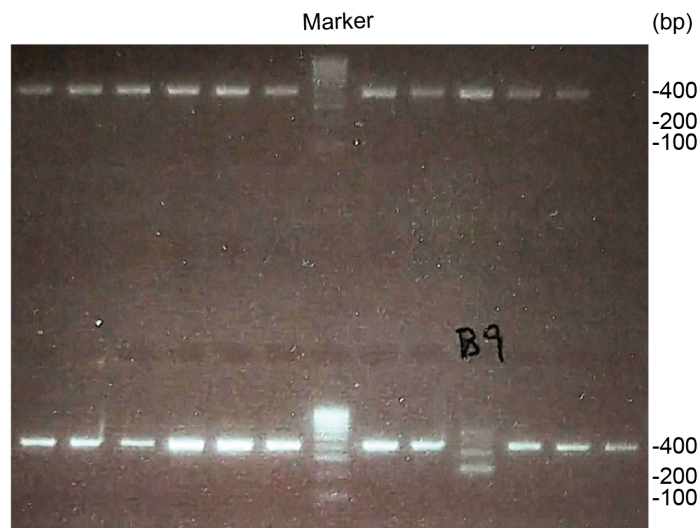
A**B****C**

Figure 4.9. Introduction of SV40 poly(A) signal in EL4 near Gm13218 TSS.

A) Strategy in inserting DNA sequence via HDR after CRISPR cleavage. The sgRNA bears an RNA aptamer S1m, which can be bound by streptavidin and in turn recruit a biotinylated single-strand oligodeoxynucleotide (ssODN) repair template, can guide Cas9 to target region and increase rate of HDR due to the increased local concentration of repair template. B) Schematic of SV40 poly(A) signal insertion to 144 bp downstream of Gm13218 TSS. The ssODN template used contains the poly(A) signal flanked by 50 nt homology arm on each side. C) PCR screening to isolate biallelic poly(A) signal knock-in clone. WT band is 400 bp, poly(A) incorporated allele would have bands of expected size of 180 bp and 490 bp. Clone B9 appears to have biallelic knock-in despite the unexpectedly large size of both alleles.

Due to the unexpected sizes, Sanger sequencing was performed on both alleles of the clones and aligned to a theoretical, expected sequence when precise HDR occurred (Fig. 4.10A). The SV40 polyadenylation signal was indeed successfully incorporated into both alleles. However, surprisingly, both alleles suffer extensive unintended duplications of homology arm of repair template and small random insertions around the poly(A) signal. While highly unexpected, Gata3 and Gm13218 expression in this clone #B9 were still analysed (Fig. 4.10B, C). Transcription of Gm13218 was successfully reduced to 8% of the original, however, expression of Gata3 was largely unaltered, which is in line with the previous experiments suggesting that transcription Gm13218 is not required for the maintenance of Gata3 expression.

A

```

EL4POLYA TCGGTAACCTGGCCAGGTGCTTTCTGGAGTCCTTTAAGCAGGAGACTAACTCTA-----
ALLELE-1 TCGGTAACCTGGCCAGGTGCTTTCTGGAGTCCTTTAAGCAGGAGACTAACTCTA-----
ALLELE-2 TCGGTAACCTGGCCAGGTGCTTTCTGGAGTCCTTTAAGCAGGAGACTAACTCTAGGTGCT

EL4POLYA -----ACTGTCGCAGTGAAAAAATGCTT
ALLELE-1 -----ACTACTGTCGCANTGAAAAAATGCTT
ALLELE-2 TTCTGGAGTCCTTTAAGCAGGAGACTAACTCTA---ACTGTCGCAGTGAAAAAATGCTT
SV40 Polyadenylation signal

EL4POLYA TATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTAAGCTGCAATAAACA
ALLELE-1 TATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTAAGCTGCAATAAACA
ALLELE-2 TATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTAAGCTGCAATAAACA

EL4POLYA AGTTAGGTAATGAAGGAGACATGTTGAATTGGCGAGGGTTGAATCAAG-----
ALLELE-1 AGTTAGGTAATGAAGGAGACATGTTGAATTGGCGAGGGTTGAATCAAGGGTAAGCTTC
ALLELE-2 AGTTAGGTAATGAAGGAGACATGTTGAATTGGCGAGGGTTGAATCAAAAAAATGCTTA

EL4POLYA -----
ALLELE-1 TTCTTGCAAAGCCTGATGACCTGACTTTGATGACCCAGCCCCACATTAGAGGAGAGAA
ALLELE-2 TTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTAAGCTGCAATAAACAAG

EL4POLYA -----
ALLELE-1 CTGACTCCCCAAGTCTGTCTTTGTCTCCTTCTCCACAGGCCCTGTTTCTAGCTATTCA
ALLELE-2 TTAGGTAATGAAGGAGACATGTTGAATTGGCGAGGGTTGAATCAAGAACACTGTCAGGT

EL4POLYA -----
ALLELE-1 AGTTGCTAAGTTCCTTGACTCTGTCTCAACTGTAAATACTCAATTAAATGTTTCA
ALLELE-2 AATGAAGGAGACATGTTGAATTGGCGAGGGTTGAATCAAG-----

EL4POLYA -----
ALLELE-1 GGGAGAAATGAAGTTTCCGGGTTCCCAAGGCCGAGGATCATTTGGCATCAGTTTCAGTGGAG
ALLELE-2 -----

EL4POLYA -----
ALLELE-1 GATGTGTCATATGGCTACCTGGAGTCANANTCTCTTANGTACTTCAGACTGTCTAGCATC
ALLELE-2 -----

EL4POLYA -----
ALLELE-1 TTGAGATGCAGCCAAGATTGGCTGCAAAATGTCANGTAATGAAGGAGACATGTTGAATTG
ALLELE-2 -----

EL4POLYA -----AACACAGAAAACCATCTATCCATCAGGGACCTT
ALLELE-1 GCGAGGGTTTGAATCAAGAACACAGAAAACCATCTATCCATCAGGGACCTT
ALLELE-2 -----AACACAGAAAACCATCTATCCATCAGGGACCTT

```

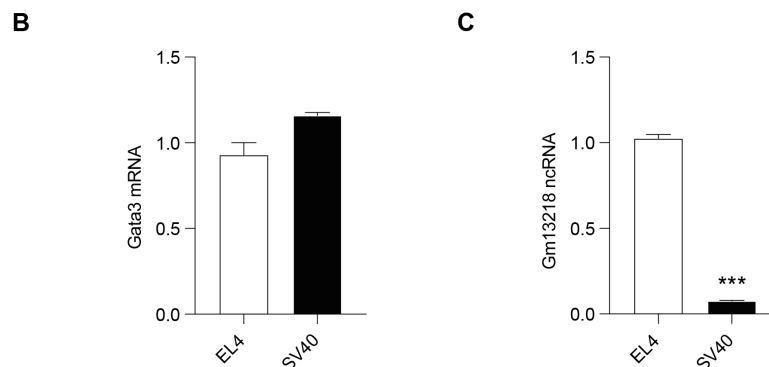


Figure 4.10. Introduction of poly(A) signal reduces Gm13218 transcription but not Gata3.

A) SV40 poly(A) signal insertion in EL4. Targeted region in Clone #B9 was sanger-sequenced. EL4POLYA represents the expected sequence after knock-in, whereas ALLELE-1 and -2 are actual allelic sequences from clone #B9. Aligned sequences, including SV40 poly(A) signal, are in black. B, C) Gata3 and Gm13218 expression levels in clone #B9 with Actb as endogenous control. Shown are mean and SEM from three independent experiments with *** $p < 0.001$ from t -test comparing to control.

The promoter of Gm13218 does not affect steady state Gata3 transcription in EL4

Another way to interrogate the function of Gm13218 including the role of transcription involves the direct removal of the lncRNA promoter. As have mentioned before, there are main two accessible regions along the Gata3 enhancer, G2-1 and G2-2 (Fig. 4.11A), where G2-1 coincides with Gm13218 TSS and is thus promoter of Gm13218. Therefore, I further dissected the regulatory functions of the two cis-elements via CRISPR/Cas9. Unexpectedly, removal of G2-1 did not disrupt Gata3 transcription, albeit expression of Gm13218 was suppressed down to 10% (Fig. 4.11B, C). Intriguingly, deletion of G2-2 led to downregulation of both, with 25% and 15% reduction of Gata3 and Gm13218, respectively. This suggests that G2-2 appears to regulate both Gata3 and Gm13218 expression.

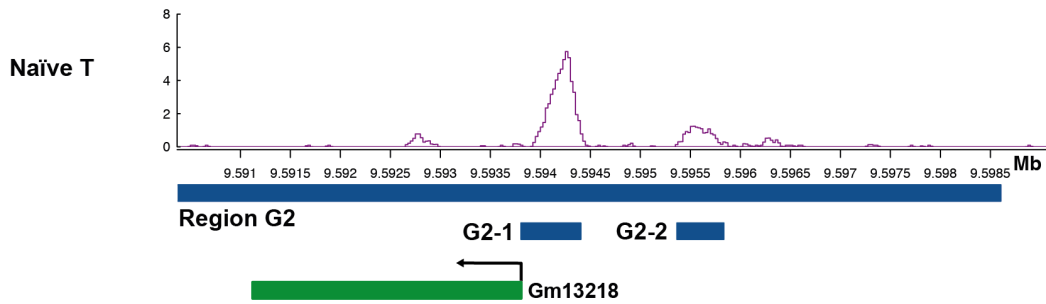
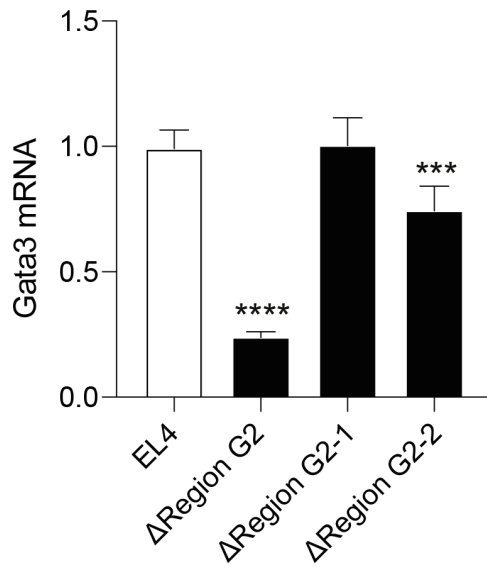
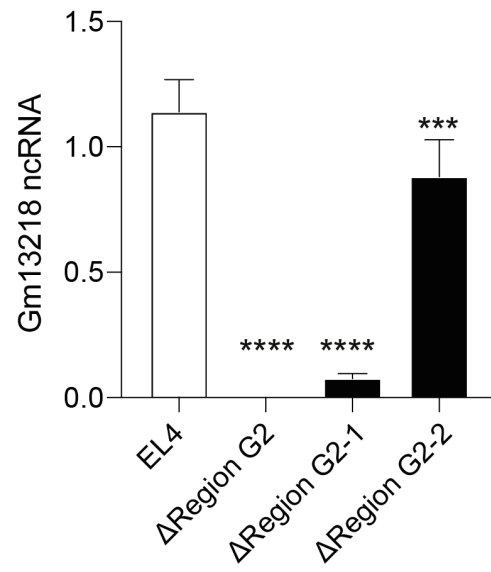
A**B****C**

Figure 4.11. Promoter of Gm13218 does not regulate steady state Gata3 in EL4.

A) ATAC-seq chromatin accessibility profile of Gata3 enhancer, designated as G2, for splenic naïve T cells. The two prominent open regions are designated as G2-1 and G2-2, represented by the shorter blue bars. Green bar illustrates the Gm13218 locus with arrow indicates TSS and transcription directionality. B, C) Gata3 and Gm13218 expression levels in various deletional mutants, transcript levels were normalised to Actb endogenous control. Shown are mean and SEM from 6-10 RNA samples, *** $p < 0.001$ and **** $p < 0.0001$ from ANOVA with Dunnett test comparing to EL4 control.

Exogenous source of Gm13218 modestly upregulates Gata3 in A20 B cells

Other than removing endogenous lncRNA or stopping transcription, another way for functional dissection is to provide an exogenous source. Although so far Gm13218 transcripts appear to have minimal influence to the steady state Gata3 level in EL4, it could potentially be functional during the establishment of the 3D chromatin environment, especially given the highly spatio-temporal patterning of lncRNA. I therefore attempted to ectopically express Gm13218 in a B cell line A20 that expresses negligible levels of Gata3.

While in previous CRISPRi, expression of dCas9-KRAB was performed using lentiviral vectors, lncRNA overexpression appears to be difficult using the same strategy. Usually, with lentiviral vector multiple proteins can be over-expressed from a single mRNA precursor by exploiting the RNA motif – internal ribosome entry site (IRES) (Morgan et al., 1992) – or the self-cleaving peptide sequences such as P2A and T2A (Hu et al., 2009; Ibrahimi et al., 2009). However, while a common approach, the constraint of only one mRNA per vector, imposed by the nature of retroviral reverse transcription, poses a problem to overexpressing non-coding RNA if a selection marker is needed. Indeed, there are strategies overexpressing short hairpin RNA (shRNA) using lentiviral vectors (ter Brake et al., 2008), also as in the case of overexpressing sgRNA in the previous dCas9-KRAB experiments. However, these shRNAs and sgRNAs are driven by H1 or U6 promoters bound by RNA polymerase III, and therefore usually lack 5'-cap and polyadenylation tail, and thus considerably differs from the aforementioned Pol II-driven nature of lncRNA Gm13218.

To circumvent this, I adopted the *PiggyBac* transposon system (Yusa et al., 2011), in which a hyperactive *PiggyBac* transposase excises DNA fragment flanked by specific inverted terminal repeats (ITR), and translocate to motif 5'-TTAA-3' anywhere in the genome (Fig. 4.12A). From the sequences derived from RACE, I cloned the full length Gm13218, driven by CMV promoter and terminated by a bGH polyadenylation signal, and a TagBFP selection marker under the control of PGK promoter, into a *PiggyBac* transposon vector via Gibson assembly (Gibson et al., 2009). Stable transfectants of A20 revealed a successful ectopic

overexpression of Gm13218 (Fig. 4.12C), and interestingly this led to a modest 1.7-fold upregulation of Gata3 (Fig. 4.12B).

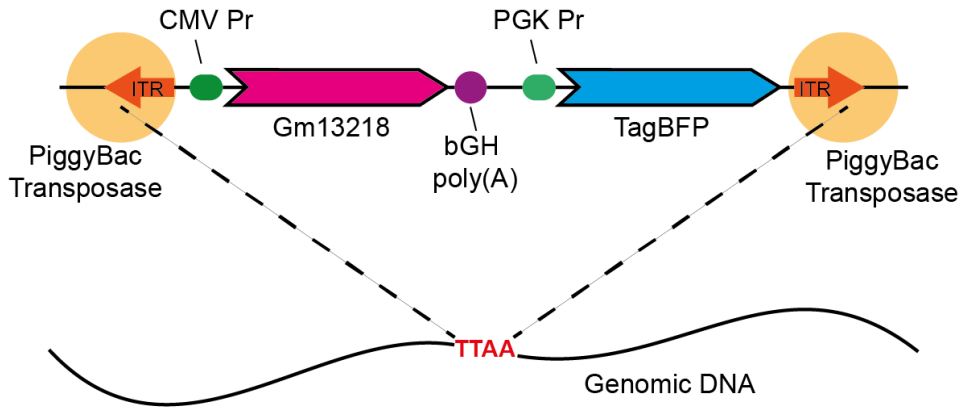
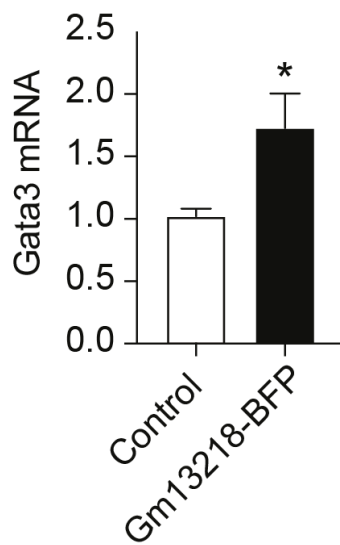
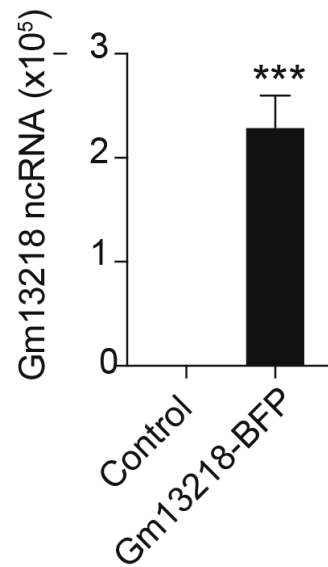
A**B****C**

Figure 4.12. Ectopically overexpression of Gm13218 modestly upregulate Gata3 in A20.

A) Mechanism of *PiggyBac* transposition. Transgenes including *Gm13218* and selection maker *TagBFP*, arranged as two separate transcription units, are flanked by ITRs, which are recognised by *PiggyBac* transposase. Transgenes would then be excised and translocated to 5'-TTAA-3' motif in the genome. B, C) *Gata3* and *Gm13218* levels upon ectopic and stable overexpression of *Gm13218*. Levels normalised to *Actb* endogenous control. Shown are mean and SEM of data from six samples. * $p < 0.05$ and *** $p < 0.001$ from *t*-test comparing to control overexpressing only *TagBFP*.

CRISPRa via SunTag-VP64 and SAM, and DNA demethylation failed to activate Gm13218

With the recent expansion of CRISPR toolbox (Section 1.3), CRISPR activation (CRISPRa) has also been developed to interrogate cis-regulatory elements (Simeonov et al., 2017), where it was hypothesized that gene transcription could be activated by recruiting a strong transcriptional activator to the corresponding enhancer, even when enhancer is in an inactive state in the assayed cells. It is therefore interesting to test this hypothesis by adopting CRISPRa in A20 in the attempt to activate Gm13218 and also Gata3.

Among various CRISPRa strategies, in this trial I deployed the two most effective approaches – the SunTag system (Tanenbaum et al., 2014) and the synergistic activation mediator (SAM, Fig. 4.13A, B) (Chavez et al., 2016; Konermann et al., 2015). With ten repeats of GCN4 epitope fused to dCas9, SunTag is able to recruit ten VP64 domain to confer gene activation, whereas in SAM, in addition to a single VP64 attached to dCas9, the MS2 aptamer utilised is capable of recruiting four pieces of p65 and Hsf1. Both approaches were carried out via a lentiviral transduction manner. However, the stable transfectants of both systems failed to significantly upregulate Gm13218 (Fig. 4.13C, D) – there were only very slight, if not negligible, activation. This could be possibly due to the DNA methylation at the Gm13218 TSS that was observed in B cells (Fig. 4.2). As such, I attempted to demethylate the CpG motif along the TSS by employing SunTag system to recruit multiple DNA demethylating agent, the TET1 catalytic domains (Fig. 4.13E) (Morita et al., 2016). This strategy has been demonstrated to mediate gene activation. Unfortunately, by targeting TET1 to Gm13218 TSS, the approach could not activate Gm13218 (Fig. 4.13F). Plausibly, this might be due to the absence of lineage-specific transcription factors to bind to the enhancer, even though the region is now demethylated.

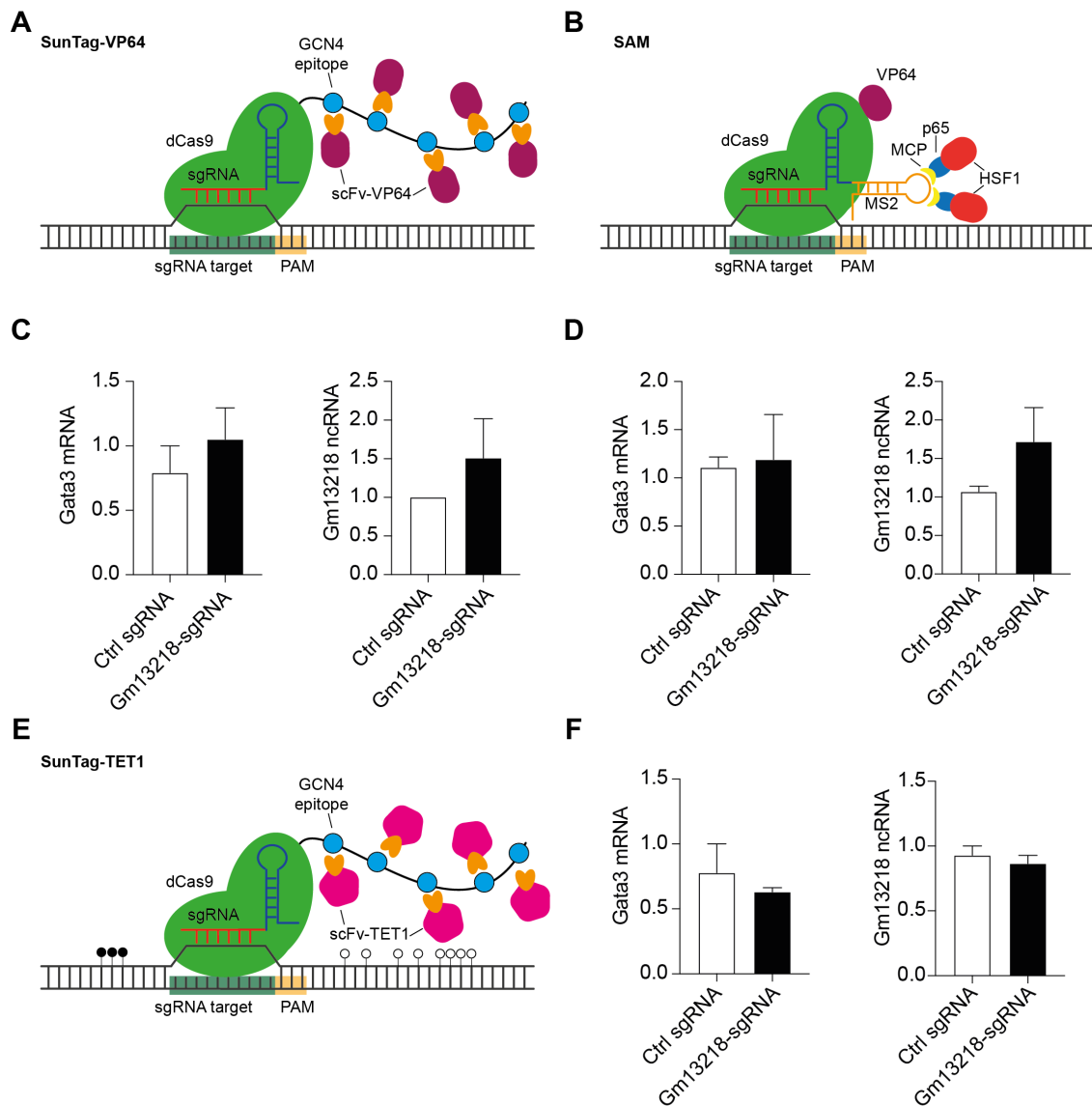


Figure 4.13. CRISPR activation and demethylation failed to activate Gm13218 in A20.

A) SunTag-VP64 activation system, dCas9 fusing to repeats of GCN4 epitopes which recruit multiple copies of VP64. B) SAM, MS2 aptamer bearing sgRNA recruits multiple p65 and Hsf1, assisted also by a single VP64 attached to dCas9. C) Gata3 and Gm13218 expression levels with SunTag-VP64. D) Gata3 and Gm13218 expression levels with SAM. E) SunTag-TET1 DNA demethylation system, repeats of GCN4 recruiting multiple TET1 catalytic domain. Open lollipops represent unmethylated CpG, whereas closed lollipops denote methylated CpG. F) Gata3 and Gm13218 expression levels with SunTag-TET1. All expression levels were normalised to Actb. Shown are mean and SEM from two independent experiments.

Combining CRISPRa with DNA demethylating agent activates Gm13218 but not Gata3

To circumvent the apparent absence of transcriptional activators after DNA methylation, I attempted to combine SunTag-TET1 system with SAM (Fig. 4.14A). I modified the SAM system by replacing VP64 with GCN4 epitope repeats, thus becoming capable of recruiting multiple TET1 while the MS2 aptamer draws the transcriptional activators p65 and Hsf1. Through this modification, Gm13218 was successfully activated in the B cell line A20, transcription of which was upregulated to 80-fold (Fig. 4.14C). However, Gata3 expression remained largely unchanged (Fig. 4.14B).

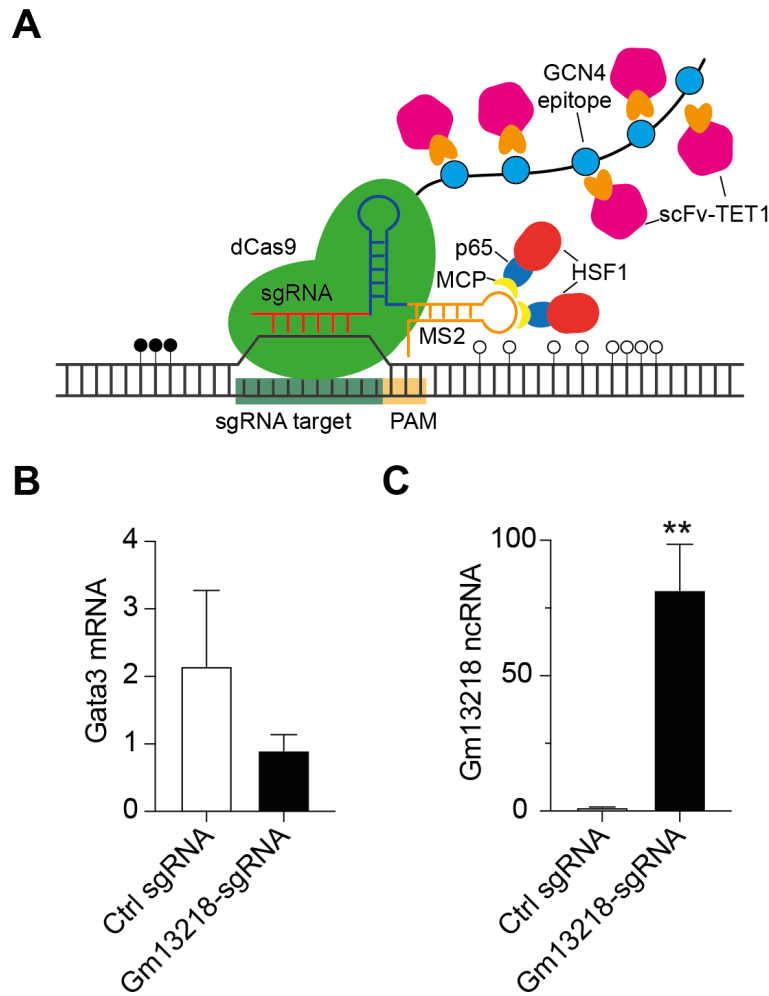


Figure 4.14. Combining SAM with SunTag-TET1 activates Gm13218 but not Gata3.

A) A modified approach to both demethylate CpG while activating transcription. VP64 fusing to dCas9 in SAM is replaced with GCN4 repeats, which then thereby recruit TET1 while MS2 aptamer recruits p65 and Hsf1. B, C) Gata3 and Gm13218 expression levels after DNA methylation and activation, levels were normalised to Actb. Shown are mean and SEM from three independent experiments with ** $p < 0.01$ from *t*-test comparing to control.

4.2. Discussion

Within Gata3 TAD I found a lncRNA Gm13218 transcribed from a previously described enhancer of Gata3 (Hosoya-Ohmura et al., 2011; Ohmura et al., 2016), with the expression highly correlated with Gata3. As a gene regulatory unit, TADs regulate appropriate enhancer-promoter interaction. However, there is a considerable debate around if TADs orchestrate or synchronise the underlying gene transcription (Ibrahim and Mundlos, 2020; Schoenfelder and Fraser, 2019). Indeed, several studies found a positive correlation between TADs and gene expression during differentiation (Dixon et al., 2015; Le Dily et al., 2014; Nora et al., 2012; Symmons et al., 2014). However, active and inactive genes were also found to co-exist in majority of TADs (Le Dily et al., 2014). Similarly, gene expression level can differ drastically within TADs (Hnisz et al., 2016). Nonetheless, it appears that genes that are responsive to external stimuli are more likely to be coordinated within TADs. Using a breast cancer cell line, a study demonstrated about 20% of TADs responded to hormonal stimulation in a concerted manner, with underlying genes being activated or suppressed as a discrete unit (Le Dily et al., 2014). Accordingly, it is of interest to speculate developmentally regulated genes would exhibit a similar pattern. Herein, as the two main transcribed members in the TAD, Gm13218 and Gata3 appear to be co-regulated during T cell development, and also during T_H2 differentiation under the IL-4 stimulus.

Since Gm13218 is transcribed at the enhancer region, it thus superficially appears to be an eRNA. However, Gm13218 defies some other general properties of eRNAs. Transcription at enhancers usually generates bidirectional transcripts (2D-eRNAs), and at a lesser extent, unidirectional transcripts (1D-eRNAs) (Koch et al., 2011; Natoli and Andrau, 2012). While 2D-eRNAs are generally non-polyadenylated and short with length less than 2 kb, 1D-eRNAs are polyadenylated and long, usually over 4 kb in length (Chan et al., 2020a). Gm13218 is the predominant RNA species near the Gata3 enhancer and is polyadenylated, apparently fitting the properties of a 1D-eRNA. However, it is relatively short with a length of around 690 nt (Chan et al., 2020a). More importantly, RNA splicing rarely occurs in eRNAs (Koch et al., 2011), yet

Gm13218 is alternatively spliced, generating multiple RNA species with slight variations at the first exon-intron junction. On the other hand, a sub-group of lncRNAs, later termed as activating ncRNA (ncRNA-a), was identified to possess enhancer-like properties (Lai et al., 2013; Orom et al., 2010). These ncRNA-a contrast with eRNAs in that ncRNA-a are spliced and polyadenylated RNA species, and are unidirectionally transcribed from promoter-like regions (Chan et al., 2020a). As such, it overall appears that Gm13218 is an activating lncRNA that coincides near the Gata3 enhancer.

Many studies have shown knockdown of eRNAs (Lam et al., 2013; Li et al., 2013b; Melo et al., 2013; Mousavi et al., 2013) or stopping transcription (Engreitz et al., 2016; Ho et al., 2006; Isoda et al., 2017; Kornienko et al., 2013; Latos et al., 2012; Ling et al., 2004) result in the reduction of gene expression of the nearby protein-coding genes. However, knockdown of Gm13218, initial silencing of the Gm13218 promoter, stopping Gm13218 transcription or even deletion of Gm13218 promoter had no apparent effects on steady state Gata3 transcription in EL4 thymoma cell line, which expressed high levels of Gata3. In contrast, overexpression of Gm13218 appears to modestly upregulate Gata3 in a B cell line A20. This suggests Gm13218 might have a role in initiating Gata3 expression, rather than a role in maintaining transcription once it has been established. lncRNAs or eRNAs are well known to exhibit very high tissue- or developmental stage-specificity (Fatica and Bozzoni, 2014), with functions only being critical in limited developmental transitions or timeframes. For instance, near the Bcl11b enhancer a lncRNA ThymoD is expressed at DN2 stage to activate the transcription of Bcl11b (Isoda et al., 2017). However, addressing the role of Gm13218 in establishing Gata3 expression in the appropriate setting such as early T cell development or T_H2 differentiation has thus far proven technically challenging (Chan et al., 2020a). Unfortunately, the rapid upregulation of Gm13218 within 24 hours in T_H2 conditions provides only a very limited timeframe and thereby makes it difficult to exploit ASO, overexpression or CRISPR-related techniques that I have readily employed in cell lines. As such, experiments that involve deletion of Gm13218 from the germ line of mice are most likely needed to address the role of Gm13218 in establishing Gata3 expression during early T cell development.

Apart from this spatiotemporal pattern, the importance of lncRNAs are further complicated by the function diversity. As mentioned earlier (Section 1.1, Fig. 1.4), a number of studies have recently demonstrated that lncRNAs or eRNAs are able to modulate gene expression in a plethora of ways. The RNA transcripts can act in cis or trans to recruit various histone modifiers (Nagano et al., 2008; Pandey et al., 2008; Rinn et al., 2007; Wang et al., 2011), or interact with cohesin or mediator complexes to directly assist in chromatin looping (Hsieh et al., 2014; Lai et al., 2015; Lai et al., 2013; Li et al., 2013b). Intriguingly, in some instances the act of transcription *per se* was instead shown to be important. The elongation process of RNA polymerase II can recruit histone modifiers (Ng et al., 2003; Wilson et al., 1996), displace nucleosomes (Hughes et al., 2012; Valouev et al., 2011; Weiner et al., 2010) or evict CTCF and cohesin complex (Busslinger et al., 2017; Heinz et al., 2018; Lefevre et al., 2008; Lengronne et al., 2004), and thereby alter the local chromatin accessibility or three-dimensional configuration. Moreover, the transcription activities of enhancers are capable of relocating the underlying loci to active nuclear compartments such as transcription factories or condensates, which then facilitates enhancer-promoter wirings and hence gene activation (Furlong and Levine, 2018). There is also growing evidence which suggests that RNA transcripts are capable of assisting the formation of phase-separated condensates (Roden and Gladfelter, 2020). Thus, given these complexities the exact functional mechanism of Gm13218 still requires further elucidation.

The prolonged CRISPR silencing mediated by dCas9-KRAB is able to downregulate both Gm13218 and Gata3. While the downregulation of Gm13218 could be mediated by either the occlusion of transcription factors at the promoter or the deposition of histone marks, gene silencing of Gata3 is presumably achieved through the deposition and spreading of H3K9me3 marks to genetically nearby or looped regions (Thakore et al., 2015). Notably, the spreading is very likely to be TAD-specific as TAD boundaries are known to insulate the spreading of chromatin modifications (Narendra et al., 2015). As such, dCas9-KRAB mediated effects are probably TAD specific. As already known, Gata3 is expressed in tissues of haematopoietic, neural, urogenital and cardiac origins,

with different tissues utilising different enhancers (Hasegawa et al., 2007; Hosoya-Ohmura et al., 2011; Lakshmanan et al., 1999; Moriguchi et al., 2018; Ohmura et al., 2016). While the otic and nephric enhancers are located at 571 kb 3' and 113 kb 5' of Gata3 promoter respectively (Hasegawa et al., 2007; Moriguchi et al., 2018), the Gata3 TAD documented here in T cells does not span these regions. It is therefore very likely that Gata3 locus in inner ear and kidney would possess their own distinct topology. As such, by targeting Gm13218, the dCas9-KRAB silencing of Gata3, mediated by spreading of H3K9me3 marks, would only be limited to cells of haematopoietic origin.

Simeonov et al. (2017) has raised the hypothesis where gene activation can be achieved by guiding strong transcriptional activators to the corresponding enhancer in any cells in which the enhancer is inactive. By utilising CRISPRa in the form of a single VP64 attached to dCas9, the study succeeded in identifying several cis-elements of activation-induced CD69 and IL2RA genes in T cells (Simeonov et al., 2017). In contrast, adopting the same strategy failed to activate a T cell-restricted gene in the B cell line A20, through guiding CRISPRa machineries to Gm13218 promoter. While it is possible that activation could be accomplished by targeting other locations along region G2, the failure in even activating the local Gm13218 alone has revealed caveats or complications of, and therefore rejected, the aforementioned hypothesis and suggests that DNA methylation of the target site may preclude gene activation. CRISPRa conducted in SAM or SunTag-VP64 systems, which is more potent than the format consisting of a single VP64 (Chavez et al., 2016), still failed to counteract the silencing effects conferred by DNA methylation, let alone to activate the corresponding gene in long-range. More importantly, in situation where the promoter of the corresponding gene is as well methylated, or the enhancer-promoter loop still requires other cis-elements to pre-establish, it is certainly impossible for the gene to be activated, simply by recruiting transcription activators to the associating enhancer.

While the Gm13218 promoter (G2-1) does not modulate Gata3 expression in steady state, the deletion of the second prominent accessible regions (G2-2) only led to moderate downregulation of Gata3, which is in stark contrast to the drastic

reduction of Gata3 when the entire G2 region is deleted. This suggests that the transcription factor binding sites are scattered across the other open regions, with each separately interacting with the promoter and conferring additive or synergistic modulation. Thus, this also appears to be the reason of the absence of a stable, persistent loop in the Gata3 TAD. Intriguingly, the deletion of G2-2 also led to downregulation of Gm13218. Since the deletion of Gm13218 promoter did not completely abrogate Gm13218 transcription, it could be speculated that G2-2 act as an alternative promoter of Gm13218, as occurred quite commonly in lncRNA (Ziegler and Kretz, 2017). A second plausible speculation could be that G2-2 acts as the enhancer of both Gata3 and Gm13218 – enhancer sharing by multiple promoters is known to be common in genome (Chepelev et al., 2012; Sanyal et al., 2012). In this case, it could be speculated that Gm13218 and Gata3 may have no direct modulatory relationship but are instead regulated by a common enhancer, and Gm13218 could actually serve other unknown functions. Nevertheless, the regulatory circuitry in the Gata3 TAD still needs further disentanglement.

The Gata3 TAD and the synteny between Gata3 and lncRNA was found to be conserved in human. Several reports have recently addressed the evolutionary conservation of TADs (Fudenberg and Pollard, 2019; Harmston et al., 2017; Krefting et al., 2018; Rowley et al., 2017). Through whole genome analyses of different vertebrate species, it was found that the evolutionary DNA rearrangement events occurred most frequently at TAD boundaries (Krefting et al., 2018). Conversely, rearrangements that break the synteny within a TAD is very rare. This indicates TADs are critical regulatory units in which the underlying 3D interactions and hence gene expression patterns are crucial and therefore well preserved through million years of evolution (Chan et al., 2020a). As such, the conservation of synteny and the long-range contact between Gata3 and lncRNA further illustrates the importance of the lncRNA.

Furthermore, the conservation of TAD, synteny and expression pattern in human can potentially be harnessed for manipulative or therapeutic purposes. Gata3 is implicated as a mediator driving allergic asthma (Ano et al., 2013; KleinJan et al., 2014) and T cell lymphoma (Manso et al., 2016; van Hamburg et al., 2008; Zhang

et al., 2016). Strategies have therefore been developed to target Gata3 protein as well as coding mRNA (Garn and Renz, 2017; Nakamura and Hoshino, 2005). However, Gata3 can be expressed elsewhere and thus the interference would have broad and unpredicted consequences. In contrast, lncRNAs or cis-elements of developmentally regulated transcription factors display a much higher tissue- and stage-specificity. For instance, the enhancer-promoter interaction and TAD of Gata3 should be exclusive to T cells and other related immune cells such as NK cells. As such, it is expected that targeting the enhancers through dCas9-KRAB could lead to a more cell type specific modulating effect. With the ever-growing toolbox of genome editing techniques, the deciphering of cis-regulome in a three-dimensional context could pave the way for a finely tuned stage- or lineage-specific adjustment of transcription (Chan et al., 2020a).

5. Characterising genome reorganisation during B cell clonal expansion and differentiation

As explored and discussed in the previous chapters, different lineages and stages of immune cells possess distinct genome architecture, which correspondingly supports the underlying transcriptional programme. However, the process of how this chromatin topology is reconfigured during cellular differentiation and proliferation remains largely unclear. During this developmental transition not only are the cells undergoing transcriptional and architectural changes but also concomitantly they are going through cell division. The challenge is especially profound during activation of B cells, in which cellular changes and proliferation burst within a very brief period of time. In response to antigen stimulation, resting naïve B cells exit G_0 quiescent phase and enter cell cycle. This is accompanied by increases in RNA and protein synthesis, histone acetylation and enlargement of cell volume so dramatic that it is easily observable under a conventional microscope (Chan et al., 2020b; Jaehning et al., 1975; Pogo et al., 1966, 1967; Rawlings et al., 2011). Several recent studies have further shown during this early activation chromatin decondenses, and this requires the transcription factor Myc and energy input in the form of ATP, illustrating it as a highly energy-dependent process (Kieffer-Kwon et al., 2017; Kouzine et al., 2013; Nie et al., 2012). Intriguingly, the first cell division does not occur until in general around 30 hours post-activation, in stark contrast to the subsequent rapid cell division every 6-10 hours thereafter (Hawkins et al., 2009; Turner et al., 2008). These proliferating B cells ultimately differentiate into the effector cells – the antibody secreting plasmablast or plasma cells (Chan et al., 2020b).

Under normal cellular turnover, chromatin condenses and compacts, and becomes devoid of any chromatin loops, TADs and compartments during mitosis (Gibcus et al., 2018; Nagano et al., 2017; Naumova et al., 2013). The three-dimensional structures are then faithfully restored to inherit and resume the precedent conformation after cell division, likely by a recently described phenomenon known as mitotic bookmarking (Oomen and Dekker, 2017). Thus, a more complicated picture emerges during B cell clonal expansion and differentiation, when abolishment and reestablishment of genome architecture

are also accompanied by reconfiguration of structures to support a different transcription programme. While a number of studies have characterised the architectural changes during B cell activation (Bortnick et al., 2020; Bunting et al., 2016; Chaudhri et al., 2020; Javierre et al., 2016; Johanson et al., 2018b; Kieffer-Kwon et al., 2017; Kieffer-Kwon et al., 2013), none has addressed the restructuring event with respect to cell division and cell cycle. As such, the current study aimed to elucidate the genome reorganisation events, during B cell activation, clonal expansion and differentiation, in relation to cell division and cell cycle.

5.1. Results

B cells undergo first cell division after 30-33 hours upon LPS stimulation

To determine the time at which the first cell division occurs upon activation, primary naïve resting B cells, freshly isolated from spleen, were stained with a cell division tracker dye CellTrace Violet (CTV) before being stimulated with lipopolysaccharide (LPS). Upregulation of CD69 at 12 hours after stimulation indicated a majority of population has been activated (Fig. 5.1A). A subsequent time-course analysis with flow cytometry revealed an absence of cell division in the first 28 hours, and the first cell division occurred at around 30-33 hours post-stimulation (Fig. 5.1B). By day 4 a large proportion of B cells has already gone through seventh cell divisions. Intriguingly, there are some undivided cells persisted throughout the entire course of proliferation.

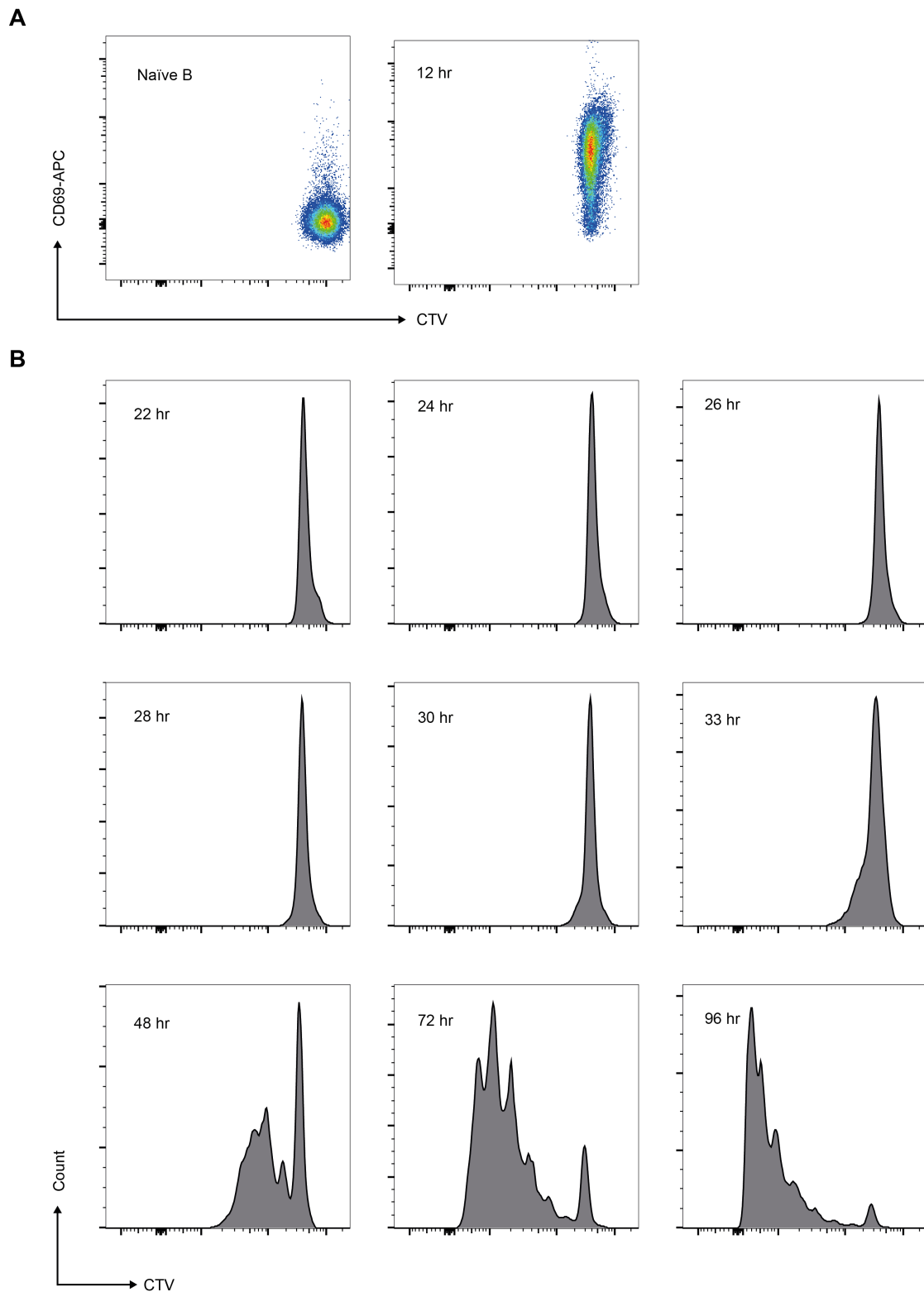


Figure 5.1. B cells undergo cell division after 30-33 hours upon LPS stimulation.

A) Flow cytometry analysis of naïve and activated B cells with activation marker CD69. B) Flow cytometry analysis of B cells at different time after stimulation with LPS. Cells were stained with CellTrace Violet (CTV) to track cell division.

Different stages of activated B cells were harvested for Hi-C and RNA-seq

In a previous study from our lab it was demonstrated, through *in situ* Hi-C, that naïve B cells, activated B cells that have proliferated for 4 days, and differentiated plasmablasts individually possess distinct genome architecture (Johanson et al., 2018b). In order to investigate the topological changes during early activation phase in relation to cell division, I activated naïve resting B cells with LPS, and FACS-sorted activated and undivided cells at 3-, 10- and 33-hour post stimulation (Fig. 5.2B). *In situ* Hi-C was subsequently performed.

To also relate the topological remodelling events to transcriptional changes, the above activated, undivided cells were also harvested for RNA-seq experiments. Since the previous study has not documented any transcriptomic profiles, I therefore also isolated naïve resting B cells (Fig. 5.2A), 4-day activated expanded B cells and differentiated plasmablast (Fig. 5.2C) for RNA-seq.

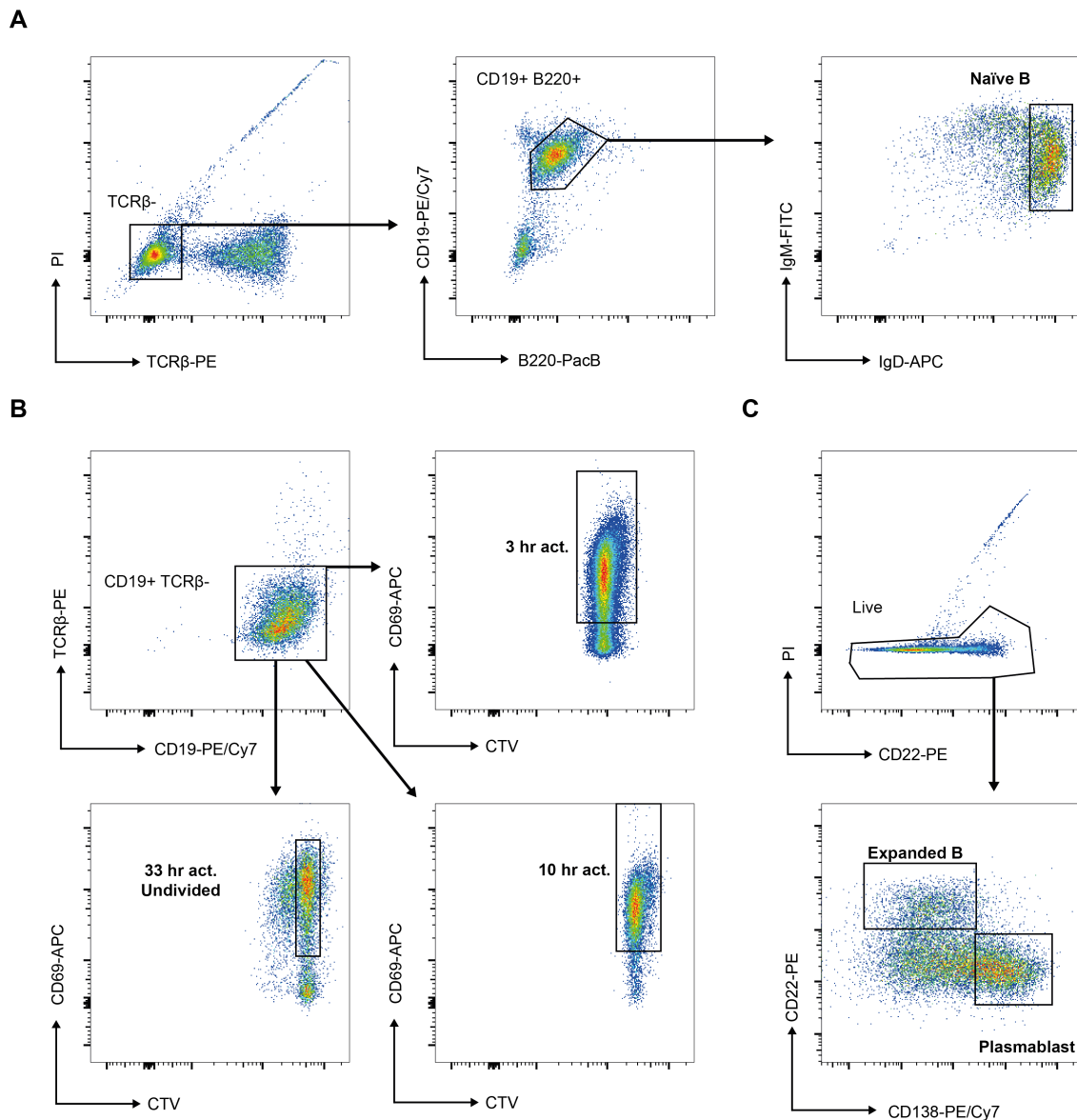


Figure 5.2. Isolation of different stages of activated or differentiated B cells.

A) Gating strategy for naïve B cells. Splenocytes with TCR β - CD19+ B220+ IgM+ IgD^{hi} were sorted as naïve B cells. B) Activated, undivided B cells were sorted at 3-, 10- and 33-hours after LPS stimulation as TCR β - CD19+ CD69+ CTV^{hi} population. C) Activated, expanded B cells (CD22+ CD138-) and differentiated plasmablast (CD22- CD138+) were isolated after 4 days of activation and proliferation.

Two discrete waves of genome restructuring across B cell activation and differentiation

Hi-C was performed in duplicates for each activation or differentiation stage, with changes in genome organisation, named differential interactions (DIs), determined using *diffHic* pipeline (Lun and Smyth, 2015). In brief, *diffHic* aligns read-pairs, which represent the captured DNA-DNA interactions, to the reference genome that has been partitioned into discrete bins for quantification. This is followed by identification of statistically significant difference in interaction intensity, represented as DIs, between samples.

Using this analysis pipeline, it was revealed that the genome architecture stayed very much invariant for the first 10 hours after LPS stimulation (Fig. 5.3) – there are only very few DIs in the first 3 hours, and almost no change for the next 7 hours. Intriguingly, this is followed by a massive, discrete wave of restructuring event prior to the first cell division between 10- and 33-hours post stimulation, with 10628 DIs. This demonstrates that genome is able to undergo global topological remodelling without the need of cell division.

Despite spending roughly an additional 60 hours in culture and having undergone many cell divisions, the expanded B cell population showed very little alteration in genome architecture to their 33-hr counterparts (Fig. 5.3). Interestingly, this is followed by another discrete wave of restructuring event upon the terminal differentiation into plasmablast, with 6784 DIs. As such, throughout the final stage of B cell differentiation, there are two discrete waves of genome reorganisation events, with the first occurs prior to first cell division, and the second happens upon final differentiation. Additionally, it overall appears that genome reorganisation occurs as two discrete waves, with chromatin undergoing restructuring in a global and concurrent fashion.

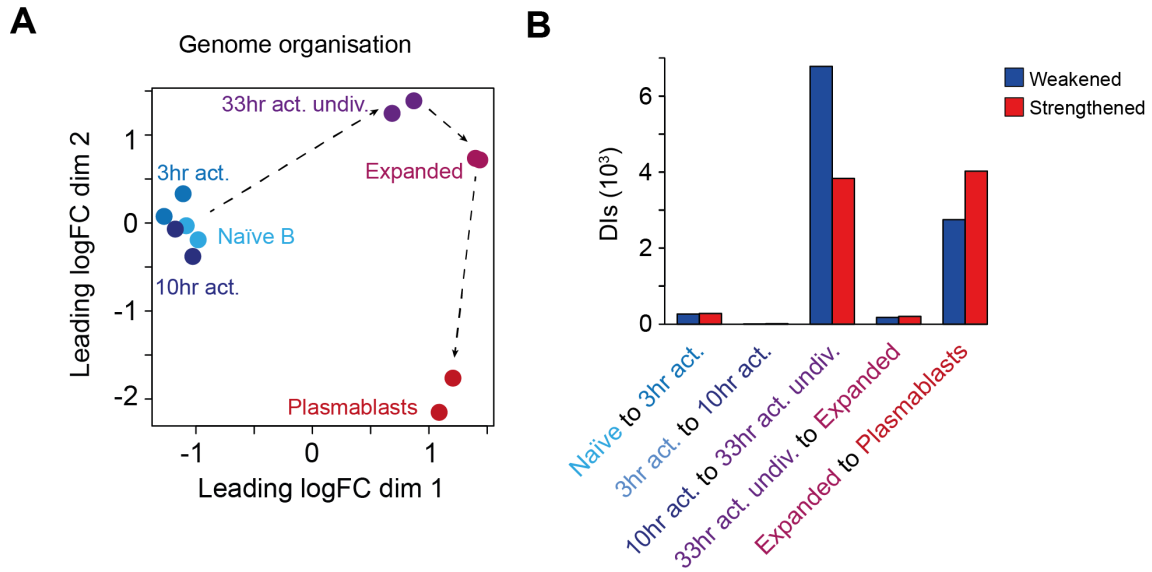


Figure 5.3. Two discrete waves of genome reorganisation events across B cell activation with first occurs prior to first cell division.

A) Multidimensional-scaling plot of genome architecture of different activation and differentiation stages of B cells. Dashed lines illustrate the developmental trajectories. B) Quantification of differential interactions (DIs) between developmental transitions of activated B cells. Blue indicates weakening of structures during the corresponding transition, whereas red represents strengthening of structures.

The first major genome reorganisation event occurs prior to first cell division

Illustrating the first wave of chromatin restructuring event, prior to the first cell division, are structures around *Twistnb* and *Prdm1* loci (Fig. 5.4). *Twistnb* encodes a component of RNA polymerase I complex, with promoter interacting with a distal site roughly 2 Mb apart. This loop persists for the first ten hours of B cell activation, but is significantly weakened just prior to the first cell division during the first wave of restructuring, and is completely lost afterwards.

Prdm1, on the other hand, encodes the transcription factor Blimp-1 which is very critical for plasmablast differentiation. The interaction between the *Prdm1* promoter and a distal site 600 kb away is strengthened upon the first restructuring event, and persists thereafter (Fig. 5.4). Intriguingly, although Blimp-1 is expressed only immediately before the final plasmablast differentiation process (Shaffer et al., 2002; Shapiro-Shelef et al., 2003), this loop is already established prior to first cell division.

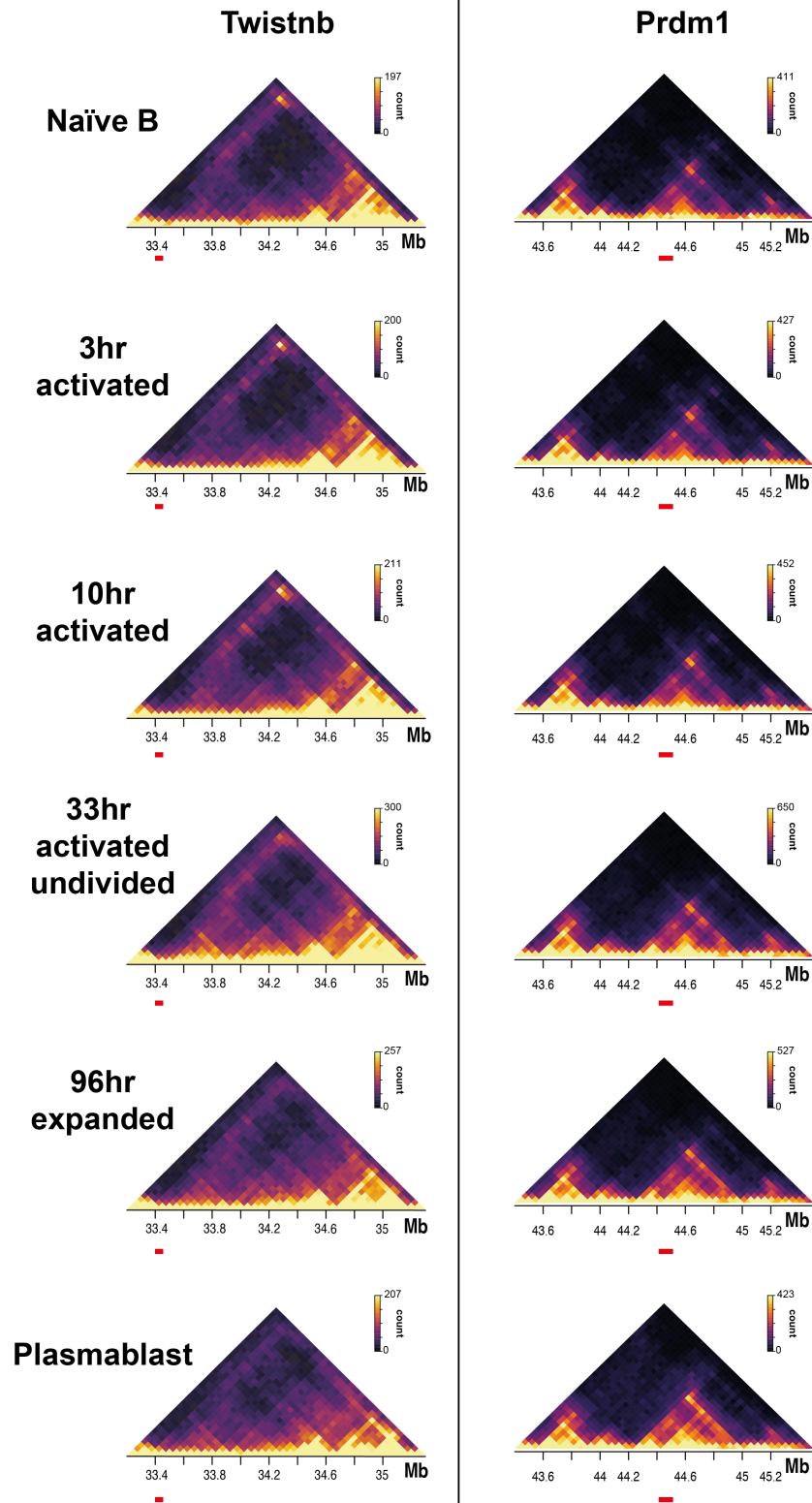


Figure 5.4. Genome reorganisation events prior to first cell division.

Hi-C contact matrices around *Twistnb* (left, chr12) and *Prdm1* (right, chr10) loci in different stages of activated and differentiated B cells. Red bars below the map indicate the corresponding loci.

The second major genome reorganisation occurs upon differentiation into plasmablast

Examples of the second wave of chromatin restructuring, upon the terminal differentiation into plasmablasts, are structures around *Ebf1*, *Bcl6* and *Myc* loci (Fig. 5.5). *Ebf1* promoter interacts with several distal regions, these structures exist before and shortly after activation, and are lost upon plasmablast differentiation. Similarly, structures around *Bcl6* persist throughout the entire period of B cell activation, only to disappear upon plasmablast differentiation.

Although genome structures around *Myc* show a higher degree of temporal dynamics during B cell activation, interactions upstream of *Myc* (61.1-62 Mb, Fig. 5.5) are significantly lost upon the terminal differentiation.

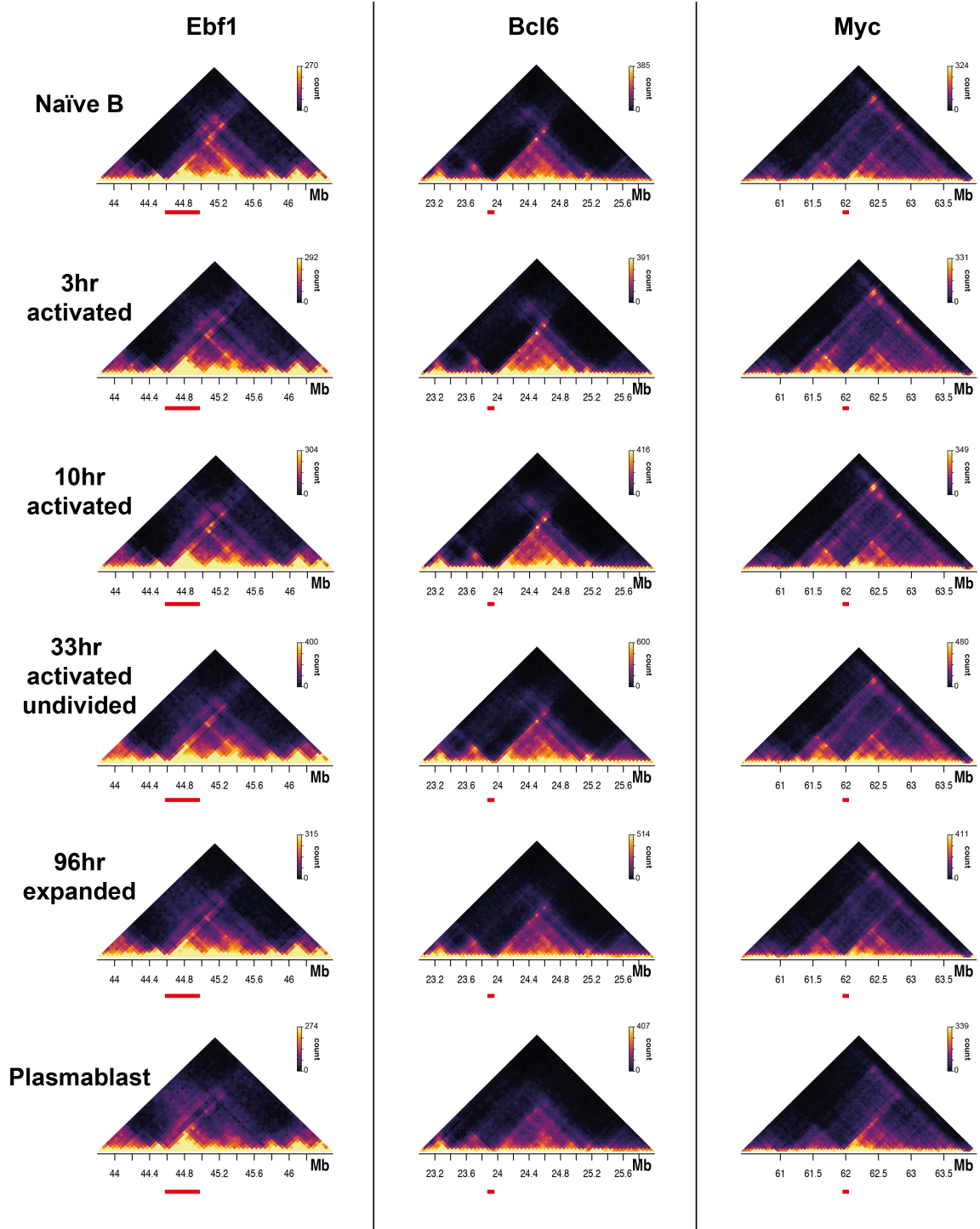


Figure 5.5. Genome reorganisation events upon terminal differentiation. Hi-C contact matrices around *Ebf1* (left, chr11), *Bcl6* (middle, chr16) and *Myc* (right, chr15) loci in different stages of activated and differentiated B cells. Red bars indicate the corresponding loci.

TADs increase in size immediately prior to the first cell division

I next examined how larger genomic structures were altered during B cell activation and differentiation. Further analysis of data with *diffHic* uncovered that a higher proportion of TAD boundaries became weakened just prior to the first cell division, at 33-hour post stimulation (Fig. 5.6A), whereas roughly similar number of boundaries are strengthened or weakened during other transitions. By using another analytical package TADbit (Serra et al., 2017), which allows the detection and quantification of TADs across the entire genome, it was revealed that TADs are generally larger in size just prior to the first cell division (Fig. 5.6B). Intriguingly, while there are similar number of TADs of size larger than 800 kb in different activation stages (Fig. 5.6C), B cells at 33-hour harbours relatively less TADs of size smaller than 800 kb (Fig. 5.6D). A further examination of individual DIs using *diffHic* at different transitions has revealed that long-range interactions are weakened just prior to the first division (Fig. 5.6E). Thus, overall it appears that the loss of TAD boundaries and long-range interaction at this stage led to the loss of TADs that are <800 kb, and hence a general increase in TAD size. This might have reflected that TAD insulation is weakened by the DNA replication process, or in preparation to mitosis (Nagano et al., 2017).

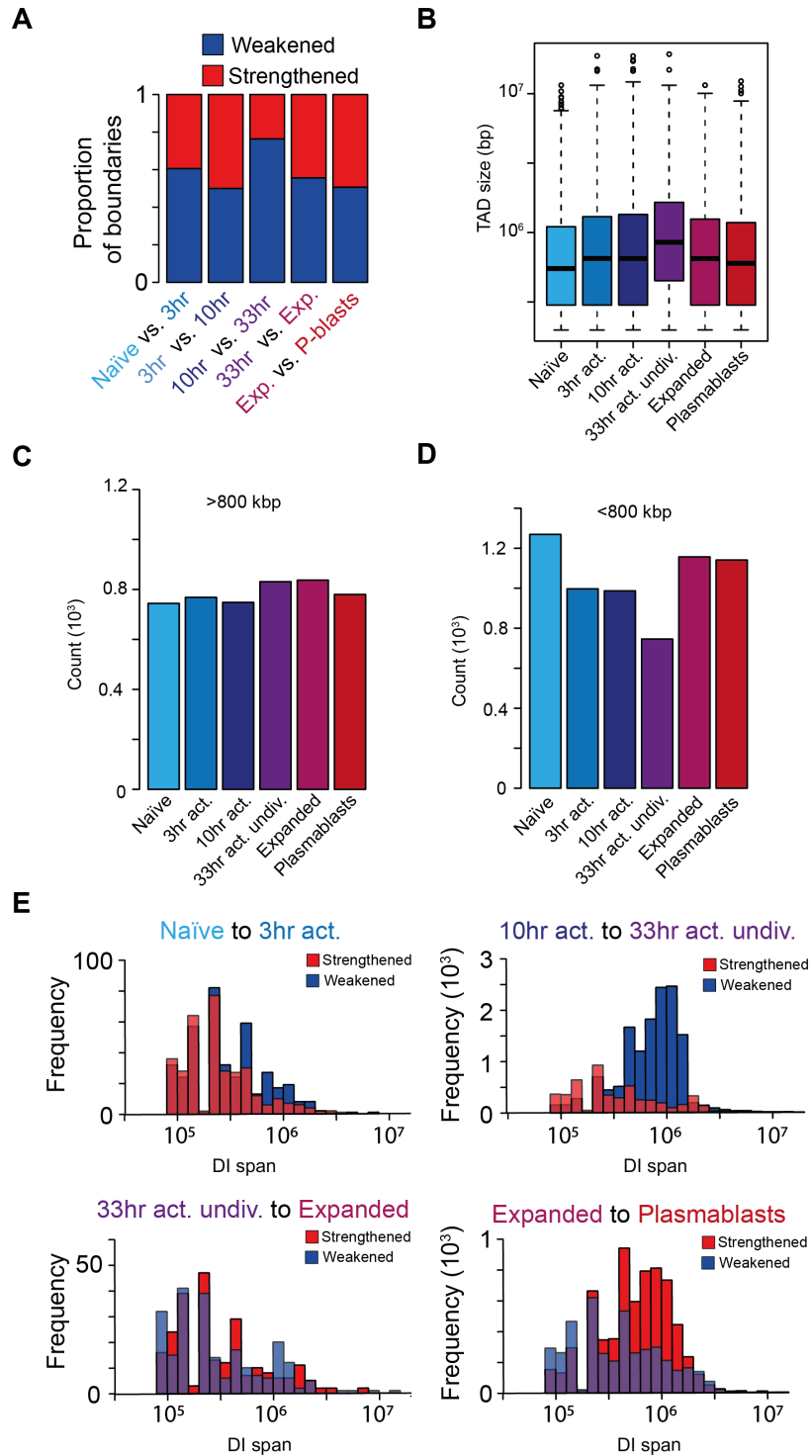


Figure 5.6. TAD increase in size immediately prior to first cell division.

A) Proportion of TAD boundaries strengthened or weakened at different transitions during B cell activation and differentiation. B) Distribution of TAD size at each developmental stage. C) Number of TADs of size larger than 800 kb at each developmental stage. D) Number of TADs of size smaller than 800 kb at different stages. E) Distribution of DI spans, i.e. distance between anchors of DIs, that are strengthened and weakened at different transitions.

Compartments are largely invariant across B cell activation and differentiation

As introduced and discussed in chapter 1, compartments are also important topological structures besides TADs. Genes have been documented to undergo compartmental switch between A- and B- compartments during stem cell differentiation (Dixon et al., 2015) as well as early B- and T- cell development (Hu et al., 2018; Isoda et al., 2017; Lin et al., 2012). It is therefore informative to investigate how compartments changes during B cell activation, by using principal component analysis (PCA). Surprisingly, there were very few compartment switches before the first cell division (Fig. 5.7), and also very few genes had undergone A/B compartment switching during clonal expansion and terminal differentiation into plasmablasts – less than 10 genes had ever switched compartments during the process.

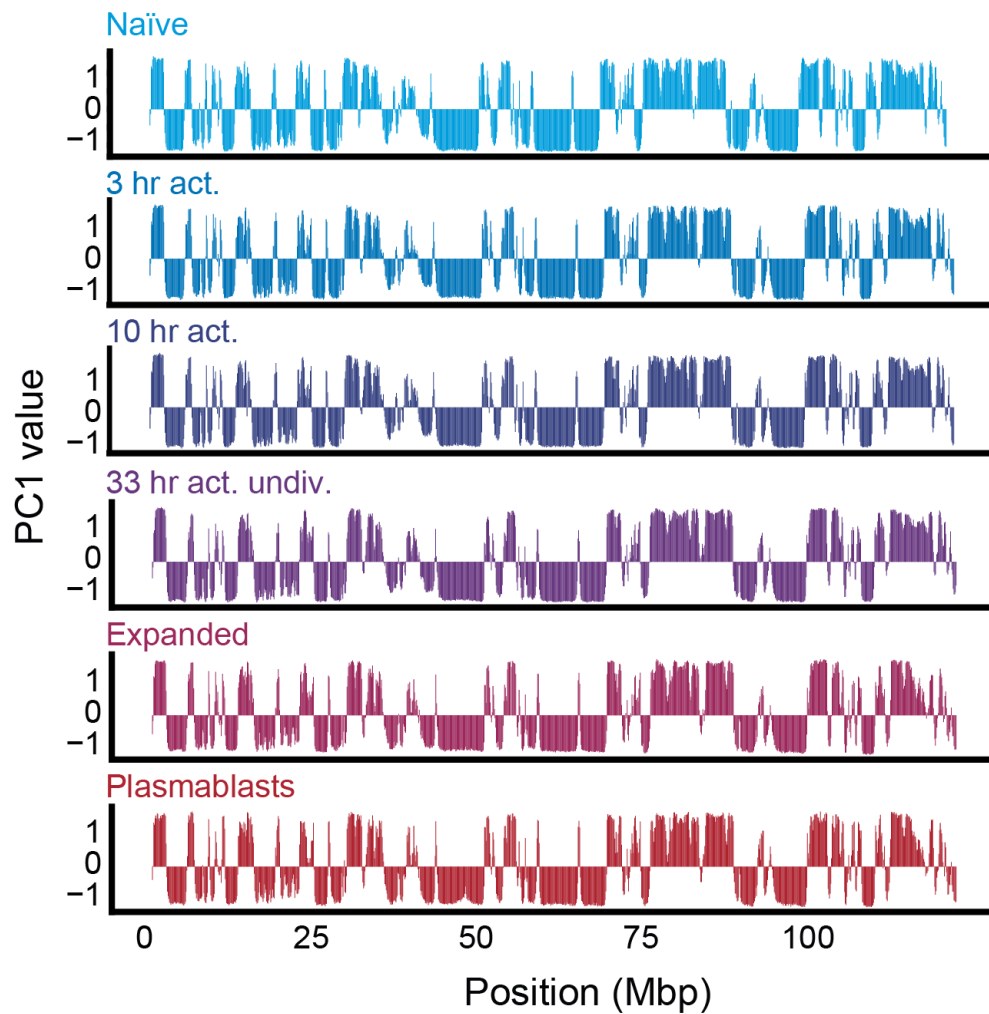


Figure 5.7. Compartments are largely invariant across B cell activation and differentiation.

Compartment profiles showing the first principal components of the entire chromosome 12 for different B cell activation and differentiation stages. A-compartments are of positive PC1 scores while B-compartments bear negative PC1 scores.

Dynamic transcriptional changes throughout the course of B cell activation and differentiation

Next, to relate the chromatin restructuring events with transcriptional changes, transcriptomic data obtained from RNA-seq was analysed. It was revealed that transcriptional changes burst as early as 3-hour post activation (Fig.5.8), despite the absence of genome reorganisation events – there are 5838 differentially expressed genes (DEGs) within this 3-hour period. Surprisingly, a further 3963 DEGs were observed in the next 7 hours, and another 1877 genes have their expression altered when B cells were about to transit through the first cell division. Transcriptional changes did not stop thereafter, with 1371 DEGs observed during the clonal expansion, followed by another 3340 DEGs upon the terminal differentiation into plasmablasts. Thus, it appears that transcription profiles are much more dynamics and undergo changes throughout the entire course of B cell activation and differentiation. This is in stark contrast to genome architecture, which is restructured only at particular stages.

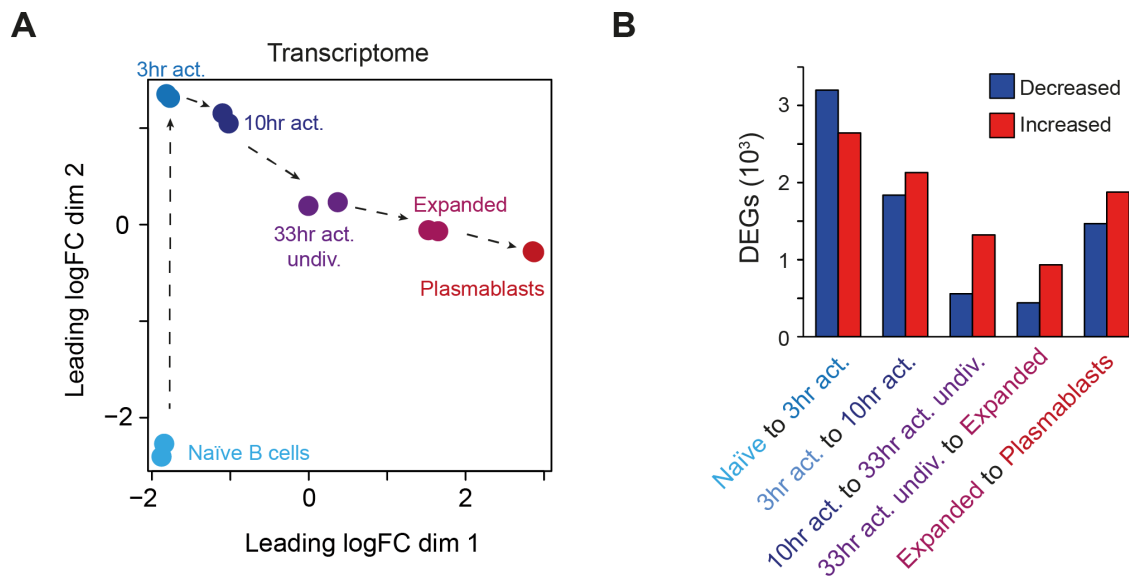


Figure 5.8. Dynamic transcriptional changes throughout the course of B cell activation.

A) Multidimensional-scaling plot of transcription profiles of different activation and differentiation stages of B cells. Dashed lines illustrate the developmental trajectories. B) Quantification of differentially expressed genes (DEGs), calculated as fold change significantly above 1.5 (Treat FDR<0.05), between developmental transitions of activated B cells. Blue indicates weakening of structures during the corresponding transition, whereas red represents strengthening of structures.

Gene ontology unveils the stage-specific cellular changes across B cell activation

The dynamics of transcriptional changes across B cell activation can be illustrated by certain important B cell genes (Fig. 5.9A). For instance, Blimp-1, encoded by *Prdm1*, the master regulator of plasma cell differentiation (Shaffer et al., 2002; Shapiro-Shelef et al., 2003), is upregulated upon the plasmablast stage. Conversely, a number of B cell signature genes, such as *Ebf1*, *Pax5* and *Bcl6*, are suppressed upon differentiation into plasmablast. On the other hand, *Myc*, an important transcription factor involved in metabolism and also for amplifying transcription (Nie et al., 2012), is mostly upregulated during the post-stimulation period before the first cell division. Whereas *Aicda*, encoding the activation-induced cytidine deaminase (AID) which is involved in class switch recombination and somatic hypermutation (Muramatsu et al., 2000), is only upregulated during the clonal expansion period.

The massive and dynamic transcriptional changes also prompted us to investigate the details at each transition. With gene ontology analysis, it was uncovered that during the first 3 hours upon LPS stimulation gene expression involving ribosome biogenesis and RNA metabolism are significantly altered (Fig. 5.9B). For the next 7 hours, a number of metabolic genes instead have the expression changed. This appears to fit with the upregulation of *Myc*, which regulates transcription and cellular metabolism. Intriguingly, when B cells were approaching the first cell division at 33 hours post-stimulation, it was revealed that transcriptional changes switched the focus to DNA related functions – replication, chromatin assembly and conformation change. This, again, is consistent with the genome reorganisation event occurred at this transition. Lastly, upon terminal differentiation into plasmablast, transcriptional changes are mainly found to be genes involved in protein trafficking and ER stress, which is presumably reflecting the massive production and secretion of antibodies in plasmablasts.

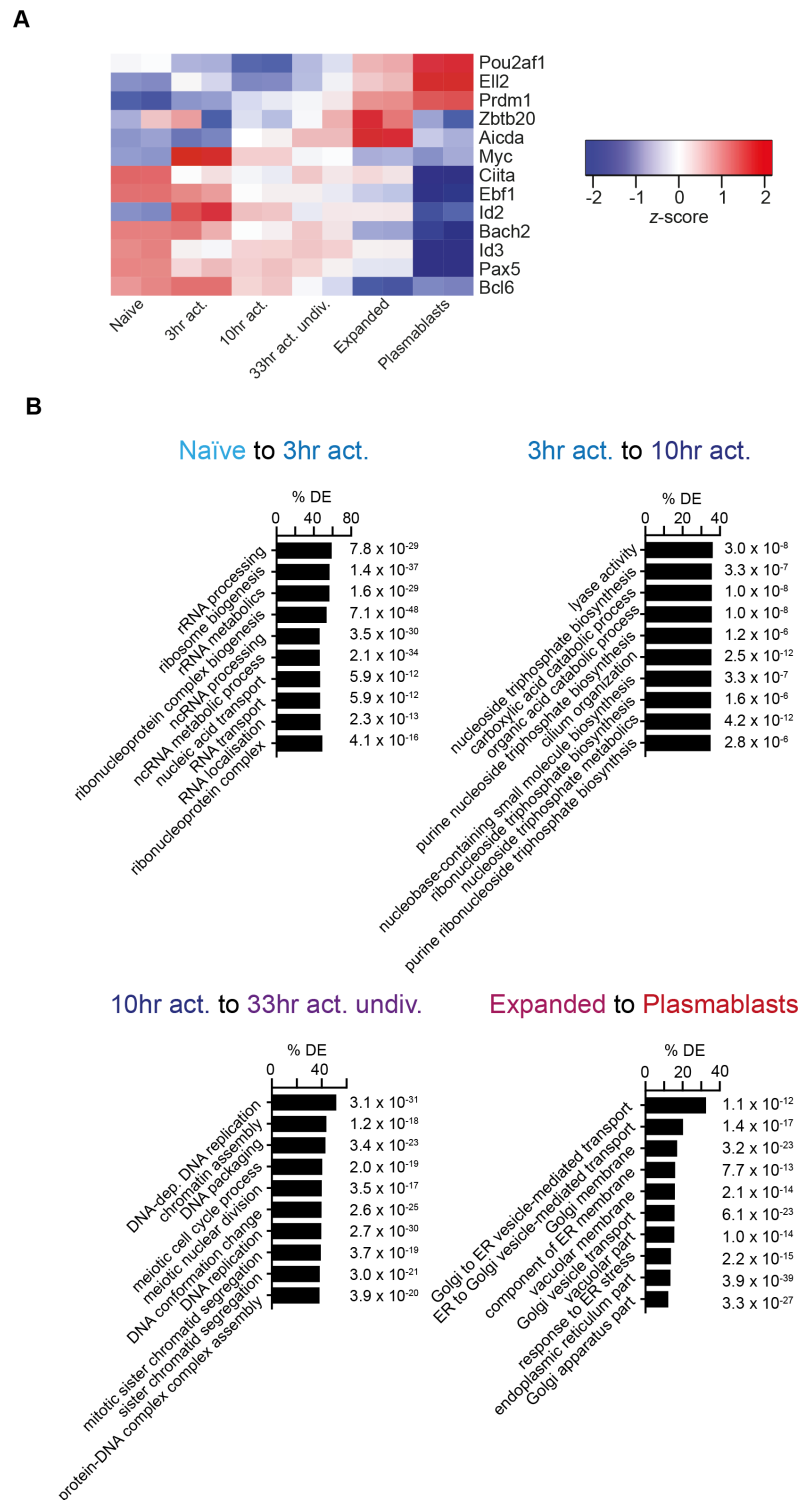


Figure 5.9. Gene ontology unveils the stage-specific cellular changes across B cell activation.

A) Heat-map showing relative transcription levels of select genes across B cell activation. B) Gene ontological terms with genes increasing in differential expression overrepresented at the indicated transitions. Groups ranked according to the percentage of genes that are differentially expressed.

Remodelling of interactions around promoters correlates with transcriptional changes

To correlate genome reconfiguration with transcriptional changes, DNA interactions involving all protein-coding promoters were bioinformatically identified and isolated. As such, the promoter interactivity switch at a particular transition, or herein referred as differentially interacting promoters (DIPs), can then be correlated with the DEGs at the same transition. For a number of B cell genes, it was shown that interactivity does reflect expression (Fig. 5.10A). For instance, during the course of B cell activation, expression of *Pax5* and *Bcl6* decreases along with the loss of interactions around the promoters, whereas interactivity involving *Prdm1* promoter increases with gene expression.

At a global level, it was revealed that during the two genome reconfiguration events – immediately prior to the first cell division and upon plasmablast differentiation, the chromatin remodelling around promoters does significantly correlate with gene expression (Fig. 5.10B, C). In other words, increase in interactivity around promoter is generally accompanied as well by an increase in expression.

In contrast to the above two transitions, transcriptional changes during the first 3 hours post-stimulation does not associate with any genome reorganisation event. Presumably the transcription bursts are driven by the pre-established chromatin structures. Indeed, upon analysis it was unveiled that the existing loop structures in naïve resting B cells correlates well with the transcriptional changes undergone within the 3 hours of activation (χ^2 test, $P = 2.46 \times 10^{-6}$).

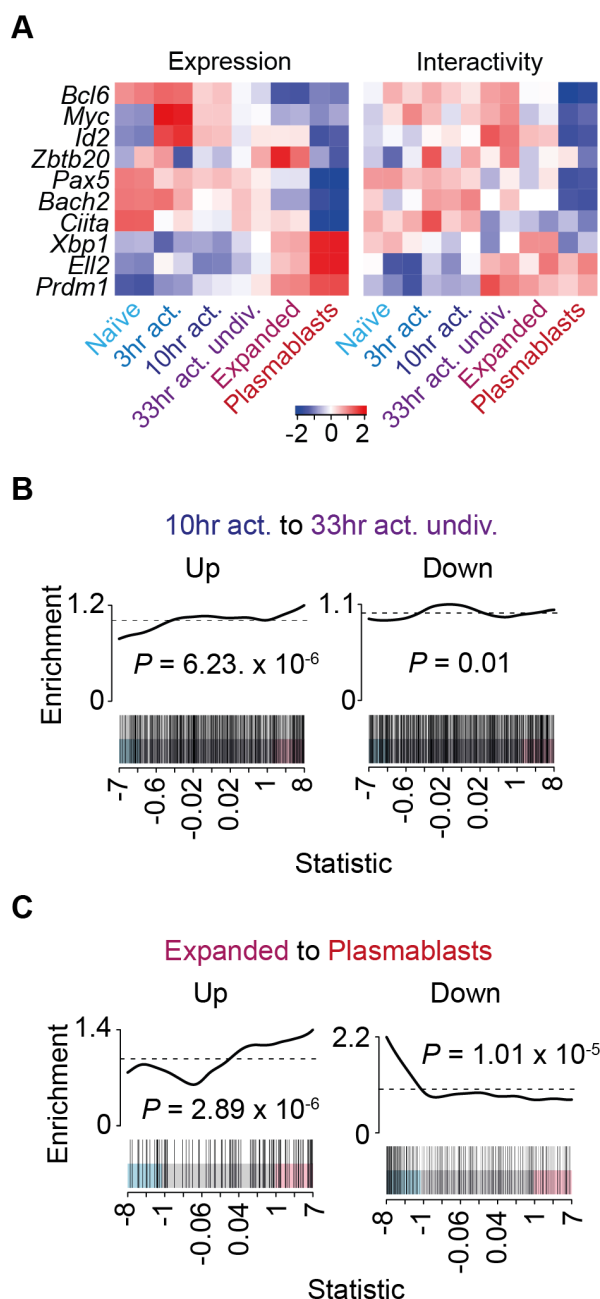


Figure 5.10. Remodelling of interactions around promoters correlates with transcriptional changes.

A) Heatmaps showing relative changes in expression and promoter interactions in select B cell genes across B cell activation and differentiation. B, C) Barcode enrichment plots showing the correlation between promoters that significantly increase/decrease in associating interactions with increase/decrease in gene expression, at the transition from 10- to 33-hour post-stimulation, and upon plasmablast differentiation, as indicated. Genes expressions are ranked on the plot in an ascending order (most downregulated to most upregulated) from left to right. *P*-value was calculated with the fry test.

Genome reorganisation occurs before S phase prior to first cell division

While it has been established that the first genome restructuring occurs prior to the first cell division, it remains unclear if the restructuring event was executed in a particular phase of cell cycle or throughout interphase. To tackle this, a cell cycle indicating system – the Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI) – was exploited (Sakaue-Sawano and Miyawaki, 2012; Tomura et al., 2013). With two fluorescent proteins – mKO2 (red) and mAG (green), cell cycle can be visually stratified into four different colours. Red colour denotes majority of G₁ phase, whereas green inclusively indicates S, G₂ and M phases, with the presence (yellow) and absence (colourless) of both fluorescent proteins mark the transitions between G₁-S and M-G₁, respectively. In addressing the genome reorganisation event in relation to cell cycle, akin to my previous approach, I stimulated naïve B cells from the FUCCI mice with LPS for 33 hours, and isolated the activated but undivided cells that were still in the G₁ (red) phase (Fig. 5.11A). This population has yet to initiate DNA replication, or enter S phase, and is therefore denoted as pre-S. Intriguingly, it was unveiled that this pre-S population has already undergone a vast majority of genome reconfiguration (Fig. 5.11B, C). Thus, it appears that genome reorganisation at this transition precedes, and is therefore independent of, DNA replication.

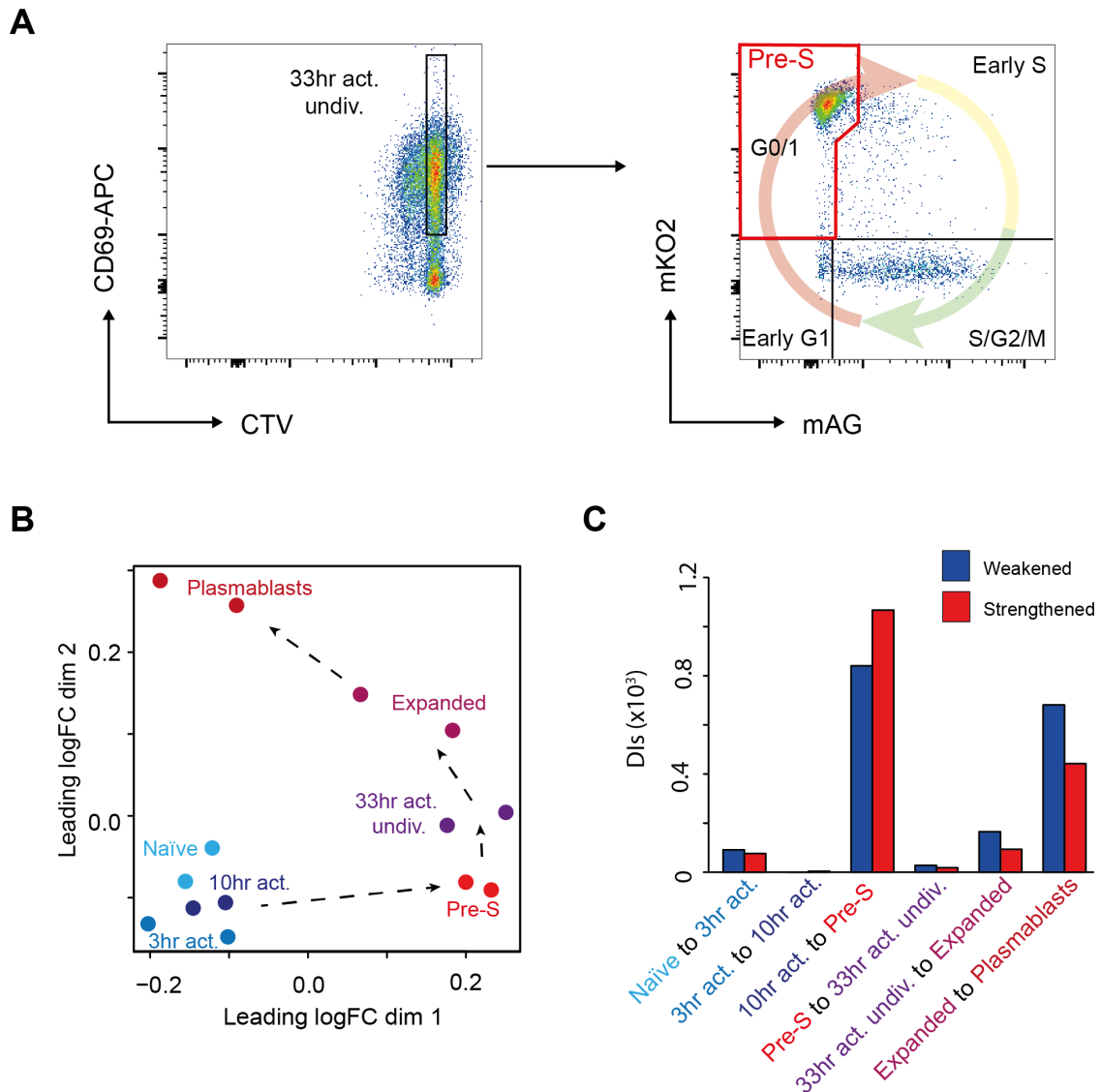


Figure 5.11. The first genome reconfiguration event occurs prior to DNA synthesis phase.

A) Gating strategy for isolating undivided, activated B cells at the G₁ phase at 33-hour post-stimulation. B) Multidimensional-scaling plot of genome architecture of different activation and differentiation stages of B cells. Dashed lines indicate the developmental trajectories. B) Quantification of differential interactions (DIs) between developmental transitions of activated B cells. Blue indicates weakening of structures during the corresponding transition, whereas red represents strengthening of structures.

Terminal differentiation into plasmablast associates with prolonged time in G₁ phase

Since the first genome reconfiguration event occurs during a prolonged G₁ phase, it is therefore interesting to investigate if the second wave of reorganisation upon plasmablast differentiation is related to specific stages of cell cycle. Proliferating cells are no longer synchronous in cell division and differentiation is largely independent of division number, it is thus challenging to isolate a population that is immediately about to differentiate. I therefore explored this by analysing the cell cycle status of the proliferating cells in different stages (Fig. 5.12). As illustrated previously, clonally expanding activated B cells and plasmablasts can be identified with CD22 and CD138 surface markers. In addition, there is an intermediate, transitional population known as pre-plasmablast that has downregulated CD22 and has yet upregulated CD138 (Kallies et al., 2007; Minnich et al., 2016). As such, with CTV and FUCCI system, the cell cycle phases of the proliferating cells in each differentiation stage were analysed.

It was found that over time, a higher proportion of the expanded B cell population were found in G₁ phase, suggesting these cells are spending an increasing time in this phase (Fig. 5.13A). The trend is also true regarding the pre-plasmablast population (Fig. 5.13B), however, the proportion of differentiated plasmablasts in G₁ phase was reduced, suggesting these cells may have begun to rapidly cycle again (Fig. 5.13C). Intriguingly, the transitional pre-plasmablast, compared with other stages, spent the highest amount of time in G₁ phase over the 5-day period (Fig. 5.13D-F). Once undergone differentiation, plasmablasts significantly spent less time in G₁. Thus, it appears that the terminal differentiation, during which the second genome reconfiguration event occurs, is generally associated with a lengthening of G₁ phase, especially at the transitional pre-plasmablast stage.

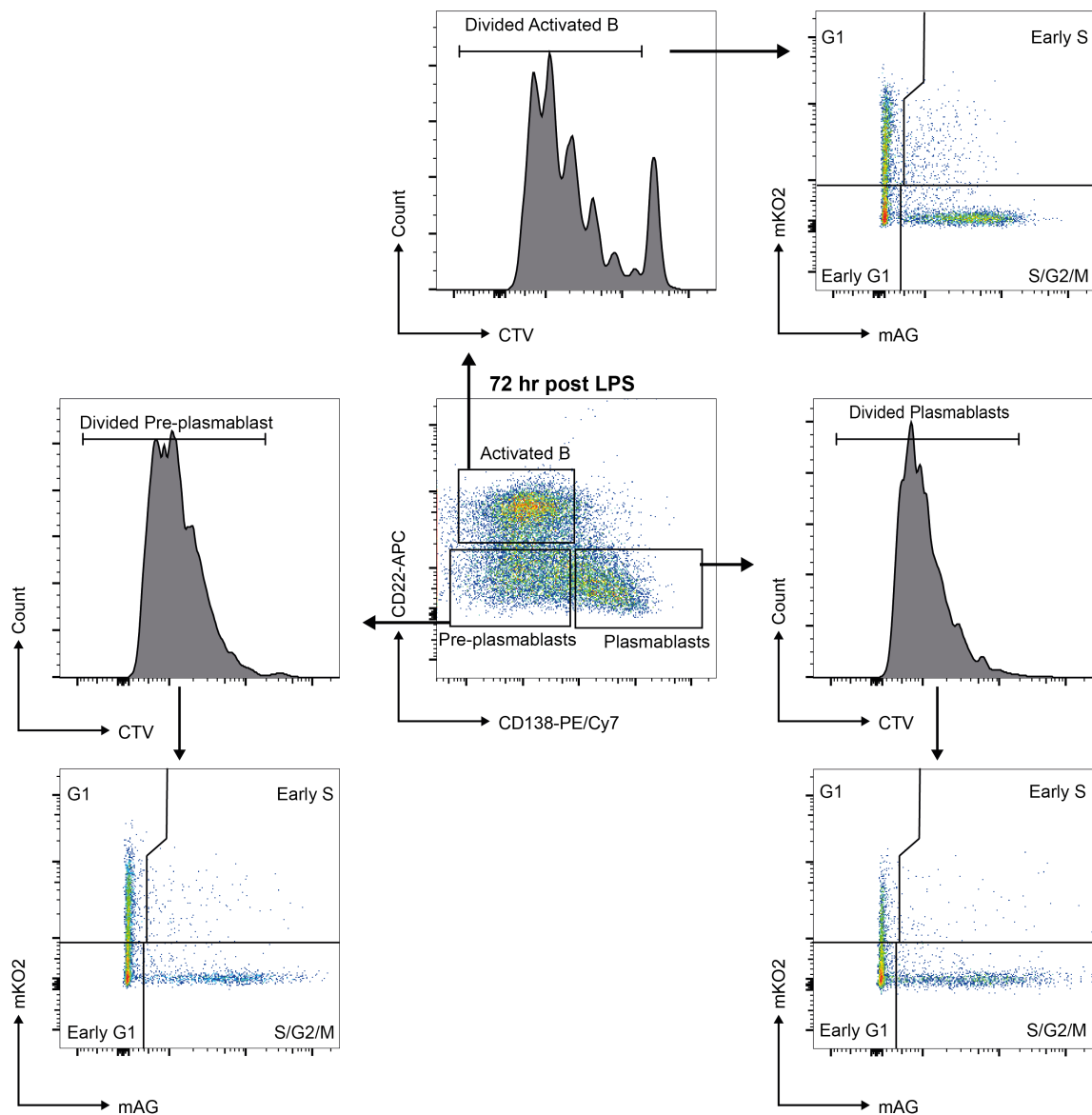


Figure 5.12. Cell cycle analysis of LPS activated FUCCI B cells over a five-day period.

Flow cytometry analysis showing strategies for determining cell cycle phases of proliferating activated B cells, pre-plasmablasts and plasmablasts in a five-day time course study. Shown is example from culture day 3 post LPS stimulation.

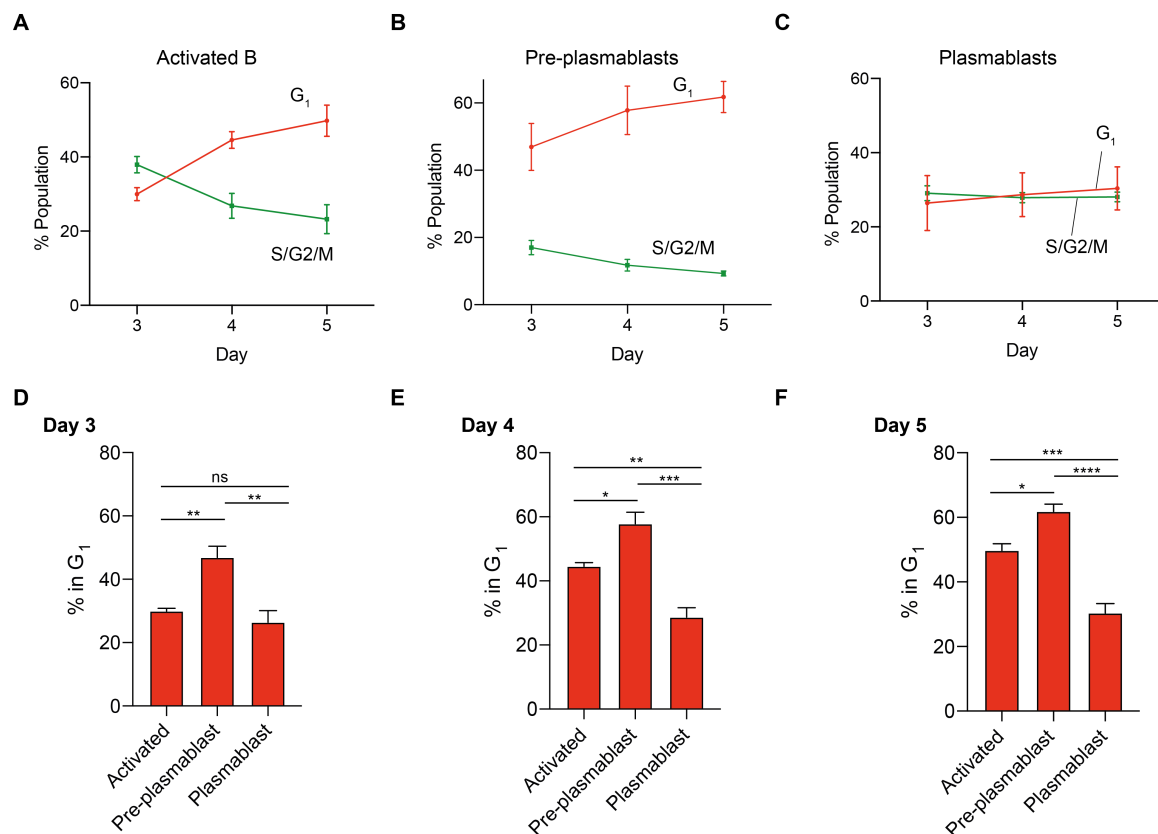


Figure 5.13. Terminal differentiation into plasmablast associates with prolonged G₁ phase.

A) The percentage of FUCCI activated B cells in G₁ or S/G₂/M phases of the cell cycle from day 3 to day 5 post stimulation. B) The percentage of FUCCI pre-plasmablasts in G₁ or S/G₂/M phases of the cell cycle from day 3 to day 5 post-stimulation. C) The percentage of FUCCI plasmablasts in G₁ or S/G₂/M phases of the cell cycle from day 3 to day 5 post stimulation. Shown are mean with SD from four independent experiments. D, E, F) The proportion of G₁ cells in activated B cell, pre-plasmablast and plasmablast population at day 3, 4 and 5 post-stimulation, as indicated. Shown are mean and SEM from four independent experiments, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$ from one-way ANOVA with post-hoc Tukey test.

5.2. Discussion

Here it was discovered that genome reconfiguration occurs in two discrete waves during B cell activation and terminal differentiation, whereupon chromatin undergoes restructuring in a global, concurrent manner within a relatively limited timeframe. This is in stark contrast to transcription, which is continuously altered throughout the developmental process. Genome conformational changes involve establishment and abolishment of loops, loop extrusion via the action of cohesin, as well as deposition and removals of histone modifications. These processes require a large amount of energy input (Kieffer-Kwon et al., 2017) and are therefore under stringent control. Our transcriptomic profiling of cells during the first ten hours after stimulation, during which the chromatin topology remained unaltered, may indirectly reflect the stringency of the process of genome reorganisation. In the hours shortly after activation, B cells appear to dedicate their energy towards transcription, RNA processing and ribosome biogenesis, with focus switched to other metabolic processes several hours after. Only after these metabolic investments do B cells initiate the DNA-related processes such as conformational changes, chromatin remodelling and others. Thus, it appears that there is a stepwise cellular process that precedes genome reconfiguration.

The first wave of genome reorganisation was unveiled to occur prior to the first cell division, and even before the first DNA replication event. Since the discovery of the complete abolishment of 3D structures, due to massive chromatin condensation, during mitosis (Gibcus et al., 2018; Nagano et al., 2017; Naumova et al., 2013), it was suggested that the faithful restoration of structures in non-differentiating cells is attributed to the mitotic bookmarking (Oomen and Dekker, 2017). The factors involved in this phenomenon were proposed to be the existing histone variants, histone modifications, DNA methylation, non-coding RNAs and transcription factors, which remain adhered to the condensed chromatin during mitosis, and hence guide the faithful re-establishment of structures upon mitotic exit. This has raised a number of questions and interest as to how 3D structures reorganise and how bookmarking is altered during cellular differentiation. One speculation is that the reorganisation is unlikely to happen in G₁ phase (Oomen and Dekker, 2017), given that chromatin has limited diffusion rate (Dekker and

Mirny, 2016; Rosa and Everaers, 2008) and occupies a larger space in interphase nucleus. Contrarily, mitosis appears to be a more convenient window for cells to lay down new bookmarks and reorganise the genome when chromatin start to decondense (Gollosi et al., 2017; Oomen and Dekker, 2017). Alternatively, DNA replication has been speculated to be another window of opportunity (Bird et al., 1998; Richter et al., 1999; Weintraub et al., 1978; Wolffe, 1991). However, the data herein suggests otherwise, that the genome restructuring precedes mitosis and DNA replication, and occurs instead in G₁ phase. As such, it appears that the reconfiguration is partitioned from these two similarly physically challenging, DNA-related processes. In addition, it also suggests that the new mitotic bookmarks are laid down before cell division, along with the genome restructuring event.

In agreement with a number of previous studies (Bunting et al., 2016; Javierre et al., 2016; Johanson et al., 2018b; Kieffer-Kwon et al., 2017; Kieffer-Kwon et al., 2013), it was found that the pre-mitotic genome restructuring is associated with remarkable compartment stability. The lack of compartmental changes suggests that the majority of architectural changes are mainly DNA loops. Intriguingly, this is in striking contrast to stem cells and early lymphocyte differentiation, in which an extensive A/B compartment switch occurs during lineage specification (Dixon et al., 2015; Hu et al., 2018; Isoda et al., 2017; Lin et al., 2012). Compared to loop formation via the active process of cohesin, compartment switch is presumably a slower process as genes in the inactive compartment will need to flip or relocate to a permissive environment before any transcription could happen. This is relatively an unfavourable process for naïve B cells, which require a prompt reaction upon pathogen detection. This swift response is also enabled by the pre-established 3D structures poised for transcription in naïve B cells, which is reflected in the huge transcriptional burst in the absence of conformational changes observed shortly after activation. The pre-existing structures in naïve B cells were found to correlate with the rapid transcriptional changes upon stimulation. This suggests that the chromatin structures or loops pre-existed in naïve B cells are poised and permissive for driving transcription, enabling rapid gene activation once triggered. In describing the enhancer-promoter communication, de Laat and Duboule (2013) have proposed that there are two

separate modes of action, which differs only in the time of execution. Enhancers can be instructive, in which gene transcription is activated immediately upon *de novo* enhancer-promoter contact. Enhancers can also be permissive, where loops are established but remain poised, with transcription delayed until triggered. Fitting this model to the initial transcriptional burst driven by the existing structures, it appears that the enhancers pre-established in naïve B cells work in a permissive manner. On the other hand, during the two reorganisation events the formations of new loops, including *de novo* enhancer-promoter contacts, were revealed to be positively correlated with transcriptional changes. The observation of correlation is consistent with other previous studies (Chaudhri et al., 2020; Javierre et al., 2016; Kieffer-Kwon et al., 2017). This suggests that the predominant function of these *de novo* loops is to drive and augment gene expression, and they therefore appear to function in an instructive manner (Chan et al., 2020b).

Since the first wave of architectural changes are independent of DNA synthesis and mitosis, it is thus conceivable that the genome is reconfigured, at least partially, by the process of transcription or the actions of transcription factors. As already discussed in depth in previous chapters (Section 1.1, 4.2), lncRNAs and the progression of RNA polymerase can mediate 3D restructuring of chromatin in a myriad of ways. On the other hand, transcription factors can themselves remodel chromatin conformation in a transcription-independent manner. A notable example is the loop extrusion process mediated by CTCF and cohesin (Busslinger et al., 2017; Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017). In addition, it has previously demonstrated in our lab that the B cell transcription factor Pax5 rewires genome during lineage commitment independent of active transcription (Johanson et al., 2018b). This illustrates that certain lineage-specific transcription factors are also able to reorganise genome without the need for transcription (Chan et al., 2020b).

Upon clonal expansion, which starts after the first genome restructuring event, it was demonstrated that genome topology stays remarkably similar, in spite of the time and number of generations that each cell has experienced. This suggests the loops and structures that were established prior to the first division are

preserved over many successive cell divisions (Chan et al., 2020b). This appears to be in line with many cellular studies showing that B and T lymphocytes fate decisions for terminal differentiation are made within a day after antigen stimulation (Corbin and Harty, 2004; Iezzi et al., 1998; Kaech and Ahmed, 2001; Mercado et al., 2000; Rush and Hodgkin, 2001; Turner et al., 2008; van Stipdonk et al., 2003; van Stipdonk et al., 2001; Williams and Bevan, 2004; Wong and Pamer, 2001), probably prior to the first cell division, and the stimuli received afterwards could no longer intervene. It is thus speculated that the loops and other 3D structures established prior to the first division would pre-program and dictate the subsequent transcription strength and profile, and clonally expanding cells then become somewhat refractory to external signals thereafter as the genome architecture is already established (Chan et al., 2020b).

The second wave of genome reorganisation event was shown to occur upon terminal differentiation into plasmablast. Although it remains elusive if the restructuring was carried out in G₁ phase, cell cycle analysis showed a lengthening of G₁ phase in activated B cells. Intriguingly, the lengthening is most prominent at the transitional pre-plasmablast stage. Slowing the cell division times, which presumably leads to a prolonged G₁ phase, has been shown to increase the differentiation rate of B cells (Zhou et al., 2018). Recently, it was also demonstrated that the terminal differentiation of another cell type – adipocytes – occurs in a prolonged G₁ phase (Zhao et al., 2020). Through knocking down cyclin D1 to increase the G₁ duration, differentiation into adipocytes was enhanced. On the other hand, while a study of human naïve B cells has revealed the transcriptional switch of plasmablast differentiation occurs in S phase (Caron et al., 2015), it is conceivable that the reconfiguration of genome structures instead occurred in G₁ phase as it needs to precede and hence elicit such a transcriptional switch. Perhaps the utilisation of novel and single-cell techniques such as single-cell RNA-seq (Usoskin et al., 2015) and ATAC-seq (Cusanovich et al., 2018), together with SPRITE (Quinodoz et al., 2018) and GAM (Beagrie et al., 2017) could allow us to pinpoint the answer.

There are other outstanding questions regarding B cells immunity remain to be answered. While LPS stimulation can lead to plasmablast differentiation, stimulus

provided by unmethylated CpG motif does not trigger differentiation due to the failed upregulation of Blimp-1 (Fairfax et al., 2007). It is therefore interesting to determine if this failure of differentiation is attributed to the genome architectural differences between LPS and CpG stimulated B cells.

A large body of research has been focussed on understanding the cellular and molecular events governing B cell activation and terminal differentiation into antibody-secreting cells (Chan et al., 2020b). However, very little is known about how the genome is wired in long-lived memory cells and how this contributes to their rapid recall upon rechallenge, largely due to the lack of good surface markers and their low abundance. Current literature suggests that memory B cells still possess a considerable amount of B cell signature and gene programme. For example, transcriptional profiling has shown a relatively similar gene expression landscape between memory and naïve B cells, with noticeable elevations of some anti-apoptotic genes and costimulatory molecules in memory cells (Bhattacharya et al., 2007). Moreover, during a recall, memory B cells are able to promptly differentiate into plasmablasts and produce antigen-specific antibody, yet surprisingly they are also capable of undergoing a second round of somatic hypermutation and class switch recombination to modify the antibody specificity and isotype (Dogan et al., 2009). Thus, it is conceivable that the genome organisation of memory B cells would be different to terminally differentiated plasmablasts, while arranging similarly to naïve or activated B cells. Nevertheless, how such chromatin conformation facilitates a rapid memory response is entirely unknown. However, it is likely that it would involve some pre-existing loops poised for rapid activation. Again, although hampered by the scarcity and heterogeneity of memory B cells, the aforementioned novel and single-cell techniques can help elucidate the molecular identity of memory cells (Chan et al., 2020b).

6. General Discussion

Genomic structures and the relationship with gene transcription

After many decades of research, the mechanisms underlying how cells with almost identical genomic sequences are able to differentiate into distinct cell types are only beginning to be revealed. Apart from lineage-specific transcription factors and a myriad of epigenetic layers, the three-dimensional structure of genome has recently been uncovered to serve an important role in regulating gene transcription (chapter 1). Indeed, this discovery has provided explanations for how distal enhancers interact with promoters and hence augment transcription, and how insulators block such communication. However, the importance of newly described structures such as TADs and compartments was just started to be explored. Certain studies have linked pathological consequences to TAD disruption (Flavahan et al., 2016; Lupianez et al., 2015), however, expression of many genes was found to persist upon removal of majority of TADs (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017). As such, the link between these higher-order structures and transcription appears to be more complicated than anticipated. While it is well known that promoters alone do possess certain degree of basal activities (Maston et al., 2006), it is thus conceivable to speculate that TADs are probably more important in conferring modulation on top of this basal activity. In other words, TADs are critical in regulating the change. In agreeing with this, the positive correlation between TADs and gene expression was mainly found to associate with differentiation (Despang et al., 2019; Dixon et al., 2015; Nora et al., 2012; Symmons et al., 2014) and upon external stimuli (Le Dily et al., 2014; Stik et al., 2020). Consistently, in the current research, which extends from our previous study (Johanson et al., 2018b), lineage-specific TADs were found to form around developmentally regulated immune loci (chapter 3), and gene expression within the Gata3 TAD was shown to respond to IL-4 signal in a concerted fashion (chapter 4). However, it remains to be elucidated how frequent such lineage-specific TADs exist in a particular lineage and what genes do they generally harbour, as well as whether concerted gene expression is broadly observed in these TADs.

The complexity further lies at the mutual relationship between genome organisation and transcription. As previously discussed, the non-coding RNAs and the process of transcription elongation are able to alter local chromatin landscape, interfere with loop formation and relocate chromatin to a different nuclear compartment, hence affecting gene transcription (chapter 1.1, Fig. 1.4, chapter 4). Additionally, the act of transcription, regardless of the coding potentials, has been suggested to help, plausibly via the phase-separation mechanism, segregate the genome into TAD, or more specifically the compartmental domain (chapter 1.1-1.2). Thus, it appears that transcription can intriguingly feedback and in turn modulate genome conformation. Given that more than 90% of the genome has potential to be transcribed (Consortium et al., 2007; Djebali et al., 2012; Wilusz et al., 2009), the contribution made by transcription is therefore immense and should not be neglected. While there are quite a number of approaches developed, as well those adopted in this research in the study of Gm13218 (chapter 4), to functionally dissect the non-coding RNA transcripts (Kaikkonen and Adelman, 2018), the effects of transcription elongation is much harder to disentangle. The spatiotemporal patterning of non-coding RNAs has complicated this further. Recently, a computational model has been developed to simulate the chromatin topology driven by the loop extrusion process of cohesin and CTCF (Nuebler et al., 2018). It would therefore be interesting and informative if a similar model can be developed to simulate how the elongation process might shape chromatin architecture. In addition, through this perhaps it is also possible to annotate or differentiate between CTCF-dependent loop domain and the CTCF-independent compartmental domain. Given sufficient data, possibly the model could even be extended to determine how chromatin conformation is altered during the course of development, allowing plausible pinpoint of the most critical window of time of functional importance exerted by certain cis-regulatory elements, for example the non-coding RNA Gm13218. While such a model is currently lacking, the nature of Gata3 TAD (chapter 3, 4) remains largely unknown. However, the lack of a stable, persistent loop, as implicated in the absence of corner dots, in Gata3 TAD might suggest that the influence by loop extrusion is not strong, and thus transcription itself could perhaps play a more prominent role. Nonetheless, it would be interesting to investigate how this TAD comes into being during early

development, and how Gm13218 might contribute to this process. Hi-C and transcriptomic data of different thymic T cells provided from Hu et al. (2018) could possibly help elucidate this question.

Requirement for precision genome editing to examine mediators of transcription

The genome structures, as demonstrated, can be disrupted, or manipulated through various CRISPR based approaches (chapter 1.3, Fig. 1.7, chapter 3 & 4). This is made possible by the recent blossoming of a number of sophisticated epigenome-editing methods via dCas9. Contrarily, the deletion of cis-regulatory elements (chapter 3), and introduction of polyadenylation signals (chapter 4), are still associated with undesirable indels and DNA duplication. The imprecise editing is thought to be attributed to the generation of DSBs and the nature of the corresponding DNA repair processes – NHEJ and HDR (Anzalone et al., 2020). To address this several new developments of precise editing have instead exploited other DNA repair pathways such as the mismatch repair (MMR) and base excision repair (BER), with no DSBs generated. For instance, through the attachment of deaminase to the Cas9 nickase – Cas9 variant which cleaves only one strand of DNA (Cong et al., 2013) – the resultant base editing can allow for precise mutation of single nucleotide (Anzalone et al., 2020; Rees and Liu, 2018). Whereas the attachment of reverse transcriptase to the nickase and incorporation of repair template into sgRNAs – referred as prime editing – permits precise small DNA insertion with substantially less indels and duplication (Anzalone et al., 2019). These innovative approaches will presumably enable the studies of single nucleotide polymorphism (SNP) or mutations in a particular transcription factor binding site or loop anchor. Moreover, transposition can also be employed in the CRISPR platform, without generation of any DNA cleavage (Chen and Wang, 2019; Klompe et al., 2019; Strecker et al., 2019). Transposase together with catalytically inactive Cas proteins allow large DNA fragments to be integrated at the desired region. It therefore appears to be an appealing alternative approach as to the insertion of a polyadenylation signal to interfere with transcription elongation. Perhaps with further optimisation and sophistication of this approach a genome wide poly(A) insertion screening could then possibly be conducted, to identify any non-coding transcription processes that are important in regulating 3D genome architecture as well as transcription.

Waves of cell cycle related genome reorganisation during B cell differentiation

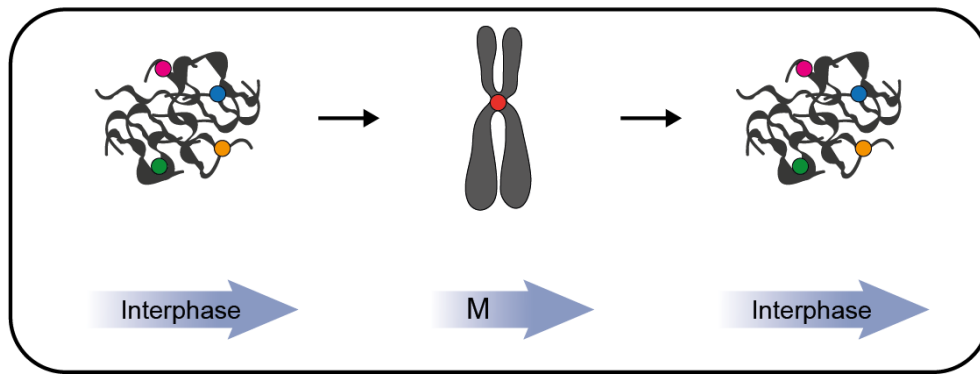
During activation of B lymphocytes, the genome was found to undergo reconfiguration prior to the first DNA synthesis as well as mitosis (chapter 5, Fig. 6.1, 6.2). As such, the reconfiguration occurs in G₁ phase and is therefore partitioned from these two challenging processes. Also, it appears to suggest that a single individual cell has potential to adopt a different chromatin conformation without the need to divide. It is then of interest to know if this is a property of other cell types and differentiation processes. In addition, the pre-mitotic reconfiguration further implicates the mitotic bookmarks – factors essential for epigenetic memory – should be laid down during this prolonged growth period (Fig. 6.1), such that the newly adopted chromatin conformation can be inherited in the subsequent generation. Consistent with this, extensive histone modifications, one of the mitotic bookmarks to be proposed (Oomen and Dekker, 2017), have been documented immediately upon B cell activation (Kieffer-Kwon et al., 2017; Pogo et al., 1966, 1967).

One unique feature regarding B cell activation is that the prompt response does not involve any broad-scale A/B compartment switching (chapter 5). While loop formation including promoter-enhancer contact, could be facilitated by active extrusion process of cohesin (Davidson et al., 2019; Kim et al., 2019b), the A/B compartment switch, as well as the relocation of DNA to a different nuclear compartment, should conceivably take more time. This is seen in the differentiation of stem cells, early T- and B-cell development (Dixon et al., 2015; Hu et al., 2018; Lin et al., 2012). Thus, it still remains to be elucidated if the compartment switch and DNA relocation during differentiation could occur in individual cell, and also if they occur in G₁ phase of cell cycle.

Genome reconfiguration of activated B cells consists predominantly of chromatin loops (chapter 5), however, the mechanism underlying such formation has yet to be ascertained. Nonetheless, both loop extrusion and transcription are presumably involved in the remodelling of 3D chromatin. Histone modifications have been documented in the early activated B cells (Kieffer-Kwon et al., 2017; Pogo et al., 1966, 1967) and such change in the local chromatin landscape could

then interfere with the process of loop extrusion, for example the loading of cohesin and how it extrudes. On the other hand, the massive transcription amplification, mediated by Myc (Nie et al., 2012), upon activation, is probably accompanied by the global emergence of transcription condensates. As such, these phase-separated foci could potentially strengthen existing loops and facilitate long-range chromatin interactions. This is exemplified by the rapid relocation of *IgH* and *Myc* loci to a shared transcription factory immediately upon B cell stimulation (Osborne et al., 2007).

Normal turnover



Upon differentiation

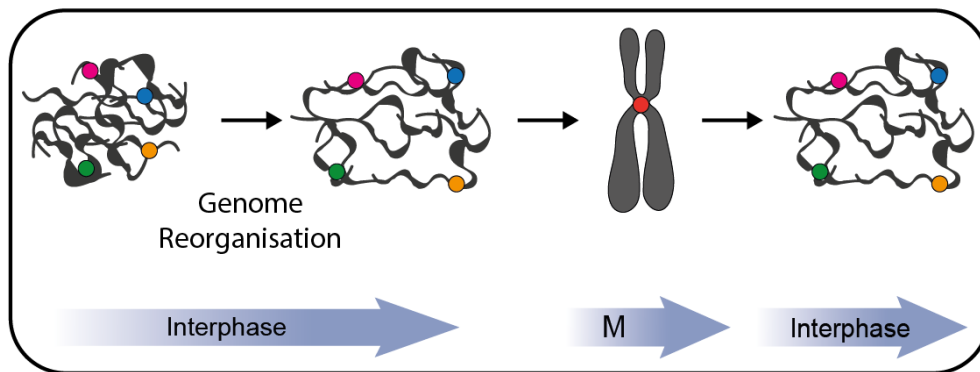


Figure 6.1. During differentiation the genome is reorganised prior to mitosis.

Top) Under normal turnover without any differentiation or stimulating signals, genome in cells becomes highly condensed and hence any 3D structures are then erased. Short after mitotic exit, 3D genome architecture would be faithfully re-established, in the exact same conformation as in the parent cells. Bottom) Having stimulated to differentiate, cells would first undergo extensive genome-wide chromatin reconfiguration, with different mitotic bookmarks being laid down. The new conformation is then inherited through the subsequent cell proliferation.

The second genome reorganisation event during B cell activation occurs upon terminal plasmablast differentiation (chapter 5, Fig. 6.2). This reconfiguration was shown to associate with a prolonged G₁ phase. However, it has not been determined if this event occurs definitively in G₁, and if it is coupled to cell division or not. It has been shown that plasma cell differentiation is accompanied by extensive DNA demethylation (Barwick et al., 2016). It is known that certain chromatin loops and interactions are sensitive to DNA methylation status (Ito et al., 2013) and therefore the demethylation process of DNA can potentially affect how chromatin is actively reorganised. In particular it can also affect how chromatin is reorganised in relation to cell division, because of the nature of the DNA demethylation process. Due to the very strong covalent bond between the cytosine and methyl group, there is no known enzymatic way in mammals to directly cleave such a bond (Moore et al., 2013). Passive dilution of methylated DNA through replication and cell division appears to be one mechanism (Kagiwada et al., 2013; Ohno et al., 2013). However, DNA can also be indirectly demethylated through deamination by the TET enzymes with following faithful DNA repair of the resulting mismatch (Cortellino et al., 2011; Tahiliani et al., 2009). This process requires energy in the repair of DNA lesions, but is independent of cell division. As such, if the passive process is involved in differentiation, it would then be conceivable that the genome reorganisation is also coupled to replication. Given that differentiation rate increases with cell division number, yet differentiation could still occur after only 3-4 rounds of cell division (Hasbold et al., 2004; Hodgkin et al., 1996), this might suggest a mix of active and passive processes are involved. Indeed, several studies have demonstrated that the TET enzymes are required for plasma cell differentiation (Fujii et al., 2020; Schoeler et al., 2019). For instance, TET is important in controlling IRF4 expression (Fujii et al., 2020). However, it is not known if looping is involved, and also in general if DNA demethylation during plasma cell differentiation is associated with chromatin looping. Overall, it still remains to be explored how genome reorganisation during terminal differentiation could be affected by DNA methylation, and how it is related to cell division and cell cycle.

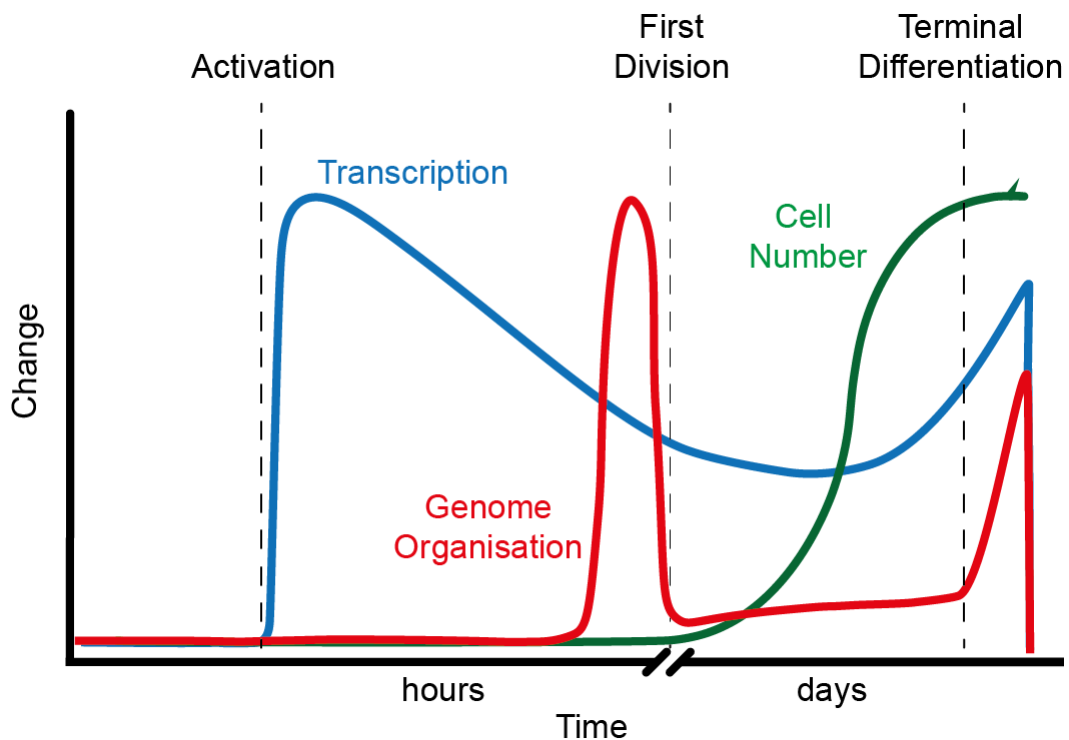


Figure 6.2. Cellular events during B cell clonal expansion and differentiation.

Upon stimulation, massive transcriptional changes prepare B cells for subsequent expansion and differentiation, associating with various metabolic changes. Immediately prior to the first cell division, B cells undergo first wave of reorganisation of genome structures, which remains relatively unaltered, and is therefore inherited, in the subsequent clonal expansion over many days and cell generations. The second wave of genome reconfiguration occurs upon the terminal differentiation into plasmablasts. Blue line denotes transcription, red indicates genome organisation whereas green depicts cell number (Chan et al., 2020b).

Overall conclusions

The derivation of the functionally and phenotypically distinct array of immune cell types is controlled by non-coding DNA, which serves as a regulator that controls the expression of a fraction of a genome at a given time. Some of these non-coding DNAs, or cis-regulatory elements, were demonstrated, in this thesis as well as in the work of others, to achieve gene regulation through three-dimensional genome interactions. Intriguingly, the transcription resulting from these 3D structures can reciprocate and reinforce the existing chromatin structure, either by proteins or non-coding RNA transcripts, or even by the act of transcription itself. On the other hand, during differentiation, the remodelling of chromatin occurs in mid-late G₁ phase of cell cycle in a global, concurrent fashion, implicating it as an important and tightly regulated process. As such, among various cellular processes, chromatin conformation appears to be well partitioned from the other DNA-related processes such as DNA synthesis and mitosis, but form an intricate, complex and reciprocal network with transcription. Therefore, the understanding of such regulatory network could provide immense insights of how cellular diversity arises, as well as the potential defects if this network is compromised. With a plethora of CRISPR-based techniques increasingly becoming more sophisticated and effective, such regulatory network, or perhaps the identity or functions of a cell, can potentially be finely and accurately manipulated.

Reference

- Abarrategui, I., and Krangel, M.S. (2006). Regulation of T cell receptor-alpha gene recombination by transcription. *Nat Immunol* 7, 1109-1115.
- Abarrategui, I., and Krangel, M.S. (2007). Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination. *EMBO J* 26, 4380-4390.
- Afshar, R., Pierce, S., Bolland, D.J., Corcoran, A., and Oltz, E.M. (2006). Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *J Immunol* 176, 2439-2447.
- Aiden, E.L., and Casellas, R. (2015). Somatic Rearrangement in B Cells: It's (Mostly) Nuclear Physics. *Cell* 162, 708-711.
- Albarracin, F., and Fonseca, R. (2011). Plasma cell leukemia. *Blood Rev* 25, 107-112.
- Allepuz-Fuster, P., O'Brien, M.J., Gonzalez-Polo, N., Pereira, B., Dhoondia, Z., Ansari, A., and Calvo, O. (2019). RNA polymerase II plays an active role in the formation of gene loops through the Rpb4 subunit. *Nucleic Acids Res* 47, 8975-8987.
- Alt, F.W., Yancopoulos, G.D., Blackwell, T.K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., and Baltimore, D. (1984). Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J* 3, 1209-1219.
- Anderson, M.K., Hernandez-Hoyos, G., Dionne, C.J., Arias, A.M., Chen, D., and Rothenberg, E.V. (2002). Definition of regulatory network elements for T cell development by perturbation analysis with PU.1 and GATA-3. *Dev Biol* 246, 103-121.
- Ano, S., Morishima, Y., Ishii, Y., Yoh, K., Yageta, Y., Ohtsuka, S., Matsuyama, M., Kawaguchi, M., Takahashi, S., and Hizawa, N. (2013). Transcription factors GATA-3 and RORgammat are important for determining the phenotype of allergic airway inflammation in a murine model of asthma. *J Immunol* 190, 1056-1065.
- Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* 38, 824-844.

Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149-157.

Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Ronnerblad, M., Hrydziuszko, O., Vitezic, M., *et al.* (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* 347, 1010-1014.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* 144, 214-226.

Bao, Z., Jain, S., Jaroenpuntaruk, V., and Zhao, H. (2017). Orthogonal Genetic Regulation in Human Cells Using Chemically Induced CRISPR/Cas9 Activators. *ACS Synth Biol* 6, 686-693.

Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* 34, 135-141.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712.

Barwick, B.G., Scharer, C.D., Bally, A.P.R., and Boss, J.M. (2016). Plasma cell differentiation is coupled to division-dependent DNA hypomethylation and gene regulation. *Nat Immunol* 17, 1216-1225.

Bashkirova, E., and Lomvardas, S. (2019). Olfactory receptor genes make the case for inter-chromosomal interactions. *Curr Opin Genet Dev* 55, 106-113.

Bassing, C.H., Swat, W., and Alt, F.W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell* 109 Suppl, S45-55.

Beagan, J.A., Gilgenast, T.G., Kim, J., Plona, Z., Norton, H.K., Hu, G., Hsu, S.C., Shields, E.J., Lyu, X., Apostolou, E., *et al.* (2016). Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell* 18, 611-624.

Beagan, J.A., and Phillips-Cremins, J.E. (2020). On the existence and functionality of topologically associating domains. *Nat Genet* 52, 8-16.

Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.M., Branco, M.R., *et al.* (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543, 519-524.

Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang, Y., Li, J.B., Senaratne, T.N., Williams, B.R., *et al.* (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc Natl Acad Sci U S A* 109, 21301-21306.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405, 482-485.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 387-396.

Berg, O.G., and von Hippel, P.H. (1987). Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *J Mol Biol* 193, 723-750.

Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 313, 1642-1645.

Bhattacharya, D., Cheah, M.T., Franco, C.B., Hosen, N., Pin, C.L., Sha, W.C., and Weissman, I.L. (2007). Transcriptional profiling of antigen-dependent murine B cell differentiation and memory formation. *J Immunol* 179, 6808-6819.

Bian, Q., Khanna, N., Alvikas, J., and Belmont, A.S. (2013). beta-Globin cis-elements determine differential nuclear targeting through epigenetic modifications. *J Cell Biol* 203, 767-783.

Bird, J.J., Brown, D.R., Mullen, A.C., Moskowitz, N.H., Mahowald, M.A., Sider, J.R., Gajewski, T.F., Wang, C.R., and Reiner, S.L. (1998). Helper T cell differentiation is controlled by the cell cycle. *Immunity* 9, 229-237.

Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. *Science* 281, 60-63.

Blobel, G. (1985). Gene gating: a hypothesis. *Proc Natl Acad Sci U S A* 82, 8527-8529.

Blobel, G. (2010). Three-dimensional organization of chromatids by nuclear envelope-associated structures. *Cold Spring Harb Symp Quant Biol* 75, 545-554.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., *et al.* (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol* 25, 833-840.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., *et al.* (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842-1855 e1816.

Bolland, D.J., Wood, A.L., Afshar, R., Featherstone, K., Oltz, E.M., and Corcoran, A.E. (2007). Antisense intergenic transcription precedes Igh D-to-J recombination and is controlled by the intronic enhancer Emu. *Mol Cell Biol* 27, 5523-5533.

Bolland, D.J., Wood, A.L., Johnston, C.M., Bunting, S.F., Morgan, G., Chakalova, L., Fraser, P.J., and Corcoran, A.E. (2004). Antisense intergenic transcription in V(D)J recombination. *Nat Immunol* 5, 630-637.

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Muller, S., Eils, R., Cremer, C., Speicher, M.R., and Cremer, T. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 3, e157.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. *Nat Rev Genet* 17, 661-678.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A., and Cavalli, G. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171, 557-572 e524.

Bortnick, A., He, Z., Aubrey, M., Chandra, V., Denholtz, M., Chen, K., Lin, Y.C., and Murre, C. (2018). An inter-chromosomal transcription hub activates the unfolded protein response in plasma cells. *bioRxiv*, 295915.

Bortnick, A., He, Z., Aubrey, M., Chandra, V., Denholtz, M., Chen, K., Lin, Y.C., and Murre, C. (2020). Plasma Cell Fate Is Orchestrated by Elaborate Changes in Genome Compartmentalization and Inter-chromosomal Hubs. *Cell Rep* 31, 107470.

Brady, B.L., Steinell, N.C., and Bassing, C.H. (2010). Antigen receptor allelic exclusion: an update and reappraisal. *J Immunol* 185, 3801-3808.

Braikia, F.Z., Oudinet, C., Haddad, D., Oruc, Z., Orlando, D., Dauba, A., Le Bert, M., and Khamlichi, A.A. (2017). Inducible CTCF insulator delays the IgH 3' regulatory region-mediated activation of germline promoters and alters class switching. *Proc Natl Acad Sci U S A* 114, 6092-6097.

Branco, M.R., and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* 4, e138.

Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R., and Nasmyth, K. (1985). Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* 41, 41-48.

Bronshtein, I., Kanter, I., Kepten, E., Lindner, M., Berezin, S., Shav-Tal, Y., and Garini, Y. (2016). Exploring chromatin organization mechanisms through its dynamic properties. *Nucleus* 7, 27-33.

Brown, K.E., Baxter, J., Graf, D., Merkenschlager, M., and Fisher, A.G. (1999). Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol Cell* 3, 207-217.

Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M., and Fisher, A.G. (1997). Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* 91, 845-854.

Buchwalter, A., Kaneshiro, J.M., and Hetzer, M.W. (2019). Coaching from the sidelines: the nuclear periphery in genome regulation. *Nat Rev Genet* 20, 39-50.

Buckley, M.S., and Lis, J.T. (2014). Imaging RNA Polymerase II transcription sites in living cells. *Curr Opin Genet Dev* 25, 126-130.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* 14, 838-853.

Bulyk, M.L. (2003). Computational prediction of transcription-factor binding site locations. *Genome Biol* 5, 201.

Bunting, K.L., Soong, T.D., Singh, R., Jiang, Y., Beguelin, W., Poloway, D.W., Swed, B.L., Hatzi, K., Reisacher, W., Teater, M., *et al.* (2016). Multi-tiered Reorganization of the Genome during B Cell Affinity Maturation Anchored by a Germinal Center-Specific Locus Control Region. *Immunity* 45, 497-512.

Burren, O.S., Rubio Garcia, A., Javierre, B.M., Rainbow, D.B., Cairns, J., Cooper, N.J., Lambourne, J.J., Schofield, E., Castro Dopico, X., Ferreira, R.C., *et al.* (2017). Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol* 18, 165.

Busslinger, G.A., Stocsits, R.R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., and Peters, J.M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* 544, 503-507.

Cai, Y., Zhang, Y., Loh, Y.P., Tng, J.Q., Lim, M.C., Cao, Z., Raju, A., Li, S., Manikandan, L., and Tergaonkar, V. (2019). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *bioRxiv*, 684712.

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* 131, 706-717.

Cannavo, E., Khoueiry, P., Garfield, D.A., Geeleher, P., Zichner, T., Gustafson, E.H., Ciglar, L., Korbel, J.O., and Furlong, E.E. (2016). Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol* 26, 38-51.

Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T., Maeda, T., Paw, B.H., and Orkin, S.H. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J Biol Chem* 289, 21312-21324.

Cardozo Gizzi, A.M., Cattoni, D.I., Fiche, J.B., Espinola, S.M., Gurgo, J., Messina, O., Houbbron, C., Ogiyama, Y., Papadopoulos, G.L., Cavalli, G., *et al.* (2019). Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol Cell* 74, 212-222 e215.

Carlson-Stevermer, J., Abdeen, A.A., Kohlenberg, L., Goedland, M., Molugu, K., Lou, M., and Saha, K. (2017). Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat Commun* 8, 1711.

Caron, G., Hussein, M., Kulis, M., Delaloy, C., Chatonnet, F., Pignarre, A., Avner, S., Lemarie, M., Mahe, E.A., Verdaguer-Dot, N., *et al.* (2015). Cell-Cycle-

Dependent Reconfiguration of the DNA Methylome during Terminal Differentiation of Human B Cells into Plasma Cells. *Cell Rep* 13, 1059-1071.

Chan, W.F., Coughlan, H.D., Iannarella, N., Smyth, G.K., Johanson, T.M., Keenan, C.R., and Allan, R.S. (2020a). Identification and characterization of the long noncoding RNA Dreg1 as a novel regulator of Gata3. *Immunol Cell Biol*.

Chan, W.F., Coughlan, H.D., Zhou, J.H., Keenan, C.R., Bediaga, N.G., Hodgkin, P.D., Smyth, G.K., Johanson, T.M., and Allan, R.S. (2019). Genome reconfiguration prior to mitosis shapes the generation of adaptive immunity. *bioRxiv*, 762757.

Chan, W.F., Johanson, T.M., and Allan, R.S. (2020b). Three-dimensional genome rewiring during the development of antibody-secreting cells. *Biochem Soc Trans* 48, 1109-1119.

Chatterjee, S., Ju, Z., Hassan, R., Volpi, S.A., Emelyanov, A.V., and Birshtein, B.K. (2011). Dynamic changes in binding of immunoglobulin heavy chain 3' regulatory region to protein factors during class switching. *J Biol Chem* 286, 29303-29312.

Chaudhri, V.K., Dienger-Stambaugh, K., Wu, Z., Shrestha, M., and Singh, H. (2020). Charting the cis-regulome of activated B cells by coupling structural and functional genomics. *Nat Immunol* 21, 210-220.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., E, P.R.I., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., *et al.* (2015). Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* 12, 326-328.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., *et al.* (2016). Comparison of Cas9 activators in multiple species. *Nat Methods* 13, 563-567.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479-1491.

Chen, D., and Zhang, G. (2001). Enforced expression of the GATA-3 transcription factor affects cell fate decisions in hematopoiesis. *Exp Hematol* 29, 971-980.

Chen, S.P., and Wang, H.H. (2019). An Engineered Cas-Transposon System for Programmable and Site-Directed DNA Transpositions. *CRISPR J* 2, 376-394.

Chen, T., Gao, D., Zhang, R., Zeng, G., Yan, H., Lim, E., and Liang, F.S. (2017a). Chemically Controlled Epigenome Editing through an Inducible dCas9 System. *J Am Chem Soc* 139, 11337-11340.

Chen, X., Yammine, S., Shi, C., Tark-Dame, M., Gondor, A., and Ohlsson, R. (2014a). The visualization of large organized chromatin domains enriched in the H3K9me2 mark within a single chromosome in a single cell. *Epigenetics* 9, 1439-1445.

Chen, Y., Lun, A.T., and Smyth, G.K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* 5.

Chen, Y., Lun, A.T.L., and Smyth, G.K. (2014b). Differential Expression Analysis of Complex RNA-seq Experiments Using edgeR. In *Statistical Analysis of Next Generation Sequencing Data*, S. Datta, and D. Nettleton, eds. (Cham: Springer International Publishing), pp. 51-74.

Chen, Y., Pal, B., Visvader, J.E., and Smyth, G.K. (2017b). Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. *F1000Res* 6, 2055.

Cheng, A.W., Jillette, N., Lee, P., Plaskon, D., Fujiwara, Y., Wang, W., Taghbalout, A., and Wang, H. (2016). Casilio: a versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labeling. *Cell Res* 26, 254-257.

Chepelev, I., Wei, G., Wangsa, D., Tang, Q., and Zhao, K. (2012). Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res* 22, 490-503.

Cho, W.K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.H., Lavis, L.D., Lionnet, T., and Cisse, II (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. *Elife* 5.

Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, II (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412-415.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., and Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361.

Chowdhury, D., and Sen, R. (2001). Stepwise activation of the immunoglobulin mu heavy chain gene locus. *EMBO J* 20, 6394-6403.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823.

Consortium, C.e.S. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012-2018.

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

Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., *et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799-816.

Corbin, G.A., and Harty, J.T. (2004). Duration of infection and antigen display have minimal influence on the kinetics of the CD4+ T cell response to *Listeria monocytogenes* infection. *J Immunol* 173, 5679-5687.

Corces, V. (2019). <https://twitter.com/CorcesVictor/status/1120698678535180289> (23 April 2019).
Twitter

Corcoran, A.E. (2010). The epigenetic role of non-coding RNA transcription and nuclear organization in immunoglobulin repertoire generation. *Semin Immunol* 22, 353-361.

Corcoran, A.E., Riddell, A., Krooshoop, D., and Venkitaraman, A.R. (1998). Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor. *Nature* 391, 904-907.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., *et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146, 67-79.

Costa, F.F. (2010). Non-coding RNAs: Meet thy masters. *Bioessays* 32, 599-608.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2, 292-301.

Cusanovich, D.A., Hill, A.J., Aghamirzaie, D., Daza, R.M., Pliner, H.A., Berletch, J.B., Filippova, G.N., Huang, X., Christiansen, L., DeWitt, W.S., *et al.* (2018). A

Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* 174, 1309-1324 e1318.

Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.M. (2019). DNA loop extrusion by human cohesin. *Science* 366, 1338-1345.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499-506.

de Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quezada, C., Verstegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* 60, 676-684.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev* 25, 1010-1022.

Dechat, T., Adam, S.A., Taimen, P., Shimi, T., and Goldman, R.D. (2010). Nuclear lamins. *Cold Spring Harb Perspect Biol* 2, a000547.

Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., *et al.* (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc Natl Acad Sci U S A* 108, 9566-9571.

Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. *Cell* 164, 1110-1121.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306-1311.

Dermitzakis, E.T., and Clark, A.G. (2002). Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. *Mol Biol Evol* 19, 1114-1121.

Despang, A., Schopflin, R., Franke, M., Ali, S., Jerkovic, I., Paliou, C., Chan, W.L., Timmermann, B., Wittler, L., Vingron, M., *et al.* (2019). Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet* 51, 1263-1271.

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., *et al.* (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* 14, 629-635.

Dillon, N., and Festenstein, R. (2002). Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet* 18, 252-258.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., *et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331-336.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376-380.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., *et al.* (2012). Landscape of transcription in human cells. *Nature* 489, 101-108.

Dogan, I., Bertocci, B., Vilmont, V., Delbos, F., Megret, J., Storck, S., Reynaud, C.A., and Weill, J.C. (2009). Multiple layers of B cell memory with different effector functions. *Nat Immunol* 10, 1292-1299.

Donda, A., Schulz, M., Burki, K., De Libero, G., and Uematsu, Y. (1996). Identification and characterization of a human CD4 silencer. *Eur J Immunol* 26, 493-500.

Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. *Nat Commun* 11, 1061.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., *et al.* (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16, 1299-1309.

Dunipace, L., Ozdemir, A., and Stathopoulos, A. (2011). Complex interactions between cis-regulatory modules in native conformation are critical for *Drosophila* snail expression. *Development* 138, 4075-4084.

Dunipace, L., Saunders, A., Ashe, H.L., and Stathopoulos, A. (2013). Autoregulatory feedback controls sequential action of cis-regulatory modules at the brinker locus. *Dev Cell* 26, 536-543.

Ebert, A., McManus, S., Tagoh, H., Medvedovic, J., Salvagiotto, G., Novatchkova, M., Tamir, I., Sommer, A., Jaritz, M., and Busslinger, M. (2011). The distal V(H) gene cluster of the IgH locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. *Immunity* 34, 175-187.

Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. *Nat Cell Biol* 10, 1106-1113.

Eden, S., and Cedar, H. (1994). Role of DNA methylation in the regulation of transcription. *Curr Opin Genet Dev* 4, 255-259.

Ellmeier, W., Sunshine, M.J., Losos, K., Hatam, F., and Littman, D.R. (1997). An enhancer that directs lineage-specific expression of CD8 in positively selected thymocytes and mature T cells. *Immunity* 7, 537-547.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539, 452-455.

Ernst, J. (2012). Mapping enhancer and promoter interactions. *Cell Res* 22, 789-790.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., and Church, G.M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* 10, 1116-1121.

Fairfax, K.A., Corcoran, L.M., Pridans, C., Huntington, N.D., Kallies, A., Nutt, S.L., and Tarlinton, D.M. (2007). Different kinetics of blimp-1 induction in B cell subsets revealed by reporter gene. *J Immunol* 178, 4104-4111.

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B.R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I., and Mirny, L.A. (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* 570, 395-399.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 15, 7-21.

Feldman, S., Achour, I., Wuerffel, R., Kumar, S., Gerasimova, T., Sen, R., and Kenter, A.L. (2015). Constraints contributed by chromatin looping limit recombination targeting during Ig class switch recombination. *J Immunol* 194, 2380-2389.

Feldman, S., Wuerffel, R., Achour, I., Wang, L., Carpenter, P.B., and Kenter, A.L. (2017). 53BP1 Contributes to Igh Locus Chromatin Topology during Class Switch Recombination. *J Immunol* 198, 2434-2444.

Flavahan, W.A., Drier, Y., Liao, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110-114.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* 15, 2038-2049.

Fudenberg, G., and Pollard, K.S. (2019). Chromatin features constrain structural variation across evolutionary timescales. *Proc Natl Acad Sci U S A* 116, 2175-2180.

Fujii, K., Tanaka, S., Hasegawa, T., Narazaki, M., Kumanogoh, A., Koseki, H., Kurosaki, T., and Ise, W. (2020). Tet DNA demethylase is required for plasma cell differentiation by controlling expression levels of IRF4. *Int Immunol*.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 354, 769-773.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science* 361, 1341-1345.

Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E., and Busslinger, M. (2004). Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* 18, 411-422.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31, 397-405.

Gall, J.G. (1981). Chromosome structure and the C-value paradox. *J Cell Biol* 91, 3s-14s.

Gao, Y., Xiong, X., Wong, S., Charles, E.J., Lim, W.A., and Qi, L.S. (2016). Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat Methods* 13, 1043-1049.

Garcia-Ojeda, M.E., Klein Wolterink, R.G., Lemaitre, F., Richard-Le Goff, O., Hasan, M., Hendriks, R.W., Cumano, A., and Di Santo, J.P. (2013). GATA-3 promotes T-cell specification by repressing B-cell potential in pro-T cells in mice. *Blood* 121, 1749-1759.

Garn, H., and Renz, H. (2017). GATA-3-specific DNAzyme - A novel approach for stratified asthma therapy. *Eur J Immunol* 47, 22-30.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The

CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67-71.

Garnier, S. (2018). viridis: Default Color Maps from 'matplotlib'.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109, E2579-2586.

Gass, J.N., Gifford, N.M., and Brewer, J.W. (2002). Activation of an unfolded protein response during differentiation of antibody-secreting B cells. *J Biol Chem* 277, 49047-49054.

Georgescu, C., Longabaugh, W.J., Scripture-Adams, D.D., David-Fung, E.S., Yui, M.A., Zarnegar, M.A., Bolouri, H., and Rothenberg, E.V. (2008). A gene regulatory network armature for T lymphocyte specification. *Proc Natl Acad Sci U S A* 105, 20100-20105.

Gibcus, J.H., Samejima, K., Goloborodko, A., Samejima, I., Naumova, N., Nuebler, J., Kanemaki, M.T., Xie, L., Paulson, J.R., Earnshaw, W.C., *et al.* (2018). A pathway for mitotic chromosome formation. *Science* 359.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343-345.

Gibson, G. (1996). Epistasis and pleiotropy as natural properties of transcriptional regulation. *Theor Popul Biol* 49, 58-89.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., *et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442-451.

Gleave, M.E., and Monia, B.P. (2005). Antisense therapy for cancer. *Nat Rev Cancer* 5, 468-479.

Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., and Bergman, Y. (2005). Epigenetic ontogeny of the Igk locus during B cell development. *Nat Immunol* 6, 198-203.

Gollosi, R., Sanders, J.T., and McCord, R.P. (2017). Genome organization during the cell cycle: unity in division. *Wiley Interdiscip Rev Syst Biol Med* 9.

Gregory, T.R. (2005). Synergy between sequence and size in large-scale genomics. *Nat Rev Genet* 6, 699-708.

Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. *Nat Rev Genet* 8, 35-46.

Grogan, J.L., Mohrs, M., Harmon, B., Lacy, D.A., Sedat, J.W., and Locksley, R.M. (2001). Early transcription and silencing of cytokine genes underlie polarization of T helper cell subsets. *Immunity* 14, 205-215.

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., Oltz, E.M., and Sen, R. (2011a). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* 147, 332-343.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., *et al.* (2011b). CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424-430.

Guo, M., Price, M.J., Patterson, D.G., Barwick, B.G., Haines, R.R., Kania, A.K., Bradley, J.E., Randall, T.D., Boss, J.M., and Scharer, C.D. (2018). EZH2 Represses the B Cell Transcriptional Program and Regulates Antibody-Secreting Cell Metabolism and Antibody Production. *J Immunol* 200, 1039-1052.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., *et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162, 900-910.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482, 339-346.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 410-413.

Hahn, M.W., and Wray, G.A. (2002). The g-value paradox. *Evol Dev* 4, 73-75.

Haines, R.R., Barwick, B.G., Scharer, C.D., Majumder, P., Randall, T.D., and Boss, J.M. (2018). The Histone Demethylase LSD1 Regulates B Cell Proliferation and Plasmablast Differentiation. *J Immunol* 201, 2799-2811.

Hakim, O., Sung, M.H., Nakayamada, S., Voss, T.C., Baek, S., and Hager, G.L. (2013). Spatial congregation of STAT binding directs selective nuclear architecture during T-cell functional differentiation. *Genome Res* 23, 462-472.

Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* 6.

Hao, N., Shearwin, K.E., and Dodd, I.B. (2017). Programmable DNA looping using engineered bivalent dCas9 complexes. *Nat Commun* 8, 1628.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405, 486-489.

Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merckenschlager, M., and Lenhard, B. (2017). Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat Commun* 8, 441.

Harr, J.C., Luperchio, T.R., Wong, X., Cohen, E., Wheelan, S.J., and Reddy, K.L. (2015). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J Cell Biol* 208, 33-52.

Hasbold, J., Corcoran, L.M., Tarlinton, D.M., Tangye, S.G., and Hodgkin, P.D. (2004). Evidence from the generation of immunoglobulin G-secreting cells that stochastic mechanisms regulate lymphocyte differentiation. *Nat Immunol* 5, 55-63.

Hasegawa, S.L., Moriguchi, T., Rao, A., Kuroha, T., Engel, J.D., and Lim, K.C. (2007). Dosage-dependent rescue of definitive nephrogenesis by a distant Gata3 enhancer. *Dev Biol* 301, 568-577.

Hawkins, E.D., Markham, J.F., McGuinness, L.P., and Hodgkin, P.D. (2009). A single-cell pedigree analysis of alternative stochastic lymphocyte fates. *Proc Natl Acad Sci U S A* 106, 13457-13462.

He, B., Chen, C., Teng, L., and Tan, K. (2014). Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci U S A* 111, E2191-2199.

Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., *et al.* (2018). Transcription Elongation Can Affect Genome 3D Structure. *Cell* 174, 1522-1536 e1522.

Helmink, B.A., and Sleckman, B.P. (2012). The response to and repair of RAG-mediated DNA double-strand breaks. *Annu Rev Immunol* 30, 175-202.

Heng, T.S., Painter, M.W., and Immunological Genome Project, C. (2008). The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* 9, 1091-1094.

Hess, S.T., Girirajan, T.P., and Mason, M.D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys J* 91, 4258-4272.

Hesslein, D.G., Pflugh, D.L., Chowdhury, D., Bothwell, A.L., Sen, R., and Schatz, D.G. (2003). Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. *Genes Dev* 17, 37-42.

Hewitt, S.L., Farmer, D., Marszalek, K., Cadera, E., Liang, H.E., Xu, Y., Schlissel, M.S., and Skok, J.A. (2008). Association between the Igk and Igh immunoglobulin loci mediated by the 3' Igk enhancer induces 'decontraction' of the Igh locus in pre-B cells. *Nat Immunol* 9, 396-404.

Hewitt, S.L., High, F.A., Reiner, S.L., Fisher, A.G., and Merkenschlager, M. (2004). Nuclear repositioning marks the selective exclusion of lineage-inappropriate transcription factor loci during T helper cell differentiation. *Eur J Immunol* 34, 3604-3613.

Hewitt, S.L., Yin, B., Ji, Y., Chaumeil, J., Marszalek, K., Tenthorey, J., Salvagiotto, G., Steinell, N., Ramsey, L.B., Ghysdael, J., *et al.* (2009). RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nat Immunol* 10, 655-664.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 33, 510-517.

Hnisz, D., Day, D.S., and Young, R.A. (2016). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* 167, 1188-1200.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. *Cell* 169, 13-23.

Ho, Y., Elefant, F., Liebhaber, S.A., and Cooke, N.E. (2006). Locus control region transcription plays an active role in long-range gene activation. *Mol Cell* 23, 365-375.

Hodgkin, P.D., Lee, J.H., and Lyons, A.B. (1996). B cell differentiation and isotype switching is related to division cycle number. *J Exp Med* 184, 277-281.

Hon, C.C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J., *et al.* (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* 543, 199-204.

Hong, J.W., Hendrix, D.A., and Levine, M.S. (2008). Shadow enhancers as a source of evolutionary novelty. *Science* 321, 1314.

Hosoya, T., Kuroha, T., Moriguchi, T., Cummings, D., Maillard, I., Lim, K.C., and Engel, J.D. (2009). GATA-3 is required for early T lineage progenitor development. *J Exp Med* 206, 2987-3000.

Hosoya-Ohmura, S., Lin, Y.H., Herrmann, M., Kuroha, T., Rao, A., Moriguchi, T., Lim, K.C., Hosoya, T., and Engel, J.D. (2011). An NK and T cell enhancer lies 280 kilobase pairs 3' to the gata3 structural gene. *Mol Cell Biol* 31, 1894-1904.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., *et al.* (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496, 498-503.

Hozumi, N., and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc Natl Acad Sci U S A* 73, 3628-3632.

Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.S., Chen, S., *et al.* (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci U S A* 111, 7319-7324.

Hsieh, T.H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* 162, 108-119.

Hsieh, T.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell*.

Hsieh, T.S., Fudenberg, G., Goloborodko, A., and Rando, O.J. (2016). Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat Methods* 13, 1009-1011.

Hu, G., Cui, K., Fang, D., Hirose, S., Wang, X., Wangsa, D., Jin, W., Ried, T., Liu, P., Zhu, J., *et al.* (2018). Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity* 48, 227-242 e228.

Hu, T., Fu, Q., Chen, P., Zhang, K., and Guo, D. (2009). Generation of a stable mammalian cell line for simultaneous expression of multiple genes by using 2A peptide-based lentiviral vector. *Biotechnol Lett* 31, 353-359.

Hubisz, M.J., Pollard, K.S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform* 12, 41-51.

Hughes, A.L., Jin, Y., Rando, O.J., and Struhl, K. (2012). A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol Cell* 48, 5-15.

Ibarra, A., Benner, C., Tyagi, S., Cool, J., and Hetzer, M.W. (2016). Nucleoporin-mediated regulation of cell identity genes. *Genes Dev* 30, 2253-2258.

Ibrahim, D.M., and Mundlos, S. (2020). The role of 3D chromatin domains in gene regulation: a multi-faceted view on genome organization. *Curr Opin Genet Dev* 61, 1-8.

Ibrahimi, A., Vande Velde, G., Reumers, V., Toelen, J., Thiry, I., Vandeputte, C., Vets, S., Deroose, C., Bormans, G., Baekelandt, V., *et al.* (2009). Highly efficient multicistronic lentiviral vectors with peptide 2A sequences. *Hum Gene Ther* 20, 845-860.

Iezzi, G., Karjalainen, K., and Lanzavecchia, A. (1998). The duration of antigenic stimulation determines the fate of naive and effector T cells. *Immunity* 8, 89-95.

Ikawa, T., Hirose, S., Masuda, K., Kakugawa, K., Satoh, R., Shibano-Satoh, A., Kominami, R., Katsura, Y., and Kawamoto, H. (2010). An essential developmental checkpoint for production of the T cell lineage. *Science* 329, 93-96.

Isoda, T., Moore, A.J., He, Z., Chandra, V., Aida, M., Denholtz, M., Piet van Hamburg, J., Fisch, K.M., Chang, A.N., Fahl, S.P., *et al.* (2017). Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. *Cell* 171, 103-119 e118.

Ito, Y., Nativio, R., and Murrell, A. (2013). Induced DNA demethylation can reshape chromatin topology at the IGF2-H19 locus. *Nucleic Acids Res* 41, 5290-5302.

Jaehning, J.A., Stewart, C.C., and Roeder, R.G. (1975). DNA-dependent RNA polymerase levels during the response of human peripheral lymphocytes to phytohemagglutinin. *Cell* 4, 51-57.

Jain, S., Ba, Z., Zhang, Y., Dai, H.Q., and Alt, F.W. (2018). CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. *Cell* 174, 102-116 e114.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Varnai, C., Thiecke, M.J., *et al.* (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369-1384 e1319.

Jhunjhunwala, S., van Zelm, M.C., Peak, M.M., Cutchin, S., Riblet, R., van Dongen, J.J., Grosveld, F.G., Knoch, T.A., and Murre, C. (2008). The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* 133, 265-279.

Jhunjhunwala, S., van Zelm, M.C., Peak, M.M., and Murre, C. (2009). Chromatin architecture and the generation of antigen receptor diversity. *Cell* 138, 435-448.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., and Schatz, D.G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* 141, 419-431.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821.

Johanson, T.M., Coughlan, H.D., Lun, A.T.L., Bediaga, N.G., Naselli, G., Garnham, A.L., Harrison, L.C., Smyth, G.K., and Allan, R.S. (2018a). Genome-wide analysis reveals no evidence of trans chromosomal regulation of mammalian immune development. *PLoS Genet* 14, e1007431.

Johanson, T.M., Lun, A.T.L., Coughlan, H.D., Tan, T., Smyth, G.K., Nutt, S.L., and Allan, R.S. (2018b). Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. *Nat Immunol* 19, 1257-1264.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 19, 187-191.

Joung, J., Engreitz, J.M., Konermann, S., Abudayyeh, O.O., Verdine, V.K., Aguet, F., Gootenberg, J.S., Sanjana, N.E., Wright, J.B., Fulco, C.P., *et al.* (2017). Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* 548, 343-346.

Kaeck, S.M., and Ahmed, R. (2001). Memory CD8⁺ T cell differentiation: initial antigen encounter triggers a developmental program in naive cells. *Nat Immunol* 2, 415-422.

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J* 32, 340-353.

Kaikkonen, M.U., and Adelman, K. (2018). Emerging Roles of Non-Coding RNA Transcription. *Trends Biochem Sci* 43, 654-667.

Kallies, A., Hasbold, J., Fairfax, K., Pridans, C., Emslie, D., McKenzie, B.S., Lew, A.M., Corcoran, L.M., Hodgkin, P.D., Tarlinton, D.M., and Nutt, S.L. (2007). Initiation of plasma-cell differentiation is independent of the transcription factor Blimp-1. *Immunity* 26, 555-566.

Kazachenka, A., Bertozzi, T.M., Sjöberg-Herrera, M.K., Walker, N., Gardner, J., Gunning, R., Pahita, E., Adams, S., Adams, D., and Ferguson-Smith, A.C. (2018). Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell* 175, 1259-1271 e1213.

Kenter, A.L., Feldman, S., Wuerffel, R., Achour, I., Wang, L., and Kumar, S. (2012). Three-dimensional architecture of the IgH locus facilitates class switch recombination. *Ann N Y Acad Sci* 1267, 86-94.

Kieffer-Kwon, K.R., Nimura, K., Rao, S.S.P., Xu, J., Jung, S., Pekowska, A., Dose, M., Stevens, E., Mathe, E., Dong, P., *et al.* (2017). Myc Regulates Chromatin Decompaction and Nuclear Architecture during B Cell Activation. *Mol Cell* 67, 566-578 e510.

Kieffer-Kwon, K.R., Tang, Z., Mathe, E., Qian, J., Sung, M.H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., *et al.* (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 155, 1507-1520.

Kim, J.H., Rege, M., Valeri, J., Dunagin, M.C., Metzger, A., Titus, K.R., Gilgenast, T.G., Gong, W., Beagan, J.A., Raj, A., and Phillips-Cremins, J.E. (2019a). LADL: light-activated dynamic looping for endogenous gene expression control. *Nat Methods* 16, 633-639.

Kim, L.K., Esplugues, E., Zorca, C.E., Parisi, F., Kluger, Y., Kim, T.H., Galjart, N.J., and Flavell, R.A. (2014). Oct-1 regulates IL-17 expression by directing interchromosomal associations in conjunction with CTCF in T cells. *Mol Cell* 54, 56-66.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182-187.

Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J., and Yu, H. (2019b). Human cohesin compacts DNA by loop extrusion. *Science* 366, 1345-1349.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. *Cell* 153, 178-192.

Kishi, J.Y., Lapan, S.W., Beliveau, B.J., West, E.R., Zhu, A., Sasaki, H.M., Saka, S.K., Wang, Y., Cepko, C.L., and Yin, P. (2019). SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat Methods* 16, 533-544.

KleinJan, A., Klein Wolterink, R.G., Levani, Y., de Bruijn, M.J., Hoogsteden, H.C., van Nimwegen, M., and Hendriks, R.W. (2014). Enforced expression of Gata3 in T cells and group 2 innate lymphoid cells increases susceptibility to allergic airway inflammation in mice. *J Immunol* 192, 1385-1394.

Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S., and Sternberg, S.H. (2019). Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* 571, 219-225.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., *et al.* (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* 18, 956-963.

Konermann, S., Brigham, M.D., Trevino, A., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500, 472-476.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., *et al.* (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583-588.

Kong, S., Bohl, D., Li, C., and Tuan, D. (1997). Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol Cell Biol* 17, 3955-3965.

Kornienko, A.E., Guenzl, P.M., Barlow, D.P., and Pauler, F.M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biol* 11, 59.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., and Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296, 158-162.

Kotelnikova, E.A., Makeev, V.J., and Gelfand, M.S. (2005). Evolution of transcription factor DNA binding sites. *Gene* 347, 255-263.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.

Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K.R., Bandle, R., Nelson, S., Nakahashi, H., Awasthi, P., Feigenbaum, L., *et al.* (2013). Global regulation of promoter melting in naive lymphocytes. *Cell* 153, 988-999.

Krefting, J., Andrade-Navarro, M.A., and Ibn-Salem, J. (2018). Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* 16, 87.

Krietenstein, N., Abraham, S., Venev, S.V., Abdennur, N., Gibcus, J., Hsieh, T.S., Parsi, K.M., Yang, L., Maehr, R., Mirny, L.A., *et al.* (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell*.

Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* 47, W171-W174.

Lai, B., Tang, Q., Jin, W., Hu, G., Wangsa, D., Cui, K., Stanton, B.Z., Ren, G., Ding, Y., Zhao, M., *et al.* (2018). Trac-looping measures genome structure and chromatin accessibility. *Nat Methods* 15, 741-747.

Lai, C.H., Chou, C.Y., Ch'ang, L.Y., Liu, C.S., and Lin, W. (2000). Identification of novel human genes evolutionarily conserved in *Caenorhabditis elegans* by comparative proteomics. *Genome Res* 10, 703-713.

Lai, F., Gardini, A., Zhang, A., and Shiekhhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399-403.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494, 497-501.

Laine, J.P., Singh, B.N., Krishnamurthy, S., and Hampsey, M. (2009). A physiological role for gene loops in yeast. *Genes Dev* 23, 2604-2609.

Lakshmanan, G., Lieu, K.H., Lim, K.C., Gu, Y., Grosveld, F., Engel, J.D., and Karis, A. (1999). Localization of distant urogenital system-, central nervous system-, and endocardium-specific transcriptional regulatory elements in the GATA-3 locus. *Mol Cell Biol* 19, 1558-1568.

Lam, D.D., de Souza, F.S., Nasif, S., Yamashita, M., Lopez-Leal, R., Otero-Corchon, V., Meece, K., Sampath, H., Mercer, A.J., Wardlaw, S.L., *et al.* (2015). Partially redundant enhancers cooperatively maintain Mammalian pomc expression above a critical functional threshold. *PLoS Genet* 11, e1004935.

Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., *et al.* (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* 498, 511-515.

Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 39, 170-182.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* 547, 236-240.

Latos, P.A., Pauler, F.M., Koerner, M.V., Senergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., *et al.* (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338, 1469-1472.

Le Dily, F., Bau, D., Pohl, A., Vicent, G.P., Serra, F., Soronellas, D., Castellano, G., Wright, R.H., Ballare, C., Fillion, G., *et al.* (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev* 28, 2151-2162.

Lefevre, P., Witham, J., Lacroix, C.E., Cockerill, P.N., and Bonifer, C. (2008). The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. *Mol Cell* 32, 129-139.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**, 573-578.

Lennon, G.G., and Perry, R.P. (1985). C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5'-nontranslatable exon. *Nature* **318**, 475-478.

Lesne, A., Baudement, M.O., Rebouissou, C., and Forne, T. (2019). Exploring Mammalian Genome within Phase-Separated Nuclear Bodies: Experimental Methods and Implications for Gene Expression. *Genes (Basel)* **10**.

Levskaya, A., Weiner, O.D., Lim, W.A., and Voigt, C.A. (2009). Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461**, 997-1001.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., *et al.* (2012a). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98.

Li, L., Zhang, J.A., Dose, M., Kueh, H.Y., Mosadeghi, R., Gounari, F., and Rothenberg, E.V. (2013a). A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. *Blood* **122**, 902-911.

Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., *et al.* (2012b). Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336-340.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., *et al.* (2013b). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516-520.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* **17**, 207-223.

Li, Y., Haarhuis, J.H.I., Sedenio Cacciatore, A., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D., and Panne, D. (2020). The structural basis for cohesin-CTCF-anchored loops. *Nature* **578**, 472-476.

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., *et al.* (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* 208, 44-53.

Liang, X.H., Sun, H., Nichols, J.G., and Crooke, S.T. (2017). RNase H1-Dependent Antisense Oligonucleotides Are Robustly Active in Directing RNA Cleavage in Both the Cytoplasm and the Nucleus. *Mol Ther* 25, 2075-2092.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L., and Ward, D.C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet* 80, 224-234.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289-293.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat Immunol* 13, 1196-1204.

Ling, J., Ainol, L., Zhang, L., Yu, X., Pi, W., and Tuan, D. (2004). HS2 enhancer function is blocked by a transcriptional terminator inserted between the enhancer and the promoter. *J Biol Chem* 279, 51704-51713.

Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., *et al.* (2017). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 355.

Lun, A.T., Perry, M., and Ing-Simmons, E. (2016). Infrastructure for genomic interactions: Bioconductor classes for Hi-C, ChIA-PET and related experiments. *F1000Res* 5, 950.

Lun, A.T., and Smyth, G.K. (2015). diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinformatics* 16, 258.

Lun, A.T., and Smyth, G.K. (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* 44, e45.

Lund, S.P., Nettleton, D., McCarthy, D.J., and Smyth, G.K. (2012). Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat Appl Genet Mol Biol* 11.

Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., *et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012-1025.

Ma, H., Tu, L.C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., and Pederson, T. (2016). Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat Biotechnol* 34, 528-530.

Magistri, M., Faghihi, M.A., St Laurent, G., 3rd, and Wahlestedt, C. (2012). Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends Genet* 28, 389-396.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., *et al.* (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* 18, 67-83.

Manis, J.P., Morales, J.C., Xia, Z., Kutok, J.L., Alt, F.W., and Carpenter, P.B. (2004). 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. *Nat Immunol* 5, 481-487.

Manso, R., Bellas, C., Martin-Acosta, P., Mollejo, M., Menarguez, J., Rojo, F., Llamas, P., Piris, M.A., and Rodriguez-Pinilla, S.M. (2016). C-MYC is related to GATA3 expression and associated with poor prognosis in nodal peripheral T-cell lymphomas. *Haematologica* 101, e336-338.

Marina-Zarate, E., Perez-Garcia, A., and Ramiro, A.R. (2017). CCCTC-Binding Factor Locks Premature IgH Germline Transcription and Restrains Class Switch Recombination. *Front Immunol* 8, 1076.

Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463, 568-571.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* 429, 571-574.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17, 3.

Martinez, A., Marin, V.G., Junquera, S.R., Martinez-Murillo, R., and Freire, M. (2005). The contributions of Santiago Ramon y Cajal to cancer research - 100 years on. *Nat Rev Cancer* 5, 904-909.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 7, 29-59.

Mateo, L.J., Murphy, S.E., Hafner, A., Cinquini, I.S., Walker, C.A., and Boettiger, A.N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568, 49-54.

Matthews, A.J., Zheng, S., DiMenna, L.J., and Chaudhuri, J. (2014). Regulation of immunoglobulin class-switch recombination: choreography of noncoding transcription, targeted DNA deamination, and long-range DNA repair. *Adv Immunol* 122, 1-57.

Mattick, J.S. (2004). RNA regulation: a new genetics? *Nat Rev Genet* 5, 316-323.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40, 4288-4297.

McCarthy, D.J., and Smyth, G.K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* 25, 765-771.

Medina, K.L., Pongubala, J.M., Reddy, K.L., Lancki, D.W., Dekoter, R., Kieslinger, M., Grosschedl, R., and Singh, H. (2004). Assembling a gene regulatory network for specification of the B cell fate. *Dev Cell* 7, 607-617.

Medvedovic, J., Ebert, A., Tagoh, H., Tamir, I.M., Schwickert, T.A., Novatchkova, M., Sun, Q., Huis In 't Veld, P.J., Guo, C., Yoon, H.S., *et al.* (2013). Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. *Immunity* 39, 229-244.

Mehra, P., Gerasimova, T., Basu, A., Jha, V., Banerjee, A., Sindhava, V., Gray, F., Berry, C.T., Sen, R., and Atchison, M.L. (2016). YY1 controls Emu-3'RR DNA loop formation and immunoglobulin heavy chain class switch recombination. *Blood Adv* 1, 15-20.

Mele, M., and Rinn, J.L. (2016). "Cat's Cradling" the 3D Genome by the Act of LncRNA Transcription. *Mol Cell* 62, 657-664.

Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., Elkon, R., Melo, S.A., Leveille, N., Kalluri, R., *et al.* (2013). eRNAs are

required for p53-dependent enhancer activity and gene transcription. *Mol Cell* 49, 524-535.

Meng, H., and Bartholomew, B. (2018). Emerging roles of transcriptional enhancers in chromatin looping and promoter-proximal pausing of RNA polymerase II. *J Biol Chem* 293, 13786-13794.

Mercado, R., Vijn, S., Allen, S.E., Kerksiek, K., Pilip, I.M., and Pamer, E.G. (2000). Early programming of T cell populations responding to bacterial infection. *J Immunol* 165, 6833-6839.

Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu Rev Genomics Hum Genet* 17, 17-43.

Minnich, M., Tagoh, H., Bonelt, P., Axelsson, E., Fischer, M., Cebolla, B., Tarakhovsky, A., Nutt, S.L., Jaritz, M., and Busslinger, M. (2016). Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation. *Nat Immunol* 17, 331-343.

Miura, H., Takahashi, S., Poonperm, R., Tanigawa, A., Takebayashi, S.I., and Hiratani, I. (2019). Single-cell DNA replication profiling identifies spatiotemporal developmental dynamics of chromosome organization. *Nat Genet* 51, 1356-1368.

Montefiori, L., Wuerffel, R., Roqueiro, D., Lajoie, B., Guo, C., Gerasimova, T., De, S., Wood, W., Becker, K.G., Dekker, J., *et al.* (2016). Extremely Long-Range Chromatin Loops Link Topological Domains to Facilitate a Diverse Antibody Repertoire. *Cell Rep* 14, 896-906.

Moore, L.D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology* 38, 23-38.

Morgan, R.A., Couture, L., Elroy-Stein, O., Ragheb, J., Moss, B., and Anderson, W.F. (1992). Retroviral vectors containing putative internal ribosome entry sites: development of a polycistronic gene transfer system and applications to human gene therapy. *Nucleic Acids Res* 20, 1293-1299.

Morgan, S.L., Mariano, N.C., Bermudez, A., Arruda, N.L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J.F., *et al.* (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat Commun* 8, 15993.

Moriguchi, T., Hoshino, T., Rao, A., Yu, L., Takai, J., Uemura, S., Ise, K., Nakamura, Y., Lim, K.C., Shimizu, R., *et al.* (2018). A Gata3 3' Distal Otic Vesicle Enhancer Directs Inner Ear-Specific Gata3 Expression. *Mol Cell Biol* 38.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K., and Hatada, I. (2016). Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat Biotechnol* 34, 1060-1065.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* 51, 606-617.

Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., *et al.* (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet* 49, 1602-1612.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102, 553-563.

Nagano, T., Lubling, Y., Varnai, C., Dudley, C., Leung, W., Baran, Y., Mendelson Cohen, N., Wingett, S., Fraser, P., and Tanay, A. (2017). Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* 547, 61-67.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322, 1717-1720.

Nagano, T., Varnai, C., Schoenfelder, S., Javierre, B.M., Wingett, S.W., and Fraser, P. (2015). Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol* 16, 175.

Nakamura, Y., and Hoshino, M. (2005). TH2 cytokines and associated transcription factors as therapeutic targets in asthma. *Curr Drug Targets Inflamm Allergy* 4, 267-270.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347, 1017-1021.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet* 46, 1-19.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. *Science* 342, 948-953.

Nawijn, M.C., Dingjan, G.M., Ferreira, R., Lambrecht, B.N., Karis, A., Grosveld, F., Savelkoul, H., and Hendriks, R.W. (2001a). Enforced expression of GATA-3 in transgenic mice inhibits Th1 differentiation and induces the formation of a T1/ST2-expressing Th2-committed T cell compartment in vivo. *J Immunol* 167, 724-732.

Nawijn, M.C., Ferreira, R., Dingjan, G.M., Kahre, O., Drabek, D., Karis, A., Grosveld, F., and Hendriks, R.W. (2001b). Enforced expression of GATA-3 during T cell development inhibits maturation of CD8 single-positive cells and induces thymic lymphoma in transgenic mice. *J Immunol* 167, 715-723.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* 11, 709-719.

Ngan, C.Y., Wong, C.H., Tjong, H., Wang, W., Goldfeder, R.L., Choi, C., He, H., Gong, L., Lin, J., Urban, B., *et al.* (2020). Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat Genet* 52, 264-272.

Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., *et al.* (2012). c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* 151, 68-79.

Nihongaki, Y., Furuhashi, Y., Otabe, T., Hasegawa, S., Yoshimoto, K., and Sato, M. (2017). CRISPR-Cas9-based photoactivatable transcription systems to induce neuronal differentiation. *Nat Methods* 14, 963-966.

Nihongaki, Y., Kawano, F., Nakajima, T., and Sato, M. (2015). Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* 33, 755-760.

Noguchi, S., Arakawa, T., Fukuda, S., Furuno, M., Hasegawa, A., Hori, F., Ishikawa-Kato, S., Kaida, K., Kaiho, A., Kanamori-Katayama, M., *et al.* (2017). FANTOM5 CAGE profiles of human and mouse samples. *Sci Data* 4, 170112.

Nora, E.P., Caccianini, L., Fudenberg, G., Kameswaran, V., Nagle, A., Uebersohn, A., So, K., Hajj, B., Le Saux, A., and Coulon, A. (2019). Molecular basis of CTCF binding polarity in genome folding. *bioRxiv*.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930-944 e922.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., *et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381-385.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L.A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A* 115, E6697-E6706.

Nutt, S.L., Urbanek, P., Rolink, A., and Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. *Genes Dev* 11, 476-491.

O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N.J. (2004). Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* 36, 1014-1018.

Ogiyama, Y., Schuettengruber, B., Papadopoulos, G.L., Chang, J.M., and Cavalli, G. (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol Cell* 71, 73-88 e75.

Ohmura, S., Mizuno, S., Oishi, H., Ku, C.J., Hermann, M., Hosoya, T., Takahashi, S., and Engel, J.D. (2016). Lineage-affiliated transcription factors bind the Gata3 Tce1 enhancer to mediate lineage-specific programs. *J Clin Invest* 126, 865-878.

Ohno, R., Nakayama, M., Naruse, C., Okashita, N., Takano, O., Tachibana, M., Asano, M., Saitou, M., and Seki, Y. (2013). A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. *Development* 140, 2892-2903.

Oomen, M.E., and Dekker, J. (2017). Epigenetic characteristics of the mitotic chromosome in 1D and 3D. *Crit Rev Biochem Mol Biol* 52, 185-204.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., *et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143, 46-58.

Osborne, C.S., Chakalova, L., Mitchell, J.A., Horton, A., Wood, A.L., Bolland, D.J., Corcoran, A.E., and Fraser, P. (2007). Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol* 5, e192.

Osterwalder, M., Barozzi, I., Tissieres, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., *et al.* (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239-243.

Outters, P., Jaeger, S., Zaarour, N., and Ferrier, P. (2015). Long-Range Control of V(D)J Recombination & Allelic Exclusion: Modeling Views. *Adv Immunol* 128, 363-413.

Pai, S.Y., Truitt, M.L., and Ho, I.C. (2004). GATA-3 deficiency abrogates the development and maintenance of T helper type 2 cells. *Proc Natl Acad Sci U S A* 101, 1993-1998.

Pai, S.Y., Truitt, M.L., Ting, C.N., Leiden, J.M., Glimcher, L.H., and Ho, I.C. (2003). Critical roles for transcription factor GATA-3 in thymocyte development. *Immunity* 19, 863-875.

Palikyras, S., and Papantonis, A. (2019). Modes of phase separation affecting chromatin regulation. *Open Biol* 9, 190167.

Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 35, 190-194.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32, 232-246.

Pang, B., and Snyder, M.P. (2020). Systematic identification of silencers in human cells. *Nat Genet* 52, 254-263.

Park, S.K., Xiang, Y., Feng, X., and Garrard, W.T. (2014). Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. *Genes Dev* 28, 1159-1164.

Pascual-Garcia, P., Debo, B., Aleman, J.R., Talamas, J.A., Lan, Y., Nguyen, N.H., Won, K.J., and Capelson, M. (2017). Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer-Promoter Contacts. *Mol Cell* 66, 63-76 e66.

Pekowska, A., Klaus, B., Xiang, W., Severino, J., Daigle, N., Klein, F.A., Oles, M., Casellas, R., Ellenberg, J., Steinmetz, L.M., *et al.* (2018). Gain of CTCF-Anchored Chromatin Loops Marks the Exit from Naive Pluripotency. *Cell Syst* 7, 482-495 e410.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., *et al.* (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* 10, 973-976.

Perlot, T., Alt, F.W., Bassing, C.H., Suh, H., and Pinaud, E. (2005). Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc Natl Acad Sci U S A* 102, 14362-14367.

Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* 108, 13570-13575.

Pertea, M., and Salzberg, S.L. (2010). Between a chicken and a grape: estimating the number of human genes. *Genome Biol* 11, 206.

Pesu, M., Muul, L., Kanno, Y., and O'Shea, J.J. (2006). Proprotein convertase furin is preferentially expressed in T helper 1 cells and regulates interferon gamma. *Blood* 108, 983-985.

Pflueger, C., Tan, D., Swain, T., Nguyen, T., Pflueger, J., Nefzger, C., Polo, J.M., Ford, E., and Lister, R. (2018). A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. *Genome Res* 28, 1193-1206.

Phanstiel, D.H., Boyle, A.P., Araya, C.L., and Snyder, M.P. (2014). Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* 30, 2808-2810.

Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S., and Smyth, G.K. (2016). Robust Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect Differential Expression. *Ann Appl Stat* 10, 946-963.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., and Gray, J. (1988). Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc Natl Acad Sci U S A* 85, 9138-9142.

Pogo, B.G., Allfrey, V.G., and Mirsky, A.E. (1966). RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. *Proc Natl Acad Sci U S A* 55, 805-812.

Pogo, B.G., Allfrey, V.G., and Mirsky, A.E. (1967). The effect of phytohemagglutinin on ribonucleic acid synthesis and histone acetylation in equine leukocytes. *J Cell Biol* 35, 477-482.

Polstein, L.R., and Gersbach, C.A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat Chem Biol* 11, 198-200.

Pruitt, K.D., and Maglott, D.R. (2001). RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res* 29, 137-140.

Ptak, C., and Wozniak, R.W. (2016). Nucleoporins and chromatin metabolism. *Curr Opin Cell Biol* 40, 153-160.

Pulecio, J., Verma, N., Mejia-Ramirez, E., Huangfu, D., and Raya, A. (2017). CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell* 21, 431-447.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173-1183.

Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., *et al.* (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* 174, 744-757 e724.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., *et al.* (2015). In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520, 186-191.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8, 2281-2308.

Rank, G., Prestel, M., and Paro, R. (2002). Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. *Mol Cell Biol* 22, 8026-8034.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665-1680.

Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., *et al.* (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305-320 e324.

Rawlings, J.S., Gatzka, M., Thomas, P.G., and Ihle, J.N. (2011). Chromatin condensation via the condensin II complex is required for peripheral T-cell quiescence. *EMBO J* 30, 263-276.

Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* 19, 770-788.

Richter, A., Lohning, M., and Radbruch, A. (1999). Instruction for cytokine expression in T helper lymphocytes in relation to proliferation and cell cycle progression. *J Exp Med* 190, 1439-1450.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311-1323.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11, R25.

Rocha, P.P., Raviram, R., Fu, Y., Kim, J., Luo, V.M., Aljoufi, A., Swanzey, E., Pasquarella, A., Balestrini, A., Miraldi, E.R., *et al.* (2016). A Damage-Independent Role for 53BP1 that Impacts Break Order and Igh Architecture during Class Switch Recombination. *Cell Rep* 16, 48-55.

Roden, C., and Gladfelter, A.S. (2020). RNA contributions to the form and function of biomolecular condensates. *Nat Rev Mol Cell Biol.*

Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A., and Misteli, T. (2003). Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat Genet* 34, 287-291.

Roldan, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M., and Skok, J.A. (2005). Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol* 6, 31-41.

Rosa, A., and Everaers, R. (2008). Structure and dynamics of interphase chromosomes. *PLoS Comput Biol* 4, e1000153.

Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V.G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* 67, 837-852 e837.

Rudkin, G.T., and Stollar, B.D. (1977). High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence. *Nature* 265, 472-473.

Rush, J.S., and Hodgkin, P.D. (2001). B cells activated via CD40 and IL-4 undergo a division burst but require continued stimulation to maintain division, survival and differentiation. *Eur J Immunol* 31, 1150-1159.

Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* 3, 793-795.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., *et al.* (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361.

Sakaue-Sawano, A., and Miyawaki, A. (2012). [Visualizing spatiotemporal dynamics of multicellular cell-cycle progression]. *Seikagaku* 84, 47-52.

Salzman, J. (2016). Circular RNA Expression: Its Potential Regulation and Function. *Trends Genet* 32, 309-316.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., *et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* 112, E6456-6465.

Sanulli, S., Trnka, M.J., Dharmarajan, V., Tibble, R.W., Pascal, B.D., Burlingame, A.L., Griffin, P.R., Gross, J.D., and Narlikar, G.J. (2019). HP1 reshapes

nucleosome core to promote phase separation of heterochromatin. *Nature* 575, 390-394.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* 489, 109-113.

Sawada, S., Scarborough, J.D., Killeen, N., and Littman, D.R. (1994). A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. *Cell* 77, 917-929.

Scharer, C.D., Barwick, B.G., Guo, M., Bally, A.P.R., and Boss, J.M. (2018). Plasma cell differentiation is controlled by multiple cell division-coupled epigenetic programs. *Nat Commun* 9, 1698.

Schatz, D.G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol* 11, 251-263.

Schenborn, E., and Groskreutz, D. (1999). Reporter gene vectors and assays. *Mol Biotechnol* 13, 29-44.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev* 19, 697-708.

Schmutz, J., Wheeler, J., Grimwood, J., Dickson, M., Yang, J., Caoile, C., Bajorek, E., Black, S., Chan, Y.M., Denys, M., *et al.* (2004). Quality assessment of the human genome sequence. *Nature* 429, 365-368.

Schoeler, K., Aufschnaiter, A., Messner, S., Derudder, E., Herzog, S., Villunger, A., Rajewsky, K., and Labi, V. (2019). TET enzymes control antibody production and shape the mutational landscape in germinal centre B cells. *FEBS J* 286, 3566-3581.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* 20, 437-455.

Schrock, E., du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M.A., Ning, Y., Ledbetter, D.H., Bar-Am, I., Soenksen, D., *et al.* (1996). Multicolor spectral karyotyping of human chromosomes. *Science* 273, 494-497.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., and Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51-56.

Sekimata, M., Perez-Melgosa, M., Miller, S.A., Weinmann, A.S., Sabo, P.J., Sandstrom, R., Dorschner, M.O., Stamatoyannopoulos, J.A., and Wilson, C.B. (2009). CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. *Immunity* 31, 551-564.

Serra, F., Bau, D., Goodstadt, M., Castillo, D., Filion, G.J., and Marti-Renom, M.A. (2017). Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput Biol* 13, e1005665.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 148, 458-472.

Shaffer, A.L., Lin, K.I., Kuo, T.C., Yu, X., Hurt, E.M., Rosenwald, A., Giltzane, J.M., Yang, L., Zhao, H., Calame, K., and Staudt, L.M. (2002). Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* 17, 51-62.

Shandilya, J., and Roberts, S.G. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta* 1819, 391-400.

Shapiro-Shelef, M., Lin, K.I., McHeyzer-Williams, L.J., Liao, J., McHeyzer-Williams, M.G., and Calame, K. (2003). Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* 19, 607-620.

Shechner, D.M., Hacisuleyman, E., Younger, S.T., and Rinn, J.L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat Methods* 12, 664-670.

Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A.V., Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., and Chakraborty, A.K. (2019). Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol Cell* 75, 549-561 e547.

Shukla, A., and Huangfu, D. (2018). Decoding the noncoding genome via large-scale CRISPR screens. *Curr Opin Genet Dev* 52, 70-76.

Simeonov, D.R., Gowen, B.G., Boontanrart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., *et al.* (2017).

Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* 549, 111-115.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 38, 1348-1354.

Singh, B.N., and Hampsey, M. (2007). A transcription-independent role for TFIIB in gene looping. *Mol Cell* 27, 806-816.

Siu, G., Wurster, A.L., Duncan, D.D., Soliman, T.M., and Hedrick, S.M. (1994). A transcriptional silencer controls the developmental expression of the CD4 gene. *EMBO J* 13, 3570-3579.

Sood, V., and Brickner, J.H. (2014). Nuclear pore interactions with the genome. *Curr Opin Genet Dev* 25, 43-49.

Spector, D.L. (1993). Macromolecular domains within the cell nucleus. *Annu Rev Cell Biol* 9, 265-315.

Spilianakis, C.G., and Flavell, R.A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* 5, 1017-1027.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. (2005). Interchromosomal associations between alternatively expressed loci. *Nature* 435, 637-645.

Stadhouders, R., de Bruijn, M.J., Rother, M.B., Yuvaraj, S., Ribeiro de Almeida, C., Kolovos, P., Van Zelm, M.C., van Ijcken, W., Grosveld, F., Soler, E., and Hendriks, R.W. (2014). Pre-B cell receptor signaling induces immunoglobulin kappa locus accessibility by functional redistribution of enhancer-mediated chromatin interactions. *PLoS Biol* 12, e1001791.

Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y., *et al.* (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat Genet* 50, 238-249.

Stavnezer, J., and Schrader, C.E. (2014). IgH chain class switch recombination: mechanism and regulation. *J Immunol* 193, 5370-5378.

Stik, G., Vidal, E., Barrero, M., Cuartero, S., Vila-Casadesus, M., Mendieta-Esteban, J., Tian, T.V., Choi, J., Berenguer, C., Abad, A., *et al.* (2020). CTCF is

dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. *Nat Genet* 52, 655-661.

Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V., and Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 365, 48-53.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. *Nature* 547, 241-245.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* 14, 103-105.

Stubington, M.J., and Corcoran, A.E. (2013). Non-coding transcription and large-scale nuclear organisation of immunoglobulin recombination. *Curr Opin Genet Dev* 23, 81-88.

Swarts, D.C., and Jinek, M. (2018). Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip Rev RNA*, e1481.

Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res* 24, 390-400.

Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Sci Adv* 5, eaaw1668.

Taghon, T., De Smedt, M., Stolz, F., Cnockaert, M., Plum, J., and Leclercq, G. (2001). Enforced expression of GATA-3 severely reduces human thymic cellularity. *J Immunol* 167, 4468-4475.

Taghon, T., Yui, M.A., and Rothenberg, E.V. (2007). Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nat Immunol* 8, 845-855.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930-935.

Tai, T.S., Pai, S.Y., and Ho, I.C. (2013). GATA-3 regulates the homeostasis and activation of CD8⁺ T cells. *J Immunol* 190, 428-437.

Tan-Wong, S.M., French, J.D., Proudfoot, N.J., and Brown, M.A. (2008). Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Proc Natl Acad Sci U S A* 105, 5160-5165.

Tan-Wong, S.M., Wijayatilake, H.D., and Proudfoot, N.J. (2009). Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev* 23, 2610-2624.

Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., and Proudfoot, N.J. (2012). Gene loops enhance transcriptional directionality. *Science* 338, 671-675.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635-646.

Taniuchi, I., Osato, M., Egawa, T., Sunshine, M.J., Bae, S.C., Komori, T., Ito, Y., and Littman, D.R. (2002a). Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell* 111, 621-633.

Taniuchi, I., Sunshine, M.J., Festenstein, R., and Littman, D.R. (2002b). Evidence for distinct CD4 silencer functions at different stages of thymocyte differentiation. *Mol Cell* 10, 1083-1096.

ter Brake, O., t Hooft, K., Liu, Y.P., Centlivre, M., von Eije, K.J., and Berkhout, B. (2008). Lentiviral vector design for multiple shRNA expression and durable HIV-1 inhibition. *Mol Ther* 16, 557-564.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* 12, 1143-1149.

Thomas-Claudepierre, A.S., Robert, I., Rocha, P.P., Raviram, R., Schiavo, E., Heyer, V., Bonneau, R., Luo, V.M., Reddy, J.K., Borggreffe, T., *et al.* (2016). Mediator facilitates transcriptional activation and dynamic long-range contacts at the IgH locus during class switch recombination. *J Exp Med* 213, 303-312.

Ting, C.N., Olson, M.C., Barton, K.P., and Leiden, J.M. (1996). Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature* 384, 474-478.

Tiwari, V.K., McGarvey, K.M., Licchesi, J.D., Ohm, J.E., Herman, J.G., Schubeler, D., and Baylin, S.B. (2008). PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol* 6, 2911-2927.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 10, 1453-1465.

Tomura, M., Sakaue-Sawano, A., Mori, Y., Takase-Utsugi, M., Hata, A., Ohtawa, K., Kanagawa, O., and Miyawaki, A. (2013). Contrasting quiescent G0 phase with mitotic cell cycling in the mouse immune system. *PLoS One* 8, e73801.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* 302, 575-581.

Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2010). B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. *Nat Immunol* 11, 148-154.

Trimarchi, T., Bilal, E., Ntziachristos, P., Fabbri, G., Dalla-Favera, R., Tsiganos, A., and Aifantis, I. (2014). Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell* 158, 593-606.

Turner, M.L., Hawkins, E.D., and Hodgkin, P.D. (2008). Quantitative regulation of B cell division destiny by signal strength. *J Immunol* 181, 374-382.

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lonnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggstrom, J., Kharchenko, O., Kharchenko, P.V., *et al.* (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat Neurosci* 18, 145-153.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* 474, 516-520.

van Hamburg, J.P., de Bruijn, M.J., Dingjan, G.M., Beverloo, H.B., Diepstraten, H., Ling, K.W., and Hendriks, R.W. (2008). Cooperation of Gata3, c-Myc and Notch in malignant transformation of double positive thymocytes. *Mol Immunol* 45, 3085-3095.

van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* 169, 780-791.

van Steensel, B., and Furlong, E.E.M. (2019). The role of transcription in shaping the spatial organization of the genome. *Nat Rev Mol Cell Biol* 20, 327-337.

van Stipdonk, M.J., Hardenberg, G., Bijker, M.S., Lemmens, E.E., Droin, N.M., Green, D.R., and Schoenberger, S.P. (2003). Dynamic programming of CD8+ T lymphocyte responses. *Nat Immunol* 4, 361-365.

van Stipdonk, M.J., Lemmens, E.E., and Schoenberger, S.P. (2001). Naive CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. *Nat Immunol* 2, 423-429.

Verma-Gaur, J., Torkamani, A., Schaffer, L., Head, S.R., Schork, N.J., and Feeney, A.J. (2012). Noncoding transcription within the Igh distal V(H) region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells. *Proc Natl Acad Sci U S A* 109, 17004-17009.

Wang, H., Xu, X., Nguyen, C.M., Liu, Y., Gao, Y., Lin, X., Daley, T., Kipniss, N.H., La Russa, M., and Qi, L.S. (2018). CRISPR-Mediated Programmable 3D Genome Positioning and Nuclear Organization. *Cell* 175, 1405-1417 e1414.

Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J., and Wong, G.K. (2004). Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. *Nature* 431, 1 p following 757; discussion following 757.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., *et al.* (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120-124.

Wang, L., Gao, Y., Zheng, X., Liu, C., Dong, S., Li, R., Zhang, G., Wei, Y., Qu, H., Li, Y., *et al.* (2019). Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. *Mol Cell* 76, 646-659 e646.

Wang, L., Wildt, K.F., Zhu, J., Zhang, X., Feigenbaum, L., Tessarollo, L., Paul, W.E., Fowlkes, B.J., and Bosselut, R. (2008). Distinct functions for the transcription factors GATA-3 and ThPOK during intrathymic differentiation of CD4(+) T cells. *Nat Immunol* 9, 1122-1130.

Watson, M.L. (1955). The nuclear envelope; its structure and relation to cytoplasmic membranes. *J Biophys Biochem Cytol* 1, 257-270.

Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res* 20, 90-100.

Weintraub, H., Flint, S.J., Leffak, I.M., Groudine, M., and Grainger, R.M. (1978). The generation and propagation of variegated chromosome structures. *Cold Spring Harb Symp Quant Biol* 42 Pt 1, 401-407.

Weirauch, M.T., Cote, A., Norel, R., Annala, M., Zhao, Y., Riley, T.R., Saez-Rodriguez, J., Cokelaer, T., Vedenko, A., Talukder, S., *et al.* (2013). Evaluation of methods for modeling transcription factor sequence specificity. *Nat Biotechnol* 31, 126-134.

Williams, M.A., and Bevan, M.J. (2004). Shortening the infectious period does not alter expansion of CD8 T cells but diminishes their capacity to differentiate into memory cells. *J Immunol* 173, 6694-6702.

Wilson, C.J., Chao, D.M., Imbalzano, A.N., Schnitzler, G.R., Kingston, R.E., and Young, R.A. (1996). RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* 84, 235-244.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 23, 1494-1504.

Wolffe, A.P. (1991). Activating chromatin. *Curr Biol* 1, 366-368.

Wong, P., and Pamer, E.G. (2001). Cutting edge: antigen-independent CD8 T cell proliferation. *J Immunol* 166, 5864-5868.

Wuerffel, R., Wang, L., Grigera, F., Manis, J., Selsing, E., Perlot, T., Alt, F.W., Cogne, M., Pinaud, E., and Kenter, A.L. (2007). S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* 27, 711-722.

Xiang, J.F., Yin, Q.F., Chen, T., Zhang, Y., Zhang, X.O., Wu, Z., Zhang, S., Wang, H.B., Ge, J., Lu, X., *et al.* (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 24, 513-531.

Xu, C., and Corces, V.G. (2016). Towards a predictive model of chromatin 3D organization. *Semin Cell Dev Biol* 57, 24-30.

Xu, X., and Qi, L.S. (2019). A CRISPR-dCas Toolbox for Genetic Engineering and Synthetic Biology. *J Mol Biol* 431, 34-47.

Xu, Z., Zan, H., Pone, E.J., Mai, T., and Casali, P. (2012). Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat Rev Immunol* 12, 517-531.

Yancopoulos, G.D., and Alt, F.W. (1985). Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* 40, 271-281.

Yewdell, W.T., and Chaudhuri, J. (2017). A transcriptional serenAID: the role of noncoding RNAs in class switch recombination. *Int Immunol* 29, 183-196.

Yoshida, H., Lareau, C.A., Ramirez, R.N., Rose, S.A., Maier, B., Wroblewska, A., Desland, F., Chudnovskiy, A., Mortha, A., Dominguez, C., *et al.* (2019). The cis-Regulatory Atlas of the Mouse Immune System. *Cell* 176, 897-912 e820.

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P., and Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451, 202-206.

Yui, M.A., and Rothenberg, E.V. (2014). Developmental gene networks: a triathlon on the course to T cell identity. *Nat Rev Immunol* 14, 529-545.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A* 108, 1531-1536.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., and Lim, W.A. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339-350.

Zandi, S., Mansson, R., Tsapogas, P., Zetterblad, J., Bryder, D., and Sigvardsson, M. (2008). EBF1 is essential for B-lineage priming and establishment of a transcription factor network in common lymphoid progenitors. *J Immunol* 181, 3364-3372.

Zeng, Y., and Cullen, B.R. (2002). RNA interference in human cells is restricted to the cytoplasm. *RNA* 8, 855-860.

Zhang, W., Wang, Z., Luo, Y., Zhong, D., Luo, Y., and Zhou, D. (2016). GATA3 expression correlates with poor prognosis and tumor-associated macrophage infiltration in peripheral T cell lymphoma. *Oncotarget* 7, 65284-65294.

Zhang, X., Zhang, Y., Ba, Z., Kyritsis, N., Casellas, R., and Alt, F.W. (2019a). Fundamental roles of chromatin loop extrusion in antibody class switching. *Nature* 575, 385-389.

Zhang, Y., Zhang, X., Ba, Z., Liang, Z., Dring, E.W., Hu, H., Lou, J., Kyritsis, N., Zurita, J., Shamim, M.S., *et al.* (2019b). The fundamental role of chromatin loop extrusion in physiological V(D)J recombination. *Nature* 573, 600-604.

Zhao, M.L., Rabiee, A., Kovary, K.M., Bahrami-Nejad, Z., Taylor, B., and Teruel, M.N. (2020). Molecular Competition in G1 Controls When Cells Simultaneously Commit to Terminally Differentiate and Exit the Cell Cycle. *Cell Rep* 31, 107769.

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., *et al.* (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* 38, 1341-1347.

Zhou, J.H.S., Markham, J.F., Duffy, K.R., and Hodgkin, P.D. (2018). Stochastically Timed Competition Between Division and Differentiation Fates Regulates the Transition From B Lymphoblast to Plasma Cell. *Front Immunol* 9, 2053.

Zhu, J., Min, B., Hu-Li, J., Watson, C.J., Grinberg, A., Wang, Q., Killeen, N., Urban, J.F., Jr., Guo, L., and Paul, W.E. (2004). Conditional deletion of Gata3 shows its essential function in T(H)1-T(H)2 responses. *Nat Immunol* 5, 1157-1165.

Zhu, J., Yamane, H., and Paul, W.E. (2010). Differentiation of effector CD4 T cell populations (*). *Annu Rev Immunol* 28, 445-489.

Ziegler, C., and Kretz, M. (2017). The More the Merrier-Complexity in Long Non-Coding RNA Loci. *Front Endocrinol (Lausanne)* 8, 90.