Molecular determinants of chromatin accessibility at CpG islands in mouse embryonic stem cells

Hamish William King

Green Templeton College

University of Oxford



A thesis submitted for the degree of

Doctor of Philosophy

DECLARATION OF AUTHORSHIP

I declare that this thesis and research therein is my own unless otherwise acknowledged.

This thesis has not been submitted, either partially or in full, for another degree, diploma, certificate or other qualification at this University or at any other institution

Hamish William King

5th October 2016

ACKNOWLEDGEMENTS

I would like to begin this thesis by acknowledging and thanking the many people who have helped me throughout my PhD studies in a myriad of different ways.

Firstly, I am incredibly grateful to my supervisor, Prof Rob Klose, for all of his efforts to teach and guide me during these exciting four years. Thank you Rob for giving me the opportunity and inspiration to challenge myself with these exciting biological questions, and for providing an incredibly stimulating environment and supervision for my PhD. Your enthusiasm for science has proved completely infectious.

In addition to this fantastic supervision, I have been very lucky to have many wonderful colleagues who have not only helped me with experiments, but have provided a great amount of help and emotional support throughout the many challenges of the PhD. A big thank you to everyone past and present in the lab, particularly Neil, Emilia, Dave, Nathan, Anne, Nadya, Hannah, Anca, Vincenzo, Nadya, Angelika and Miles. It has been a lot of fun talking about such interesting science with you all.

I would also like to thank Doug Higgs and Jim Hughes at the WIMM for their supervision during my rotation project in their lab, in which I learnt a huge amount. Thanks also to Skirmantas Kriaucionis and Martin Cusack for their collaboration, and to Neil Brockdorff and everyone in his lab for helpful discussions and contributing to a great lab environment.

Outside of the lab, I would like to thank my friends Lars Hanssen, Jon Kerry, Guy Flint, Charlotte Koldeweij, Daisy Johnson, Gabby Penfold, Imran House, Rhiannon Murrie and Toby Phillips for their friendship and support.

I would also like to thank both my parents, Alison and David, for constantly encouraging and supporting me, and having my best interests at heart. I am very grateful for the hard work and sacrifices that you have made to help me be here today. A big thank you also to my sister Bridget, for all your help and support along the way.

Finally, I need to thank Sian. You have been there for me throughout the good and bad during the last four years, and have helped me in too many ways to count. You make me want to be better, to strive further, and to achieve greater things, and I hope that I can always be there for you in the same way.

ABSTRACT

In eukaryotic cells, transcription factors and polymerases must access DNA in the context of nucleosomes and chromatin. The accessibility of DNA sequences to such *trans*-acting factors is an important feature of gene regulatory elements, including promoters. In vertebrates, the majority of gene promoters coincide with CpG islands (CGIs), which remain free from DNA methylation and exhibit elevated CpG densities. This hypomethylated and CpG-rich state at CGI promoters is associated not only with transcriptional activity, but also with high levels of chromatin accessibility. However, the causes and consequences of such chromatin accessibility remain unclear. To address this, I have profiled chromatin accessibility in mouse embryonic stem cells (ESCs). In addition to confirming that CGI accessibility is independent of transcriptional activity, I was able to demonstrate that the loss of DNA methylation in ESCs resulted in increased chromatin accessibility at a subset of CpG-rich repetitive elements, suggesting that non-methylated CpG-rich sequences may, at least partially, facilitate open chromatin states. This was supported by preliminary work targeting bacterial CpG-rich sequences into the mouse genome, where they were sufficient to establish novel regions of chromatin accessibility. To examine potential mechanisms by which hypomethylated DNA could serve to promote chromatin accessibility, I profiled chromatin accessibility in mouse ESCs lacking various chromatin-modifying proteins which are normally enriched at CGIs, with the histone demethylases KDM2A/B linked to maintaining open chromatin at CGIs. As an alternative approach to understanding the causes of chromatin accessibility in mouse ESCs, I examined the mechanism by which the pioneer transcription factor OCT4 is able to access previously inaccessible chromatin, and reveal that it requires the chromatin remodeller BRG1 to remodel chromatin and facilitate transcription factor binding at distal regulatory elements. Ultimately, this work provides an insight into some of the molecular determinants of chromatin accessibility in mouse ESCs, although many of the consequences of such chromatin states remain unclear.

TABLE OF CONTENTS

DECLARATION OF AUTHORSHIP **1**

ACKNOWLEDGEMENTS **2**

ABSTRACT **3**

TABLE OF CONTENTS **4**

LIST OF FIGURES **9**

LIST OF TABLES **11**

LIST OF SUPPLEMENTARY FIGURES **11**

ABBREVIATIONS **12**

CHAPTER 1 - INTRODUCTION **14**

1.1 REGULATION OF GENE EXPRESSION IN EUKARYOTIC CELLS 14

1.1.1 General principles of eukaryotic gene regulation 14

1.1.2 Histone modifications at gene regulatory elements 15

1.2 LACK OF DNA METHYLATION AT CPG ISLANDS REGULATES PROMOTER FUNCTION 17

1.2.1 Cytosine methylation by DNA methyltransferases 18

1.2.2 Transcriptional repression by DNA methylation 19

1.2.3 Distribution of methylated cytosine in vertebrate genomes 19

1.2.4 Potential mechanisms for lack of DNA methylation at CGIs 20

1.2.5 Non-methylated CGI promoters are transcriptionally active 21

1.3 INTERPLAY BETWEEN DNA METHYLATION AND CHROMATIN MODIFICATIONS 22

1.3.1 Methyl-binding proteins recognise DNA methylation 22

1.3.2 ZF-CxxC domain-containing proteins bind non-methylated DNA 24

1.3.3 CGIs recruit diverse chromatin-modifying activities 24

1.3.3.1 KDM2A/B histone demethylases and FBXL19 25

1.3.3.2 CFP1 and MLL1/2 regulate H3K4me3 deposition 26

1.3.3.3 Recruitment of polycomb repressive complexes 29

1.3.4 CGIs function as unique regulatory platforms at promoters 29

1.3.4.1 Antagonism between TrxG and PcG systems 30

1.3.4.2 Feedback mechanisms to maintain TrxG/ PcG chromatin states 31
1.3.4.3 Bistability at CGIs for stable, but responsive, transcriptional states 31
1.4 CHROMATIN ACCESSIBILITY AT CGIS AND OTHER REGULATORY ELEMENTS 33
1.4.1 Definitions and measurements of chromatin accessibility 33
1.4.1.1 Nuclease hypersensitivity as a direct measure of chromatin accessibility 34
1.4.1.2 Measuring nucleosome occupancy using MNase-seq 35
1.4.1.3 Alternative chromatin accessibility protocols 36
1.4.1.4 Chromatin accessibility as altered nucleosome occupancy or structure 37
1.4.2 Chromatin structure and accessibility of CGI promoters 38
1.4.2.1 Non-methylated CGIs are hypersensitive to nucleases 38
1.4.2.2 Polycomb-occupied CGIs have limited chromatin accessibility 39
1.5 TRANSCRIPTION FACTORS AND THE CHROMATIN BARRIER 40
1.5.1 ATP-dependent chromatin remodelling complexes 41
1.5.2 Pioneer transcription factors to bind inaccessible chromatin 42
1.6 KEY AIMS AND QUESTIONS 45

CHAPTER 2 - MATERIALS AND METHODS 47

2.1 CELL LINES AND CULTURE CONDITIONS 47
2.1.1 Mouse embryonic stem cell culture 47
2.1.2 Inhibition of RNA polymerase II 48
2.1.3 Alkaline phosphatase staining 48
2.1.4 Nuclei isolation and optimisation 48
2.2 MOLECULAR BIOLOGY 49
2.2.1 Nuclear extracts 49
2.2.2 Acid extraction of histones 49
2.2.3 SDS-PAGE and Western blotting 50
2.3 EXPRESSION AND PURIFICATION OF TN5 TRANSPOSASE 51
2.3.1 Preparation of Tn5 MEDS oligonucleotides 51
2.3.2 *E.coli* expression of Tn5 transposase 51
2.3.3 Purification of Tn5 by thiol-mediated cleavage of intein-CBD 52
2.3.4 Testing *in vitro* activity of purified Tn5 transposase 53
2.4 MEASURING CHROMATIN ACCESSIBILITY 53
2.4.1 DNase I digestion and DNase-seq library generation 53
2.4.2 ATAC-seq sample and library preparation 54
2.4.3 DNase- and ATAC-qPCR 55
2.5 CHROMATIN IMMUNOPRECIPITATION 56
2.6 NUCLEAR RNA ISOLATION AND SEQUENCING 57
2.7 RECOMBINASE-MEDIATED CASSETTE EXCHANGE 58
2.7.1 Generating RMCE-HygroTK ESCs 58
2.7.2 Generation of targeting construct and cloning *S. enterica* DNA fragments 58

2.7.3 Insertion of bacterial sequence into RMCE locus 59
2.7.4 Methylation-sensitive qPCR 60
2.8 MASSIVELY PARALLEL SEQUENCING, DATA PROCESSING AND NORMALISATION 60
2.8.1 Sequencing using IonProton and Illumina technologies 60
2.8.2 Data alignment, normalisation and visualisation 61
2.8.2.1 Calibrating ATAC-seq with Drosophila control 62
2.9 ANNOTATION AND ANALYSIS OF REGULATORY ELEMENTS 63
2.9.1 Peak calling 63
2.9.2 Annotation of putative promoters and enhancers 63
2.9.3 Genome-wide correlations 63
2.9.4 Gene ontology and motif analysis 64
2.10 DIFFERENTIAL ACCESSIBILITY, FACTOR BINDING AND GENE EXPRESSION 64
2.11 REPETITIVE ELEMENTS ANALYSIS 65
2.12 ACCESSION CODES AND GENOMIC DATASETS 66

CHAPTER 3 - EXAMINING CHROMATIN ACCESSIBILITY IN EMBRYONIC STEM CELLS 70

3.1 INTRODUCTION 70
3.2 RESULTS 71
3.2.1 Application of DNase-seq to mouse embryonic stem cells 71
3.2.1.1 Optimisation of nuclei isolation for DNase-seq 72
3.2.1.2 Qualitative and quantitative assessment of DNase I digestion 74
3.2.1.3 Application of DNase-seq to mouse ESCs 75
3.2.1.4 An examination of DNase-seq library fragment sizes 77
3.2.1.5 Altered nuclei isolation aids detection of enhancer DHS 79
3.2.1.6 DNase-seq profiles are not robust across biological replicates 81
3.2.2 Using ATAC-seq to measure chromatin accessibility 83
3.2.2.1 Expression and purification of Tn5 transposase in *E. coli* 84
3.2.2.2 Application of purified Tn5 transposase to ATAC-seq of mouse ESCs 86
3.3 DISCUSSION 90
3.3.1 Technical difficulties of genome-wide chromatin accessibility assays 90

CHAPTER 4 - CHROMATIN ACCESSIBILITY AT CPG ISLANDS 93

4.1 INTRODUCTION 93
4.2 RESULTS 94
4.2.1 CpG density, DNA methylation and chromatin accessibility 94
4.2.1.1 CGIs exhibit chromatin accessibility across many cell types 95
4.2.1.2 Non-methylated DNA is associated with chromatin accessibility 97
4.2.2 Chromatin accessibility at NMI promoters is independent of transcription 100

4.2.2.1 Promoter accessibility correlates with CpG density rather than transcriptional activity. 100
4.2.2.2 NMI accessibility is maintained after RNAPII inhibition. 103
4.2.3 Loss of DNA methylation results in chromatin accessibility changes. 106
4.2.3.1 DNA methylation limits chromatin accessibility of many distal regulatory elements. 106
4.2.3.2 Methylated CGIs do not gain accessibility after loss of DNA methylation. 109
4.2.3.3 CpG-rich repeats gain chromatin accessibility upon DNA methylation loss. 110
4.2.4 Is non-methylated, CpG-rich DNA sufficient to create chromatin accessibility? 112
4.2.4.1 Development and optimisation of RMCE targeting system. 114
4.2.4.2 Targeting bacterial CpG-rich sequences into the mouse genome. 116
4.3 DISCUSSION. 120
4.3.1 Evidence for and against CpG density directly influencing chromatin structure. 121
4.3.2 DNA methylation-dependent pathways to establish or constrain accessibility. 122
4.3.3 Transcription factors to shape hypomethylated and hypersensitive chromatin. 125
4.3.4 Conclusions and future directions. 127

CHAPTER 5 - REGULATION OF CGI ACCESSIBILITY BY CHROMATIN-MODIFYING PATHWAYS. 128

5.1 INTRODUCTION. 128
5.2 RESULTS. 130
5.2.1 ZF-CxxC domain-containing proteins and chromatin accessibility at CGIs. 130
5.2.1.1 FBXL19 does not regulate chromatin accessibility in ESCs. 130
5.2.1.2 Chromatin accessibility at CFP1-dependent NMIs. 133
5.2.1.3 The KDM2 histone demethylases regulate NMI chromatin accessibility. 137
5.2.1.4 KDM2-dependent accessibility is driven by KDM2A/B-SF isoforms. 140
5.2.2 The regulation of chromatin accessibility by polycomb repressive complexes. 142
5.2.2.1 Polycomb-occupied NMIs have reduced chromatin accessibility. 143
5.2.2.2 Measuring chromatin accessibility in PRC1- or PRC2-null ESCs. 144
5.3 DISCUSSION. 151
5.3.1 Potential mechanisms for the regulation of chromatin structure at CGIs. 151
5.3.2 PcG-dependent or –independent chromatin compaction at CGIs. 154
5.3.3 Conclusions and Future Directions. 159

CHAPTER 6 - REGULATION OF CHROMATIN ACCESSIBILITY BY THE PIONEER FACTOR OCT4. 160

6.1 INTRODUCTION. 160
6.2 RESULTS. 163
6.2.1 OCT4 regulates chromatin structure and transcription factor binding. 163
6.2.1.1 OCT4 is required for chromatin accessibility at its binding sites. 163
6.2.1.2 OCT4 supports pluripotency factor binding and gene regulation. 167
6.2.2 OCT4 targets a chromatin remodeller to regulate chromatin accessibility. 170

6.2.2.1 OCT4-dependent chromatin accessibility correlates the chromatin remodeller BRG1 . 170
6.2.3 BRG1 regulates chromatin accessibility at OCT4 target sites 172
6.2.4 OCT4 recruits BRG1 to chromatin..... 175
6.2.5 BRG1 is required to mature transcription factor binding..... 176
6.2.5.1 BRG1 is required for OCT4/SOX2 DNA binding and gene regulation 177
6.2.6 Distinct modes of OCT4 binding and function during reprogramming and development..... 181
6.3 DISCUSSION..... 184
6.3.1 Cooperation between OCT4 and BRG1/BAF complexes..... 184
6.3.2 Tissue-specific chromatin remodelling for sequence-specific regulation..... 186
6.3.3 Models for maturation of transcription factor binding by BRG1 187
6.3.4 Conclusions and future directions 188

CHAPTER 7 - CONCLUSIONS AND FUTURE DIRECTIONS 189

APPENDICES 193

APPENDIX ONE – SUPPLEMENTARY FIGURES 193

REFERENCE LIST 198

LIST OF FIGURES

Figure 1.3.1 Type 1 ZF-CxxC domain-containing proteins. 23

Figure 3.2.1 Optimisation of nuclei isolation from mouse embryonic stem cells. 73

Figure 3.2.2 Electrophoresis and DNase-qPCR of ESC DNase I digestion curve. 75

Figure 3.2.3 DNase I digestion curve of mouse ESCs followed by sequencing. 76

Figure 3.2.4 Variation of size selection for DNase-seq library preparation. 78

Figure 3.2.5 DNase I digestion of mouse ESCs with altered NP-40 concentration. 80

Figure 3.2.6 Failure to robustly reproduce DNase-seq profiles. 82

Figure 3.2.7 Expression and purification of Tn5 transposase. 85

Figure 3.2.8 Trial ATAC-seq experiment profiles chromatin accessibility, but with poor alignment statistics. 87

Figure 3.2.9 ATAC-seq is a robust measure of chromatin accessibility. 89

Figure 4.2.1 CpG-rich DNA exhibits chromatin accessibility across many cell types. 96

Figure 4.2.2 Examining chromatin accessibility at non-methylated DNA genome-wide. 99

Figure 4.2.3 Relationship between gene expression and promoter accessibility. 102

Figure 4.2.4 RNAPII has a minor influence on chromatin accessibility at promoters. .. 105

Figure 4.2.5 Loss of DNA methylation results in changes in chromatin accessibility at regulatory elements. 108

Figure 4.2.6 Methylated CGIs do not gain chromatin accessibility upon DNA methylation loss. 110

Figure 4.2.7 Chromatin accessibility at CpG-rich repetitive elements is limited by DNA methylation. 113

Figure 4.2.8 Establishing recombinase-mediated cassette exchange system in ESCs. 115

Figure 4.2.9 Integration of non-methylated, CpG-rich bacterial DNA into mouse ESCs. 118

Figure 5.2.1 Loss of ZF-CxxC domain in FBXL19 does not affect chromatin accessibility. 132
Figure 5.2.2 Conditional deletion of CFP1 results in reduced H3K4me3 with no effect on NMI accessibility. 134
Figure 5.2.3 Chromatin accessibility at CFP1-dependent or CFP–independent H3K4me3 promoters. 136
Figure 5.2.4 Mouse ESC cell lines to study KDM2A/B isoform function. 139
Figure 5.2.5 Chromatin accessibility in the absence of KDM2A and KDM2B. 140
Figure 5.2.6 KDM2-SF isoforms are responsible for majority of KDM2-dependent chromatin accessibility. 142
Figure 5.2.7 Reduced chromatin accessibility at PcG-occupied NMIs. 144
Figure 5.2.8 Loss of PRC1 or PRC2 does not increase chromatin accessibility of PcG-occupied NMIs. 146
Figure 5.2.9 PcG complexes contribute to nuclease sensitivity of PcG-occupied NMIs. 148
Figure 5.2.10 CRISPR/Cas9 deletion of *Eed* in *Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ ESCs. 150
Figure 6.2.1 Deletion of OCT4 using ZHBTC4 ESCs. 164
Figure 6.2.2 OCT4 regulates chromatin accessibility at distal regulatory sites. 166
Figure 6.2.3 Loss of OCT4 results in altered pluripotency factor occupancy and down-regulation of neighbouring genes. 169
Figure 6.2.4 OCT4-dependent chromatin accessibility is correlated with BRG1. 172
Figure 6.2.5 BRG1 regulates chromatin accessibility at OCT4 target sites. 174
Figure 6.2.6 OCT4 recruits BRG1 to chromatin. 176
Figure 6.2.7 OCT4 and SOX2 require BRG1 to bind DNA and regulate gene expression. 180
Figure 6.2.8 OCT4s requirement for BRG1 chromatin remodelling reveals distinct timing of regulatory events during reprogramming and development. 183

LIST OF TABLES

Table 1 Genotype and treatment conditions for mouse embryonic stem cell lines. 47

Table 2 Antibody details for western blotting..... 51

Table 3 Primers for DNase- and ATAC-qPCR 56

Table 4 Primers for cloning *S. enterica* fragments. 59

Table 5 Published genomic datasets included in this study. 66

Table 6 Genomic datasets generated in this study. 66

Table 7 Calibrated ATAC-seq samples and sequencing depth. 69

LIST OF SUPPLEMENTARY FIGURES

Supplementary Figure 1 Reactivation of repetitive elements only partially explains changes in chromatin accessibility in DNMT TKO cells..... 193

Supplementary Figure 2 Experimental design of bacterial DNA sequences. 194

Supplementary Figure 3 Transcription factor motifs in RMCE bacterial sequences. ... 195

Supplementary Figure 4 The ZF-CxxC domain of FBXL19 is essential for occupancy at NMI promoters..... 195

Supplementary Figure 5 Redundancy between CFP1, MLL1 and MLL2 for deposition of H3K4me3. 196

Supplementary Figure 6 OCT4-dependent regulation of pluripotency transcriptional network..... 197

ABBREVIATIONS

4SU	4-thiouridine
4SU-seq	4SU pulse-labelled RNA followed by massively parallel sequencing
ATAC	assay for transposase-accessible chromatin
BioCAP	biotinylated CxxC affinity purification
CAGE	cap analysis of gene expression
CBD	chitin-binding domain
CFP1	CxxC finger protein 1
CGI	CpG island (computationally predicted)
ChIP	chromatin immunoprecipitation
ChIP-seq	ChIP followed by massively parallel sequencing
Cre-ERT	Cre recombinase-estrogen receptor fusion
DHS	DNase I hypersensitive site
DIC	differential interference contrast
DNase I	deoxyribonuclease I
DNase-qPCR	DNase I digestion followed by qPCR
DNase-seq	DNase I digestion followed by massively parallel sequencing
DNMT	DNA methyltransferase
DNMT TKO	DNA methyltransferase triple knockout
DOX	doxycycline
ESC	embryonic stem cell
FAIRE	formaldehyde assisted isolation of regulatory elements
FBXL19	F-box and leucine-rich repeat protein 19
FLAV	flavopiridol (CDK9 inhibitor)
HDAC	histone deacetylase complex
HygroTK	hygromycin phosphotransferase-thymidine kinase
IAP	intra-cisternal A-type particle
iPSC	induced pluripotent stem cell
JmjC	Jumonji C (histone demethylase domain)
KDM2A	lysine demethylase 2A (JHDM1a/FBXL11)
KDM2B	lysine demethylase 2B (JHDM1b/FBXL10)
LIC	ligation independent cloning
LRR	leucine rich-repeat
LTR	long terminal repeats
MBD	methyl binding domain
ME	mosaic end (Tn5 recognition sequences)

MeCP2	methyl-C binding protein
MEDS	mosaic end double stranded (oligonucleotides)
MLL1/2	mixed lineage leukaemia protein 1/2
MNase	micrococcal nuclease
MNase-seq	MNase digestion followed by massively parallel sequencing
NMI	non-methylated island (experimentally determined)
NuRD	Nucleosome Remodelling and Deacetylase
OKSM	OCT4, KLF4, SOX2 and c-Myc
PcG	polycomb group protein
PRC1/2	polycomb repressive complex 1/2
qPCR	quantitative real-time PCR
RMCE	recombinase-mediated cassette exchange
RNAPII	RNA polymerase II
RPKM	reads per kilobase per million
SCF	SKP-Cullin-FBox complex
SKP1	S-phase Kinase-associated protein 1
TAM	4-hydroxytamoxifen; tamoxifen
TRP	triptolide (TFIIH inhibitor)
TrxG	trithorax group protein
TSS	transcription start site
ZF-CxxC	zinc-finger CxxC
ZHBTC4	OCT4 doxycycline-sensitive transgene ESC line

CHAPTER 1 - INTRODUCTION

1.1 Regulation of gene expression in eukaryotic cells

1.1.1 General principles of eukaryotic gene regulation

In eukaryotic cells, DNA is packaged around symmetrical histone octamers to form individual nucleosomes and, more broadly, chromatin (Kornberg and Lorch, 1999). Each nucleosome consists of two molecules of histones H2A, H2B, H3 and H4, with a fifth histone, H1, binding linker DNA between nucleosomal units. In addition to providing an important mechanism to compact and organise DNA within the nucleus, the modification of nucleosomes, and chromatin structure in general, has emerged to be an essential component of transcriptional regulation in eukaryotes (Kouzarides, 2007, Li et al., 2007). In its simplest sense, chromatin represents a barrier for sequence-specific transcription factors and the transcriptional machinery to access the DNA sequence (Cockerill, 2011). Cells can overcome this by establishing nucleosome-depleted, more accessible chromatin states by the eviction or repositioning of nucleosomes, often using ATP-dependent chromatin remodelling complexes, which is important for the ability of RNA polymerases to initiate productive transcription from transcription start sites (TSS) (Struhl, 1999, Li et al., 2007, Clapier and Cairns, 2009).

In contrast to prokaryotic systems of gene expression, where the regulatory inputs for transcription are limited to a small, defined area proximal to TSS, the increases in genome size throughout eukaryotic evolution has resulted in the appearance of regulatory elements many kilobases away from their target TSS or promoter (Spitz and Furlong, 2012, Shlyueva et al., 2014). Such distal regulatory elements have been classified as enhancers or silencers of gene expression (Kim et al., 2010, Marsman and Horsfield, 2012, Spitz and Furlong, 2012), or insulators to maintain distinct chromosomal territories

(Gaszner and Felsenfeld, 2006). Although the mechanistic principles of how such distal regulatory elements influence promoter function remain speculative, it appears as if long-range interactions between distal regulatory elements and their target promoters is likely to play an important role in modulating transcriptional activity (Marsman and Horsfield, 2012, Shlyueva et al., 2014). The characterisation of chromatin structure and function of both promoters and distal regulatory elements is therefore essential to understand how eukaryotic cells regulate gene expression.

1.1.2 Histone modifications at gene regulatory elements

An important feature of eukaryotic genomes is the chemical modification of histones to create unique chromatin states at different classes of regulatory elements. Histones can be covalently modified in many different ways, including the addition or removal of methylation, acetylation, phosphorylation, ubiquitylation, sumoylation and ribosylation moieties. The majority of these modifications occur on the exposed histone tails that protrude from the core nucleosomal particle, making them more accessible to other chromatin-modifying and transcriptional factors (Peterson and Laniel, 2004, Kouzarides, 2007). Through variation in the type of chemical group used, the number of modifications on a single histone residue, and the combination of different modifications on a single histone or nucleosome, cells mark specific genomic regions in different ways, which can reflect, or instruct, transcriptional activity at these loci (Li et al., 2007, Voss and Hager, 2014). In many cases, the absence of specific histone modifications can also be important for defining chromatin states at regulatory elements. For example, promoters are enriched for tri-methylation of histone 3 lysine 4 (H3K4me3), with low levels of mono-methylation (H3K4me1) (Barski et al., 2007, Guenther et al., 2007). In contrast, distal regulatory elements, particularly enhancers, have high levels of H3K4me1 and low levels of H3K4me3 (Barski et al., 2007, Heintzman et al., 2009). Histone acetylation is also typically linked with transcriptionally active regulatory elements, while the removal of acetylation by histone deacetylases (HDACs) is often coincident with transcriptional repression

(Clayton et al., 2006). While these histone modifications tend to be limited to regulatory elements, some modifications are widely found throughout eukaryotic genomes, such as H3K36me2 and H3K9me2 (Peters et al., 2003, Robin et al., 2007, Garcia et al., 2008). Alternatively, another set of histone modifications occur co-transcriptionally, such as the deposition of H3K36me3 (Krogan et al., 2003, Tippmann et al., 2012).

While some histone modifications, particularly acetylation, have been suggested to alter chromatin structure directly through disruption of DNA-histone contacts (Anderson et al., 2001, Görisch et al., 2005, Shogren-Knaak et al., 2006), they also appear to play an important role in recruiting additional chromatin-modifying enzymes or transcriptional co-factors to active, or repressed, loci throughout the genome (Fischle et al., 2003a, Taverna et al., 2007). For example, tri-methylation of H3K9 is enriched at inactive, heterochromatic regions of the genome, where it acts to recruit heterochromatin protein 1 (HP1) to establish a transcriptionally repressive chromatin state (Nakayama et al., 2001, Grewal and Jia, 2007). Alternatively, the absence of certain histone modifications can also be important to allow other chromatin-modifying enzymes to bind and modify nucleosomes (Ooi et al., 2007, Schmitges et al., 2011). These general concepts provide an indication of how the deposition and removal of histone modifications allows cells to influence different features of chromatin at regulatory elements, potentially to affect gene expression. For more comprehensive reviews, please see (Fischle et al., 2003a, Kouzarides, 2007, Li et al., 2007, Bartke et al., 2010, Voss and Hager, 2014), while more specific and detailed examples of chromatin modifying pathways will be discussed later.

Importantly, many histone modifications appear to be present throughout the evolutionary diversity of eukaryotes (Ho et al., 2014, Schwaiger et al., 2014), although there has been a great diversification in the mechanisms to target and modify histones. For the sake of simplicity and brevity, I will therefore limit the majority of my discussion to the study of metazoan, particularly mammalian, chromatin and transcriptional regulation.

1.2 Lack of DNA methylation at CpG islands regulates promoter function

In addition to the modification of histones, the majority of CpG dinucleotides in vertebrate genomes are modified on the 5-position of cytosine by the covalent addition of a methyl group (Razin and Cedar, 1977). This CpG methylation is symmetrical, marking the cytosine of both DNA strands (Bird, 1978). The replication of symmetrically methylated CpG dinucleotides therefore results in two strands of hemi-methylated DNA, transferring the methylated state epigenetically to daughter cells during cell division (Holliday and Pugh, 1975, Wigler et al., 1981). This has become an established and important concept to understand epigenetic regulation during development (Reik et al., 2001, Bird, 2002, Smith and Meissner, 2013) and appears to be an essential feature of transcriptional regulatory networks across all vertebrate genomes (Klose and Bird, 2006, Deaton and Bird, 2011, Schubeler, 2015).

However, unlike other chromatin-modifying systems, the presence of DNA methylation systems varies across taxa. While DNA methylation is common amongst plant species (Vanyushin and Ashapkin, 2011), the methylation of invertebrate genomes varies significantly. For example, the model species *Drosophila melanogaster* and *Caenorhabditis elegans* essentially lack CpG DNA methylation (Gowher et al., 2000, Wenzel et al., 2011). Many other invertebrates contain mosaic CpG DNA methylation profiles unlike the relatively uniform genomic methylation of vertebrate genomes, although gene body methylation appears to be conserved (Tweedie et al., 1997, Feng et al., 2010, Zemach et al., 2010, Roberts and Gavery, 2012). The reasons for the evolutionary conservation or divergence in DNA methylation systems are presently unclear, but the conservation of several DNA methylation features from plants to animals, and the presence of alternative DNA methylation systems in prokaryotes, does suggest that DNA methylation is a relatively ancient genomic feature (Suzuki and Bird, 2008, Law and

Jacobsen, 2010). Intriguingly, recent work has identified the presence of modified adenines in the DNA of both *D. melanogaster* and *C. elegans* (Greer et al., 2015, Zhang et al., 2015), suggesting that species with little or no CpG methylation may exploit other mechanisms to achieve similar outcomes.

1.2.1 Cytosine methylation by DNA methyltransferases

DNA methylation is achieved through the transfer of the methyl group from S-adenosylmethionine to cytosine, and this is catalysed by a family of DNA methyltransferases (DNMTs) (Bestor, 2000). DNMT1 was the first to be identified (Gruenbaum et al., 1982), and it has since emerged to play an essential role in the maintenance of DNA methylation. It achieves this through a specific and direct interaction with hemi-methylated DNA at replication forks, with the assistance of PCNA and UHRF1 (Chuang et al., 1997, Bostick et al., 2007, Avvakumov et al., 2008, Hashimoto et al., 2008), where it is responsible for establishing the symmetrical CpG methylation of the hemi-methylated daughter DNA strands (Chuang et al., 1997, Iida et al., 2002). In addition to DNMT1, the other most important DNMTs are DNMT3A and DNMT3B, which are essential for *de novo* methylation of DNA (Okano et al., 1999). DNMT3A/B establish genome-wide DNA methylation patterns at several key transitions during development, most notably after the genome-wide erasure and re-establishment of DNA methylation during early embryogenesis and germ-cell development (Reik et al., 2001), but they are also required for the specific methylation of imprinted genes, the inactive X chromosome and genes targeted for silencing (Bird, 2002, Klose and Bird, 2006). Additionally, a related protein, DNMT3L, lacks DNA methyltransferase activity but is required for normal catalytic activity and genomic targeting of DNMT3A/B (Chédin et al., 2002, Suetake et al., 2004, Ooi et al., 2007). Together the DNMTs are required to maintain normal DNA methylation, which is essential for embryonic development and differentiation (Li et al., 1992, Okano et al., 1999).

1.2.2 Transcriptional repression by DNA methylation

Early characterisation of the DNA methylation system in mammalian cells revealed that methylation of gene promoters was closely linked to transcriptional repression (Mohandas et al., 1981, Harris, 1982, Jaenisch et al., 1985, Antequera et al., 1990, Boyes and Bird, 1992, Hsieh, 1994, Kass et al., 1997). For example, during X inactivation, the process by which once copy of the X chromosome is silenced in female mammalian cells, DNA methylation is essential to maintain silencing of the inactive X chromosome (Mohandas et al., 1981, Lock et al., 1987). Similarly, methylation of tissue-specific gene promoters was found to be coincident with gene repression (Doerfler, 1983), and loss of DNA methylation by chemical inhibitors resulted in reactivation of previously methylated and repressed genes (Harris, 1982, Jaenisch et al., 1985). Furthermore, methylated transgenes inserted into cells fail to be expressed (Stein et al., 1982, Vardimon et al., 1982, Waechter and Baserga, 1982). These early experiments therefore established a functional relevance for DNA methylation in the regulation of transcription, which remains a core principle of chromatin-based gene regulation (Deaton and Bird, 2011, Jones, 2012, Smith and Meissner, 2013).

1.2.3 Distribution of methylated cytosine in vertebrate genomes

The abundance and distribution of CpG dinucleotides throughout vertebrate genomes is not uniform. In fact, the frequency of CpG is much lower than expected by chance (Sved and Bird, 1990, Lander et al., 2001), as a result of the increased deamination rate of methylated cytosine compared to unmodified cytosine (Duncan and Miller, 1980, Shen et al., 1994). This leads to widespread cytosine to thymine transitions throughout vertebrate genomes and a subsequent under-representation of CpG dinucleotides (Sved and Bird, 1990, Arndt et al., 2003, Arndt and Hwa, 2004, Lunter and Hein, 2004). A notable exception to this evolutionary selection against CpG sequences, is that of so-called CpG islands (CGIs), which consist of short, contiguous regions of DNA, typically 1 - 2 kb, with an elevated GC content, and CpG densities approaching predicted CpG frequencies (Bird,

1986, Gardiner-Garden and Frommer, 1987, Jones, 2012). Importantly, the majority of CGIs remain free from methylation and this is likely to be essential to maintain their elevated levels of CpG dinucleotides throughout evolution (Illingworth and Bird, 2009). CGIs are found at the majority (50 to 70%) of vertebrate gene promoters (Saxonov et al., 2006, Long et al., 2013b), and given their refractory nature to DNA methylation and the relationship between DNA methylation and transcriptional repression, CGIs are believed to be an important feature of promoter regulation and function in vertebrates (Blackledge and Klose, 2011, Deaton and Bird, 2011).

1.2.4 Potential mechanisms for lack of DNA methylation at CGIs

Although the majority of CGIs remain free of DNA methylation, regardless of tissue or cell type (Weber et al., 2007, Meissner et al., 2008, Jones, 2012), the mechanisms by which this is achieved remain to be elucidated. The integration of short, CpG-rich DNA sequences into mammalian genomes has revealed that sequence features, most notably CpG density and G+C richness, appear to be important determinants of CGI hypomethylation (Mendenhall et al., 2010, Thomson et al., 2010, Lienert et al., 2011, Krebs et al., 2014, Wachter et al., 2014). Importantly, CpG-rich bacterial DNA sequences, which should lack eukaryotic transcription factor motifs, also remain non-methylated after insertion into mammalian genomes (Mendenhall et al., 2010, Krebs et al., 2014). This reliance upon CpG density appears to be conserved across vertebrates, as insertion of human DNA into mouse, or mouse DNA into zebrafish, also largely recapitulates the original DNA methylation profiles at CpG-rich promoters (Long et al., 2016).

Elevated CpG density may antagonise the DNA methylation machinery in several ways, such as through the recognition and binding of non-methylated CpG dinucleotides by a conserved family of proteins containing ZF-CxxC domains (Long et al., 2013a) (see Section 1.3.2). During germ cell development and early embryogenesis, global DNA methylation is lost (Reik et al., 2001), and the concentration of these ZF-CxxC domain-

containing proteins at CpG-rich sequences could serve to protect them from the methylation machinery during development. For example, several proteins within this family contribute towards H3K4me3 deposition, which has been shown to inhibit DNMT function *in vivo* (Ooi et al., 2007, Hu et al., 2009, Morselli et al., 2015), and deletion of another ZF-CxxC domain-containing protein, KDM2B, results in hypermethylation of a subset of CGIs (Boulard et al., 2015). The conversion of methyl-cytosine to hydroxymethyl-cytosine by TET1-3 has also been proposed to remove DNA methylation from CGIs (Tahiliani et al., 2009, Williams et al., 2011, Wu et al., 2011a, Xu et al., 2011). However, expression of TET proteins and the levels of hydroxymethyl-cytosine are highly variable between tissues (Williams et al., 2012), and DNA methylation in TET-depleted cells is only partially affected (Wu et al., 2011a, Xu et al., 2011), suggesting that this is unlikely to be a central feature of the hypomethylated state at CGIs. In addition to direct antagonism or removal of DNA methylation, several instances have been described where sequence-specific transcription factor binding is sufficient to protect DNA sequences against DNA methylation (Brandeis et al., 1994, Macleod et al., 1994, Palacios et al., 2010, Lienert et al., 2011, Stadler et al., 2011, Feldmann et al., 2013, Domcke et al., 2015). This might be especially relevant for some ubiquitously expressed factors, such as SP1 and CTCF (Macleod et al., 1994, Stadler et al., 2011), as this might explain the relatively invariant hypomethylated state of CGIs across many different cell types and tissues. Although several of these instances are well characterised, the extent to which transcription factor binding limits DNA methylation, particularly at CGIs, remains to be satisfactorily explored.

1.2.5 Non-methylated CGI promoters are transcriptionally active

Mammalian gene promoters can be classified broadly as CpG-rich or CpG-poor, and this is typically reflected by the presence or absence of a CGI. Importantly, and consistent with their refractory nature to DNA methyltransferases, CGI-associated TSS exhibit low DNA methylation levels, while CpG-poor promoters are typically hypermethylated (Saxonov et

al., 2006, Meissner et al., 2008). The coincidence of CGIs and gene promoters (Saxonov et al., 2006, Long et al., 2013b), in addition to their non-methylated state, has established CGIs as important regulatory sequences in the context of gene promoter activity (Deaton and Bird, 2011). Indeed, non-methylated CGIs not located at annotated TSS also exhibit features of transcriptional activity, such as RNA polymerase II (RNAPII) occupancy (Illingworth et al., 2010). CGI promoters exhibit high residency or "pausing" of engaged RNAPII at CGI promoters (Adelman and Lis, 2012, Kellner et al., 2015), and this appears to be a key difference between CGI and non-CGI promoters. The absence of DNA methylation at promoter-associated CGIs appears to be essential for normal transcription, and acquisition of DNA methylation is associated with gene repression (for comprehensive reviews see Bird and Wolffe, 1999, Bird, 2002, Jones, 2012), particularly during cancer and other pathologies (Costello et al., 2000, Robertson and Wolffe, 2000, Portela and Esteller, 2010).

1.3 Interplay between DNA methylation and chromatin modifications

1.3.1 Methyl-binding proteins recognise DNA methylation

Soon after the discovery of DNA methylation, it was proposed that methylated CpGs might be recognised by specific methyl-binding proteins as a means to exert the transcriptionally repressive effects of DNA methylation (Bird, 1986). This was confirmed with the characterisation of methyl-C binding protein (MeCP2) that specifically binds methylated DNA and is important for methylation-dependent gene repression (Meehan et al., 1989, Boyes and Bird, 1991, Lewis et al., 1992, Nan et al., 1997). MeCP2 also acts to link DNA methylation with specific chromatin-modifying activities, due to its association with the SIN3A-containing HDAC complex (Jones et al., 1998, Nan et al., 1998, Lyst et al., 2013) and a potential link with the Suv39h1/2 H3K9 methyltransferases (Lunyak et al., 2002,

Fuks et al., 2003). Mutation of MeCP2 results in Rett syndrome, a severe neurological disorder (Amir et al., 1999). In addition to MeCP2, additional proteins containing methyl-binding domains (MBD) were soon identified, most notably MBD1, MBD2 and MBD4 (Cross et al., 1997, Hendrich and Bird, 1998, Fujita et al., 1999, Ng et al., 1999). Similar to MeCP2, MBD2 interacts with HDACs to link DNA methylation with a hypo-acetylated state (Ng et al., 1999, Ramírez et al., 2012). In addition to MeCP2, MBD1, MBD2 and MBD4, several other transcriptional repressors (Kaiso, ZHBT4, ZHBT38) have been identified to recognise and target methylated CpGs via zinc-finger domains (Prokhortchouk et al., 2001, Yoon et al., 2003, Filion et al., 2006), and there is some evidence that these factors also function in complex with HDACs to remove histone acetylation at methylated loci (Yoon et al., 2003). Together these observations establish the capacity for DNA methylation to influence the distribution of specific histone modifications.

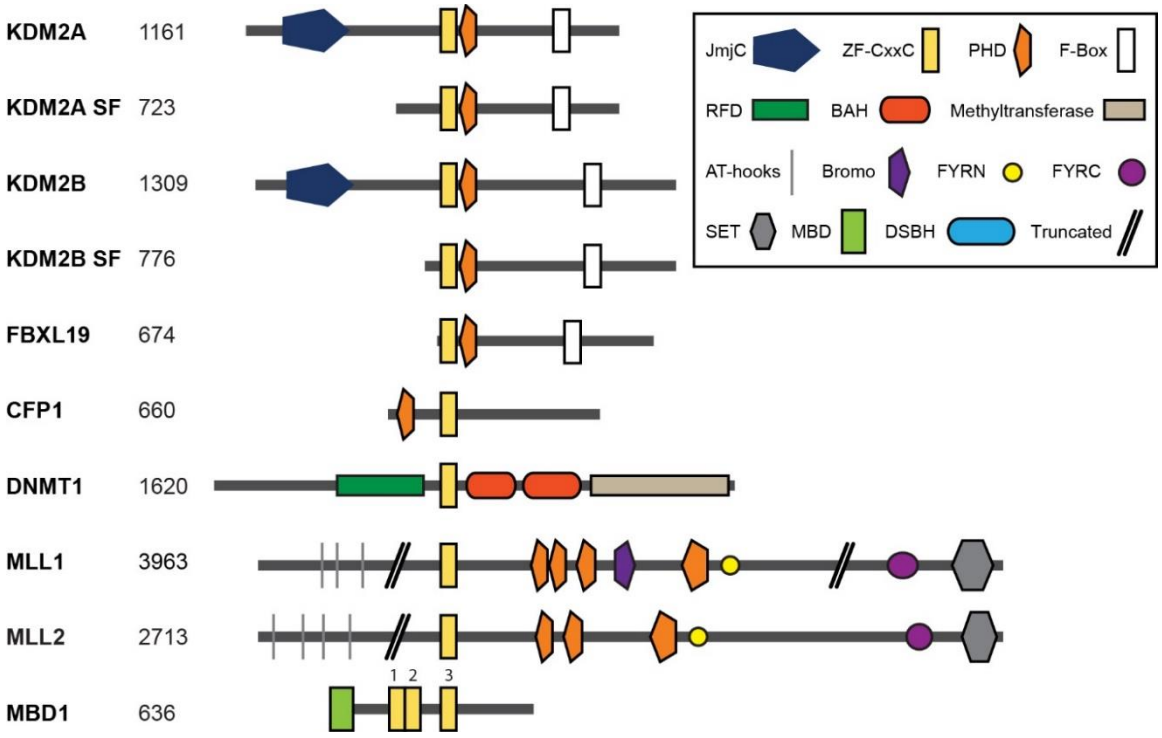


Figure 1.3.1 Type 1 ZF-CxxC domain-containing proteins.
Schematic detailing proteins identified to contain ZF-CxxC domains (Type 1). Reproduced with permission from (Long et al., 2013a).

1.3.2 ZF-CxxC domain-containing proteins bind non-methylated DNA

Non-methylated CpG dinucleotides can also be specifically recognised by various effector proteins to establish specific histone modifications at non-methylated CGIs. This was first described in the case of CxxC finger protein 1 (CFP1) (Voo et al., 2000), which recognises non-methylated DNA through a cysteine-rich zinc-finger CxxC domain (ZF-CxxC). Since the discovery of CFP1, twelve different proteins have been identified to contain ZF-CxxC domains (Figure 1.3.1), including lysine demethylase 2A (KDM2A; JHDM1a/FBXL11) and KDM2B (JHDM1b/FBXL10), mixed lineage leukaemia proteins 1 and 2 (MLL1/2), F-box and leucine-rich repeat protein 19 (FBXL19), DNMT1 and MBD1. The binding of a ZF-CxxC domain to DNA is typically characterized by two cysteine-rich clusters which coordinate with two Zn^{2+} ions in order to position a divergent linker DNA-binding interface into the major groove of the DNA helix (Long et al., 2013a). The divergent linker region of ZF-CxxC domains appears to define specificity of DNA binding, and has been used to classify ZF-CxxC domains into three different subgroups (Type 1 - 3). Of these, Type 1 ZF-CxxC domains demonstrate the highest affinity and specificity for non-methylated DNA both *in vitro* and *in vivo* (Blackledge et al., 2010, Clouaire et al., 2010, Thomson et al., 2010, Frauer et al., 2011, Farcas et al., 2012, Baubec et al., 2013, Long et al., 2013a), with the exception of MBD1 which does not require its ZF-CxxC domain to bind chromatin *in vivo* (Clouaire et al., 2010). As such, work within this thesis shall focus on the roles of different Type 1 ZF-CxxC domain-containing proteins.

1.3.3 CGIs recruit diverse chromatin-modifying activities

Many of the ZF-CxxC domain-containing proteins demonstrate histone-modifying activities, which are recruited to CGIs through their ZF-CxxC domains. Alternatively, other ZF-CxxC domain-containing proteins act to target chromatin-modifying complexes to CGIs. These typically consist of histone methyltransferases or demethylases associated with transcription activation, but also include the recruitment of polycomb repressive

complexes. Together, these combined activities result in a unique chromatin environment at CGIs.

1.3.3.1 KDM2A/B histone demethylases and FBXL19

The paralogues KDM2A and KDM2B (referred to collectively as KDM2A/B) are capable of removing lysine methylation on H3K36 through their Jumonji C (JmjC) demethylase domains, with a preference for H3K36me2 (Tsukada et al., 2006, He et al., 2008, Blackledge et al., 2010). H3K36me2 is highly abundant and is believed to mark 30-50% of total histone H3 in mammalian cells (Peters et al., 2003, Robin et al., 2007, Garcia et al., 2008). The presence of a ZF-CxxC domain in these proteins targets them to non-methylated CGIs genome-wide, where the removal of H3K36 methylation is believed to be important for maintaining a transcriptionally permissive state (Blackledge et al., 2010, Blackledge and Klose, 2011, Liang et al., 2012), and KDM2A or KDM2B-null mice are embryonically lethal (Fukuda et al., 2011, Ishimura et al., 2012, Kawakami et al., 2015). Evidence from yeast has indicated that H3K36me2 acts to repress transcription through the recruitment of the HDAC Rpd3S (Carrozza et al., 2005, Li et al., 2009a) and potentially through suppressing histone exchange (Venkatesh et al., 2012). As such, the removal of H3K36 methylation is proposed to facilitate a transcriptionally permissive chromatin state at CGIs, but evidence for a direct link between H3K36me2 and transcriptional repression is limited in mammalian systems (Sun et al., 2008, Hsia et al., 2010). Nevertheless, the histone demethylase activity of KDM2A or KDM2B appears to be required for normal transcription (Frescas et al., 2008, Pfau et al., 2008, Liang et al., 2012), but it is not clear what other non-enzymatic or structural roles KDM2A/B might contribute towards CGI chromatin and gene regulation. Intriguingly, although KDM2A and KDM2B have very similar domain structures, KDM2B interacts with the polycomb repressive complex 1 (PRC1) through a divergent C-terminal region, and is preferentially enriched over KDM2A at the subset of CGIs which coincides with repressed polycomb target genes (Farcas et al., 2012, He et al., 2013, Wu et al., 2013, Blackledge et al., 2014). The importance of

polycomb repressive systems will be discussed later in detail (Section 1.3.3.3), but for now this observation serves to highlight the potential importance of non-enzymatic roles of these enzymes, as well as regulatory functions unique to KDM2A or KDM2B, which have yet to be fully understood.

KDM2A/B exist in both long and short isoforms (Figure 1.3.1) originating from multiple unique TSS, and the short-form (SF) isoforms exhibit a similar domain organisation to another ZF-CxxC domain-containing protein, FBXL19 (Katoh and Katoh 2004, Long et al., 2013a). KDM2A/B and FBXL19 all interact separately with S-Phase Kinase-Associated Protein 1 (SKP1) (Farcas et al., 2012; Anca Farcas and Emilia Dimitrova unpublished observations, Wong et al., 2016). SKP1 typically interacts with Cullin and F-box-containing proteins to form SKP-Cullin-Fbox (SCF) complexes, which are involved in promoting E3 ligase-mediated ubiquitylation and proteasomal degradation of target proteins, where the F-box subunits provide substrate specificity (Deshaies, 1999, Jin et al., 2004). However, KDM2A/B-SKP1 and FBXL19-SKP1 appear to form atypical complexes, as they lack Cullin, which is normally required to mediate the interaction with E3 ligases and proteasomal machinery (Cardozo and Pagano, 2004). Instead, recent work has identified that KDM2B may be sufficient to act as an E3 ligase itself (Janzer et al., 2012), and preliminary work in the Klose laboratory has suggested that this may also apply to KDM2A and FBXL19 (Emilia Dimitrova). This is in fitting with previous studies implicating FBXL19 in the ubiquitylation of various cytoplasmic proteins (Zhao et al., 2012, Wei et al., 2013, Zhao et al., 2013) or KDM2B in ubiquitylating c-Fos (Han et al., 2016). Ultimately, the contribution of such activities by KDM2A/B and FBXL19 towards chromatin structure and function remain largely unclear.

1.3.3.2 CFP1 and MLL1/2 regulate H3K4me3 deposition

One of the characteristic features of CGI chromatin is the enrichment of H3K4me3 (Illingworth et al., 2010, Blackledge and Klose, 2011). Although typically associated with

transcription, both active and repressed CGIs are enriched for this histone modification (Guenther et al., 2007, Mikkelsen et al., 2007). This is driven by targeting of H3K4me3 methyltransferase activities by the ZF-CxxC domain-containing proteins MLL1, MLL2 and CFP1.

The first discovered ZF-CxxC domain-containing protein, CFP1 is a highly conserved component of SETD1 complexes (Lee and Skalnik, 2005, Lee et al., 2007a), which is one of the major H3K4me2/3 methyltransferase complexes in eukaryotic cells (Ardehali et al., 2011, Hallson et al., 2012, Bledau et al., 2014). SETD1 complexes typically consist of SETD1A or SETD1B, RbBP5 (retinoblastoma-binding protein 5), ASH2L (absent, small or homeotic 2-like), DPY-30 (dosage compensation protein 30), WDR (WD40 repeat) 5 and WDR82 (Lee and Skalnik, 2005, Lee et al., 2007a, Ruthenburg et al., 2007, Ernst and Vakoc, 2012). Through its ZF-CxxC domain CFP1 appears to recruit the SET1 complex to CGI promoters to maintain normal H3K4me3 levels (Thomson et al., 2010, Clouaire et al., 2012). CFP1 may also be targeted through alternative mechanisms, such as recognition of pre-existing H3K4me3 by its PHD domain (Eberl et al., 2013) or the co-transcriptional recruitment of CFP1/SETD1 by RNA polymerase II (Ng et al., 2003, Lee and Skalnik, 2008). While CFP1 occupies the majority of CGI promoters, it is largely absent from repressed or non-transcribed CGIs (Thomson et al., 2010; David Brown unpublished observations). Intriguingly, the presence of CFP1, and its ZF-CxxC domain, in the SETD1 complex may serve to restrict SETD1 chromatin occupancy or activity to CGI promoters (Tate et al., 2010), as loss of CFP1 or its ZF-CxxC domain results in widespread gains of ectopic H3K4me3 at distal regulatory elements (Clouaire et al., 2012; David Brown unpublished observations). Importantly, CFP1 is essential for early mouse development (Carlone and Skalnik, 2001), and CFP1-null ESCs fail to effectively differentiate *in vitro* (Carlone et al., 2005)**,** although the precise mechanisms by which it regulates CGI-associated gene expression are unknown.

The MLL family of H3K4 methyltransferases is composed of four different members (MLL1-4), which form large protein complexes with many of the same subunits as CFP1/SETD1 complexes (eg. ASH2L, WDR5, RbBP5 and DPY-30) (Ruthenburg et al., 2007). Only MLL1 and MLL2 contain ZF-CxxC domains, which bind non-methylated DNA *in vitro* (Birke et al., 2002, Bach et al., 2009), and occupy CGIs *in vivo*. MLL1 chromatin occupancy appears to be restricted to a subset of transcriptionally active CGIs (Guenther et al., 2005, Milne et al., 2005, Scacheri et al., 2006), similarly to CFP1. In contrast, recent genome-wide profiling of MLL2 in ESCs has revealed that it binds the majority of CGIs including repressed, bivalent CGI promoters (Denissov et al., 2014). In addition to their ZF-CxxC domains, other chromatin-binding modules of MLL1/2 and an interaction with RNAPII or associated factors are likely to play important roles in the genomic distribution of these complexes (Milne et al., 2005, Chang et al., 2010, Milne et al., 2010, Muntean et al., 2010). Importantly, MLL proteins are frequently subject to chromosomal translocations during leukaemogenesis (Meyer et al., 2009), and a wide variety of MLL-fusion proteins can have a range of epigenetic and transcriptional effects on cancer progression (Mohan et al., 2010, Ballabio and Milne, 2012). For example, the fusion of MLL1 with AF9 (ALL1-fused gene from chromosome 9), a component of the super-elongation complex, results in aberrant targeting of the super-elongation complex to normally silent MLL target genes (Mueller et al., 2009, Mohan et al., 2010). In general, there appears to a large degree of redundancy between MLL1 and MLL2 in their regulation of H3K4me3, with the exception of a unique role for MLL2 at bivalent or repressed CGIs (Denissov et al., 2014). However, MLL1 appears to play an essential role in haematogenesis (Yagi et al., 1998, Ernst et al., 2004), and MLL2, although essential during early mouse development (Glaser et al., 2006) is not required after embryonic day 11.5 (Glaser et al., 2009), highlighting the likelihood of unique functions for these two complexes in the regulation of CGI promoter activity.

1.3.3.3 Recruitment of polycomb repressive complexes

While non-methylated CGIs appear to be important for maintaining normal transcriptional activity of gene promoters through the recruitment of H3K4 methyltransferases and H3K36 demethylases, they can also act to recruit repressive chromatin-modifying complexes (Blackledge et al., 2015). In particular, KDM2B can form a unique version of the polycomb repressive complex 1 (PRC1), specifically with PRC1 complexes containing the non-canonical polycomb group ring finger 1 (PCGF1) (Farcas et al., 2012, He et al., 2013, Wu et al., 2013). Through its interaction with KDM2B, PCGF1-PRC1 is recruited to non-methylated CGIs, where its ability to ubiquitylate histone H2A on lysine 119 (H2AK119ub1) can lead to the recruitment of the second polycomb repressive complex, PRC2, which methylates histone H3 lysine 27 (H3K27me3) (Mendenhall et al., 2010, Blackledge et al., 2014, Cooper et al., 2014, Jermann et al., 2014). Targeting of PRC1 by KDM2B contributes to 10–20% of CGIs being occupied by PRC1/PRC2 in ESCs (Blackledge et al., 2014), often in a bivalent state where they are also occupied and modified with factors and histone modifications associated with gene activation (Mikkelsen et al., 2007, Voigt et al., 2013). Nevertheless, this occupancy of polycomb at CGI gene promoters is typically associated with reduced transcriptional activity of developmentally important genes (Lewis, 1978, Schwartz and Pirrotta, 2013). Indeed, deletion of polycomb components result in the reactivation of many polycomb-occupied promoters (Boyer et al., 2006, Endoh et al., 2008, Leeb et al., 2010, Blackledge et al., 2014). However, the precise mechanisms by which they achieve this transcriptional repression are poorly understood (Di Croce and Helin, 2013).

1.3.4 CGIs function as unique regulatory platforms at promoters

A unique feature of CGI promoters is their capacity to be targeted by chromatin-modifying pathways associated with both transcriptional activity and repression. Indeed, several of the CGI-binding ZF-CxxC domain-containing proteins are homologues of Trithorax group (TrxG) proteins, initially identified in *Drosophila*, which antagonise Polycomb group (PcG)

proteins to determine body axis specification during embryogenesis (for comprehensive reviews see Kennison, 1995, Simon and Tamkun, 2002, Schuettengruber et al., 2007). In *Drosophila* TrxG proteins are recruited to TrxG response elements (TREs) to promote transcriptional activity, while PcG proteins homologous to PRC1 and PRC2 recognise specific PcG response elements (PREs) to repress transcription. In vertebrates, these targeting systems seem to have been replaced with CGI-centred systems for recruitment of TrxG/PcG factors. Although the causes and consequences of the evolutionary divergence of TrxG/PcG genomic targeting mechanisms remains unclear, the ability for CGIs to recruit these activating and repressive factors has important implications for understanding promoter activity in vertebrate genomes.

1.3.4.1 Antagonism between TrxG and PcG systems

Classically, TrxG complexes are often associated with H3K4me3 methyltransferase (such as CFP1 and MLL1/2), acetyltransferase, or chromatin remodelling (NuRF, SWI/SNF) activities, which serve to antagonise PcG-dependent activities. This antagonism can occur via direct, or indirect, mechanisms. For example, the catalytic activity of PRC2 is allosterically inhibited by the presence of H3K4me3 (Schmitges et al., 2011) and acetylation of H3K27 prevents methylation of this residue by PRC2 (Tie et al., 2009, Reynolds et al., 2012). This may further be reinforced by recruitment of H3K27me2/3 demethylases by TrxG proteins (Lee et al., 2007b) or the presence of deubiquitylases at CGI promoters (Yang et al., 2014b). Conversely, the presence of H3K27me3 has been suggested to limit SETD1 complexes from interacting with and modifying histone H3 (Kim et al., 2013a). These mechanisms therefore serve to limit the activity, occupancy or function of the opposing regulatory system in order to limit inappropriate or spurious activity of these chromatin-modifying complexes at gene promoters.

1.3.4.2 Feedback mechanisms to maintain TrxG/ PcG chromatin states

In addition to antagonism between TrxG- and PcG-associated chromatin-modifying pathways, both systems appear to demonstrate positive feedback mechanisms to support and establish TrxG- or PcG-dominant states. For example, PRC1 complexes containing chromobox (CBX) subunits are able to recognise and bind H3K27me3, which is placed by PRC2, in order to recognise and occupy PcG target sites (Min et al., 2003, Wang et al., 2004). Similarly, PRC2 complexes can recognize, and are stimulated by, PRC1-dependent H2AK119ub1 (Blackledge et al., 2014, Cooper et al., 2014, Kalb et al., 2014). Indeed, these feedback mechanisms have been proposed to underpin chromatin-based communication between PRC1 and PRC2 to establish and maintain PcG repressive chromatin domains (Voigt et al., 2013, Steffen and Ringrose, 2014, Blackledge et al., 2015). For TrxG complexes, ZF-CxxC domain-dependent H3K4me3 methyltransferase occupancy at CGIs can be amplified by RNAPII-dependent recruitment of additional methyltransferase complexes (Ng et al., 2003, Lee and Skalnik, 2008). Furthermore, CFP1 and MLL1/2 are able to recognise and bind H3K4me3 (Chang et al., 2010, Milne et al., 2010, Eberl et al., 2013), establishing a positive feedback loop between activity and occupancy of these complexes at CGIs. Through these feedback mechanisms, TrxG- or PcG-dominant chromatin states can be established and maintained at individual CGIs.

1.3.4.3 Bistability at CGIs for stable, but responsive, transcriptional states

One of the proposed advantages of such a system is that it represents a bistable chromatin switch (Klose et al., 2013). This bistable switch consists of TrxG- or PcG-dominant chromatin states at CGIs, which, once established, are stably reinforced and maintained. At the same time, CGIs are targeted by the alternative system, in order to constantly challenge the established state. However, due to the antagonism between TrxG and PcG complexes, the established state can only be overcome with a concerted increase in the targeting, occupancy or activity of the opposing system. For example, PcG-

occupied CGIs are constantly sampled by TrxG proteins, but only upon the addition of a strong activating signal, such as transcription factors or increased RNAPII recruitment, can they overcome the established PcG state. Following this initial change in equilibrium, feedback mechanisms such as the co-transcriptional reinforcement of H3K4me3 methyltransferases, or the inhibition of PcG complex activity, allow the stable maintenance of this new TrxG state. Alternatively, if the activating signal is not maintained, reduced transcription could increase PcG complex occupancy (Riising et al., 2014), overcoming the TrxG-dominant state to support and maintain an inactive PcG chromatin state through PRC1/PRC2 feedback mechanisms. In addition to providing stable and heritable epigenetic states (ie across cell divisions), this model provides a flexibility to quickly and robustly switch chromatin states in response to increases or decreases in activating or repressive signals, ensuring a rapid alteration of gene expression profiles to environmental or developmental cues. Indeed, gene expression analyses following stimulation of macrophages by lipopolysaccharide (LPS) has revealed that not only do CGI promoters tend to be more rapidly activated (Ramirez-Carrozzi et al., 2009), but that LPS-responsive CGI promoters lose this activation more quickly (Bhatt et al., 2012). Although it is unclear how widely such principles might yet apply (Smale et al., 2013), the proposed bistable state at CGI promoters seems to provide a stable yet highly responsive system to explain the regulation of gene expression by CGIs.

1.4 Chromatin accessibility at CGIs and other regulatory elements

By its very nature, chromatin is a fundamental barrier for transcription factors and polymerases to access DNA sequences in eukaryotic cells. In particular, nucleosome positioning and occupancy at promoters can have important implications for both the initiation and elongation of RNA polymerases (Kornberg and Lorch, 1999, Kireeva et al., 2005, Jiang and Pugh, 2009). The study of nucleosome positioning at eukaryotic gene promoters from yeast to metazoans has revealed that many promoters contain nucleosome free regions that represent a relatively open chromatin state that may facilitate sequence-specific transcription factors to access DNA sequences and promote transcription (Jiang and Pugh, 2009). An appreciation for the importance of such accessibility of the DNA sequence, commonly referred to as chromatin accessibility, has emerged in recent years as a feature of chromatin-based regulatory pathways (Cockerill, 2011, Tsompana and Buck, 2014, Voss and Hager, 2014). Here I will discuss different definitions and measurements for chromatin accessibility, the relevance of chromatin accessibility towards CGI promoters, and the requirement for chromatin remodellers and transcription factors to modulate DNA access at regulatory elements.

1.4.1 Definitions and measurements of chromatin accessibility

The term chromatin accessibility typically reflects the ability of *trans*-acting factors to interact with specific DNA sequences in the context of nucleosomes (Cockerill, 2011). Defined as such, chromatin accessibility is referred to here as measure of locus-specific chromatin structures, and not, as the term has been used for previously, as a measure of global nuclear condensation or compaction. Throughout the chromatin literature, open chromatin states are often inferred by the presence, or absence, of features correlated with accessible or inaccessible chromatin, such as H3K4me3 or HP1 respectively (Gaspar-Maia et al., 2009, Wan et al., 2013). Importantly, here I focus instead upon either

the direct measurements of chromatin accessibility by nuclease sensitivity, or the indirect measurement of chromatin accessibility through nucleosome occupancy.

1.4.1.1 Nuclease hypersensitivity as a direct measure of chromatin accessibility

Although several alternative approaches for measuring chromatin accessibility have since been developed, initial observations of chromatin accessibility at gene promoters *in vivo* were made using nuclease sensitivity assays. In particular, the digestion of nuclei with limiting amounts of deoxyribonuclease (DNase) I revealed that active gene promoters were “hypersensitive” to endonuclease digestion, while inactive regions of the genome, such as gene bodies or intergenic regions, were relatively insensitive to digestion (Weintraub and Groudine, 1976, Wu, 1980). Although the question of DNase I digestion bias remains contentious (Dingwall et al., 1981, Koohy et al., 2013, Lazarovici et al., 2013, He et al., 2014), DNase I is able to endolytically cleave DNA strands in a generally non-specific manner, but with a preference for regions of open and accessible chromatin. It has been established that *de novo* identification of such open chromatin regions is an excellent indication of the presence of regulatory elements, such as enhancers, insulators or silencers using DNase I hypersensitivity approaches (Higgs et al., 1990, Chung et al., 1993). In addition to DNase I, studies have also used other endonucleases, particularly restriction enzymes, which in contrast to DNase I, recognise specific DNA sequences to interrogate chromatin accessibility at defined loci (Keshet et al., 1986, Tazi and Bird, 1990, Boyes and Felsenfeld, 1996). With the advent of next-generation sequencing, DNase I sensitivity was the first chromatin accessibility assay to be coupled with massively parallel sequencing (DNase-seq) (Sabo et al., 2004), and although restriction enzyme sensitivity approaches have also been applied (Gargiulo et al., 2009, Chen et al., 2014b), they are only able to interrogate chromatin structure at loci containing their specific sequences. Analysis of paired-end DNase-seq libraries has revealed that DNase-seq provides not only information related to nucleosome-depleted DNA, but also cleaves DNA positioned

on top of nucleosomes (Vierstra et al., 2014). Although, DNase-seq has remained the most widely used approach to measure chromatin accessibility at the genome-scale, a novel approach called assay for transposase-accessible chromatin (ATAC) has emerged recently (Buenrostro et al., 2013, Buenrostro et al., 2015). This approach, similarly to DNase-seq, relies upon sensitivity of DNA to digestion by the Tn5 transposase and generates highly similar chromatin accessibility profiles to DNase-seq. It provides several important advantages to DNase-seq (see Chapter 3 for more details), and is likely to replace DNase-seq as the most widely used chromatin accessibility assay in coming years.

1.4.1.2 Measuring nucleosome occupancy using MNase-seq

An alternative approach to measure chromatin accessibility is the use of micrococcal nuclease (MNase) to measure nucleosome occupancy and positioning. MNase endolytically cleaves DNA, but unlike DNase I digestions, it is unable to cleave nucleosomal DNA (Schones et al., 2008). MNase-protected DNA sequences therefore represent positions occupied by nucleosomes, and this approach has been widely coupled with genome wide approaches (MNase-seq) to map nucleosome positions in various cell types (Yuan et al., 2005, Johnson et al., 2006, Lee et al., 2007c, Schones et al., 2008, Weiner et al., 2010, Hu et al., 2011, Li et al., 2011b, Teif et al., 2012). Although the precise determinants of nucleosome positioning is the subject of much debate, MNase-seq has revealed a nucleosome-depleted, or nucleosome-free region at TSS across all organisms examined, and this has been used to represent an elevated level of chromatin accessibility, due to the absence of nucleosomes. However, many of these MNase-seq experiments have been performed with an excessive digestion of nucleosomes, and subsequent experiments with limiting concentrations of MNase have revealed the presence of an unstable, or fragile, nucleosome at the previously described nucleosome-depleted region (Weiner et al., 2010, Xi et al., 2011, Deng et al., 2013, Mieczkowski et al., 2016). This has partly been attributed to the presence of histone

variants like H2AZ and H3.3 at gene promoters which are more sensitive to salt concentrations and MNase digestion (Jin and Felsenfeld, 2007, Jin et al., 2009). While this clearly still corresponds to some level of chromatin accessibility, as measured by DNase I digestion or other approaches, these observations represent an important clarification of chromatin structure at gene promoters. Therefore, while MNase-related approaches can provide useful information regarding nucleosome occupancy, positioning and phasing, it is essential to understand the digestion dynamics of a specific experiment before drawing any conclusions.

1.4.1.3 Alternative chromatin accessibility protocols

Although DNase-seq and MNase-seq represent the two most widely used measures of chromatin accessibility, several other methods have been developed based upon alternative principles to nuclease digestion. For example, formaldehyde assisted isolation of regulatory elements (FAIRE)-seq relies upon the biochemical fractionation of DNA sequences which lack protein-DNA crosslinks following formaldehyde fixation and sonication, and thus removes nucleosome-bound DNA to enrich for nucleosome-free DNA sequences (Giresi et al., 2007, Song et al., 2011, Simon et al., 2012). Similarly, the propensity for open, accessible chromatin to sonicate more efficiently has been exploited in the form of Sono-Seq (Auerbach et al., 2009). Alternatively, chemical modification of nucleosomal DNA (Brogaard et al., 2012) or chemical cleavage of linker DNA between nucleosomes (Ishii et al., 2015) have also been used to map nucleosome positions. Although these assays tend to identify broadly similar accessible chromatin regions or nucleosome occupancy (Song et al., 2011, Ishii et al., 2015), the different modes of action also result in assay-specific effects, which are important to appreciate when considering the causes and consequences of chromatin accessibility.

1.4.1.4 Chromatin accessibility as altered nucleosome occupancy or structure

The interpretation of chromatin accessibility data clearly requires an appreciation of the different methods used to interrogate chromatin structure. Ultimately, these open chromatin structures are likely the result of altered nucleosome occupancy, perturbed DNA-histone interactions, the presence of histone variants and/or increased rates of histone/nucleosome turnover. Certainly, the presence or absence of nucleosomes at regulatory elements acts as the major barrier to the accessibility of DNA sequences, although intra-nucleosomal cleavage by DNase I suggests that nucleosomal DNA can still be partially accessible to cleavage, particularly at the exposed face of the minor groove on the nucleosome (Boyle et al., 2008, Winter et al., 2013, Vierstra et al., 2014). This accessibility appears to be elevated at nucleosomes at regulatory elements like promoters or enhancers, perhaps determined by the presence of destabilising histone variants (Jin and Felsenfeld, 2007, Jin et al., 2009, Hu et al., 2013) or histone modifications, like acetylation (Zentner and Henikoff, 2013). For example, histone acetylation can act to neutralize the positive charge of lysine residues in histones, potentially weakening histone-DNA contacts (Hong et al., 1993, Dion et al., 2005, Shogren-Knaak et al., 2006). This could also be an inherent feature of different DNA sequences, as has been proposed for homopolymeric sequences, such as poly(dA:dT) and poly(dG:dC) rich sequences, which are intrinsically stiff and unfavourable for nucleosome formation (McCall et al., 1985, Suter et al., 2000, Anderson and Widom, 2001, Segal and Widom, 2009). Transcription factor binding, or interaction of other *trans*-acting factors, could also alter the thermal properties of DNA on the nucleosome surface (Polach and Widom, 1996, Anderson and Widom, 2000, Sekinger et al., 2005, Albert et al., 2007, Jiang and Pugh, 2009), potentially affecting DNA-histone interactions sufficiently to facilitate nuclease sensitivity. Alternatively, the presence of H1 likely plays a role in compacting the chromatin fibre between nucleosomes by binding linker DNA sequences, and this can limit the detection of inter-nucleosomal DNA accessibility (Kamakaka and Thomas, 1990, Garrard, 1991,

Woodcock et al., 2006). While these proposed alterations to nucleosome and chromatin structure may all partially contribute to chromatin accessibility measured by nuclease sensitivity, the precise molecular features of open chromatin are likely to vary from loci to loci, and require further biochemical and genomic characterisation.

1.4.2 Chromatin structure and accessibility of CGI promoters

Soon after their discovery, it was observed that CGIs, like other active regulatory elements, exist in an open and accessible chromatin state (Weintraub and Groudine, 1976, Stalder et al., 1980, Antequera et al., 1989, Tazi and Bird, 1990). This elevated chromatin accessibility has been proposed to be important to allow CGIs to act as regulatory platforms to facilitate transcription factors and polymerases to engage with them and regulate gene expression (Blackledge and Klose, 2011, Deaton and Bird, 2011). However, the molecular determinants of this elevated chromatin accessibility have attracted little interest.

1.4.2.1 Non-methylated CGIs are hypersensitive to nucleases

Several observations have established that non-methylated DNA and CGIs are associated with elevated chromatin accessibility. Initially, single gene studies identified that CGI promoters were more susceptible to digestion by nucleases (Weintraub and Groudine, 1976, Stalder et al., 1980). Subsequent work examining the properties of chromatin released by methylation sensitive restriction endonucleases (ie fraction of chromatin associated with non-methylated DNA) confirmed that CGIs were relatively depleted of nucleosomes and enriched for acetylated histones (Tazi and Bird, 1990). Furthermore, use of a methylation-insensitive enzyme MspI was unable to efficiently digest methylated chromatin (Antequera et al., 1989), consistent with DNA methylation being associated with reduced chromatin accessibility. These observations have been extended by genome-wide chromatin accessibility, nucleosome occupancy and DNA methylation profiling (Choi, 2010, Stadler et al., 2011, Fenouil et al., 2012, Thurman et al.,

2012), confirming that hypo-methylated CGIs are well correlated with elevated chromatin accessibility and reduced nucleosome occupancy. Whether there are CGI-specific mechanisms which achieve this open and accessible chromatin state is presently unclear, although it has been suggested that CpG-rich or GC-rich DNA sequences may be inherently more accessible (Ramirez-Carrozzi et al., 2009), potentially through an enrichment of poly(dG:dC) rich sequences, which may be unfavourable to nucleosome formation (McCall et al., 1985, Segal and Widom, 2009).

1.4.2.2 Polycomb-occupied CGIs have limited chromatin accessibility

In contrast to the majority of CGIs, targeting of PcG complexes to CGIs is associated with a more compact chromatin state (Di Croce and Helin, 2013), suggesting that the accessibility of CGIs may represent an under-appreciated aspect of TrxG-PcG antagonism. PcG-dependent alteration of chromatin structure was initially observed *in vitro*, where it was demonstrated that the *Drosophila* PRC1 component Psc (mammalian homolog PCGF4) was sufficient to compact nucleosome arrays, as evidenced by electron microscopy (Francis et al., 2004, Grau et al., 2011). This relationship between PcG components and chromatin compaction has since been supported by several different lines of evidence, including altered chromatin domain (~Mb) structures (Boettiger et al., 2016), PcG-dependent long-range chromatin interactions (Eskeland et al., 2010, Schoenfelder et al., 2015), reduced nuclease accessibility profiling in wild type cells (Bell et al., 2010, Calabrese et al., 2012, Beck et al., 2014, Deaton et al., 2016), and reduced capacity for *trans*-acting factors to bind to PcG targets (Zink and Paro, 1995, McCall and Bender, 1996, Fitzgerald and Bender, 2001). Intriguingly, this refractory nature at PcG targets appears to be related to the size of the *trans*-acting factors involved, as a tetrameric “goliath” T7 RNA polymerase complex was more sensitive to PcG-mediated repression than monomeric T7 RNA polymerase (Fitzgerald and Bender, 2001). The compact PcG chromatin state at PcG targets also appears to be important for enhancing the activity of PRC2 in depositing H3K27me3 (Yuan et al., 2012), but its relevance towards

PcG-mediated gene repression at CGIs remains unclear (Simon and Kingston, 2009). Furthermore, the precise mechanisms by which chromatin accessibility is repressed or constrained at PcG-occupied CGIs has yet to be addressed *in vivo*. Indeed, given the normally hypersensitive state and transcriptionally active state of CGIs, it will be interesting and important to consider whether regulation of the chromatin accessibility at CGIS may reflect an underappreciated aspect of the TrxG/PcG antagonism in the context of gene regulation.

1.5 Transcription factors and the chromatin barrier

Sequence-specific transcription factors read DNA-encoded information to control gene transcription and define the complement of proteins a cell can produce (Spitz and Furlong, 2012, Voss and Hager, 2014). They achieve this by physically interrogating and binding specific DNA sequence motifs at promoters, enhancers, silencers or other regulatory elements, and ultimately control how RNA polymerases engage with and transcribe genes. In higher eukaryotes, transcription factors recognize short DNA sequences and often rely on combinatorial binding with other DNA-binding transcription factors to achieve affinity and specificity in highly complex genomes (Wunderlich and Mirny, 2009, Villar et al., 2014). In addition to the widespread requirement for combinatorial DNA binding, eukaryotic transcription factors are confronted with the added complexity that eukaryotic DNA is wrapped around histones to form nucleosomes and chromatin. The association of DNA with nucleosomes can therefore compete with DNA-binding transcription factors, limiting their ability to engage with the genome (Henikoff, 2008). In response to chromatin acting as a barrier to DNA recognition, eukaryotic transcription factors have evolved a series of compensatory mechanisms.

1.5.1 ATP-dependent chromatin remodelling complexes

One way the cell has evolved to deal with this competition between nucleosomes and transcription factors is to exploit the activity of large ATP-dependent chromatin remodelling complexes to mobilize nucleosomes and regulate access to DNA (Clapier and Cairns, 2009). This large family of enzymes is defined by the presence of a highly conserved helicase-like ATPase domain, which can act to reposition, unwrap or evict nucleosomes from their DNA template (Clapier and Cairns, 2009, Narlikar et al., 2013). The four main classes of chromatin remodelling complexes are characterised by their different catalytic subunits including the SWI/SNF, ISWI, CHD and INO80 families, and these have been extensively studied from yeast to human cells (for comprehensive reviews see Corona and Tamkun, 2004, Mohrmann and Verrijzer, 2005, Bao and Shen, 2007, Marfella and Imbalzano, 2007). Although the specific modes of action vary significantly, chromatin remodelling complexes generally share several common features. In addition to the DNA-dependent ATPase domain, which acts a translocating motor to disrupt histone-DNA interactions, chromatin remodellers typically exhibit a high affinity for nucleosomes in general, but also contain domains for the recognition of specific histone modifications (Clapier and Cairns, 2009). For example, the CHD family of remodellers contain multiple chromodomains (CHD), some of which have been characterised to bind H3K4me3 (Pray-Grant et al., 2005, Sims et al., 2005), while SWI/SNF remodellers often contain bromodomains to recognise acetylated histone (Marmorstein and Berger, 2001, Hassan et al., 2002). Many chromatin remodellers form complexes with other chromatin-associated factors, as demonstrated by the inclusion of CHD3/4 in the NuRD complex, which includes histone deacetylase subunits (Xue et al., 1998).

Either through direct nucleosome repositioning or through activity of associated factors, chromatin remodellers are important for the normal regulation of transcription at gene promoters. For example, CHD3/4-dependent remodelling by the NuRD complex is associated with gene repression (Linder et al., 2007, Musselman et al., 2012, Ramírez et

al., 2012), but other chromatin remodellers have been connected with the activation of transcriptional initiation or elongation (Armstrong et al., 2002, Simic et al., 2003, Skene et al., 2014). One of the ways chromatin remodellers can achieve this is by facilitating transcription factor binding, and the cooperation between transcription factors and chromatin remodelling complexes has attracted considerable attention recently (Wang et al., 1996, Olave et al., 2002, Lickert et al., 2004, Wu et al., 2007, Ho et al., 2009, Morris et al., 2014, Wang et al., 2014, Ceballos-Chávez et al., 2015, Laurette et al., 2015). Indeed, chromatin remodelling complexes are enriched at both promoters and distal regulatory elements, where they can be essential to maintain normal chromatin structure and accessibility (Morris et al., 2014, de Dieuleveult et al., 2016). However, the mechanisms by which remodellers are recruited to such elements still requires significant work, particularly with respect to the contribution of chromatin-binding domains and interactions with sequence-specific transcription factors.

1.5.2 Pioneer transcription factors to bind inaccessible chromatin

The pre-existing chromatin accessibility or nucleosome depletion at regulatory elements, potentially established by chromatin remodelling complexes, is required for the binding of many transcription factors to their cognate binding sites (Guccione et al., 2006, Lupien et al., 2008, Guertin and Lis, 2010, Biddie et al., 2011a, John et al., 2011, Wu et al., 2011b). This can be advantageous in limiting binding of transcription factors to active regulatory elements, rather than the many thousands of spurious binding sites that exist throughout the genome. This might help to allow rapid changes in gene expression by transcription factors in response to environmental stimuli (Smale et al., 2013), such as hormone treatments (John et al., 2011) or inflammatory stimuli (Ramirez-Carrozzi et al., 2009, Bhatt et al., 2012). However, this requirement for pre-existing chromatin accessibility poses a challenge for cells when they need to transition into new cell types during development and establish new regulatory elements in normally inaccessible and nucleosome-occupied chromatin.

To achieve this, cells have evolved a specialised set of transcription factors that are able to recognise and bind their DNA motif sequences in spite of the presence of nucleosomes. These factors are referred to as pioneer transcription factors, due to the fact that they are able to "pioneer" and establish new transcription factor binding events and regulatory elements in previously inaccessible chromatin (Magnani et al., 2011, Zaret and Carroll, 2011). Such activities were first suggested when it was observed that some transcription factors (eg. FoxA and GATA4) were able to bind nucleosomal DNA *in vitro* in the absence of chromatin remodellers (Cirillo et al., 2002) and genome-wide approaches have established that putative pioneers are able to bind nucleosome-occupied motif sites (Soufi et al., 2012, Raposo et al., 2015, Schulz et al., 2015, Soufi et al., 2015). Binding of pioneers to their motifs is subsequently associated with the establishment of chromatin accessibility (Raposo et al., 2015, Schulz et al., 2015), which is likely to be essential for the recruitment of non-pioneer transcription factors (Wapinski et al., 2013, Foo et al., 2014, Xu et al., 2014, Schulz et al., 2015, Iwafuchi-Doi et al., 2016). Pioneers are typically master regulatory factors required for developmental transitions or reprogramming of cells (Lupien et al., 2008, Soufi et al., 2012, Wapinski et al., 2013, Iwafuchi-Doi and Zaret, 2014). For example, OCT4 is one of the four Yamanaka transcription factors used to reprogram adult somatic cells into induced pluripotent stem cells (iPSCs) (Jerabek et al., 2014), and it appears to achieve this by recognising and binding DNA motif sequences positioned on top of nucleosomes (Soufi et al., 2012, Soufi et al., 2015). The ability to bind inaccessible chromatin and to transition it into an accessible state has emerged to be essential feature of pioneer factors (Iwafuchi-Doi and Zaret, 2014), and has led to an increased classification of transcription factors as putative pioneers. However, careful *in vivo* measurements of their precise contributions towards chromatin accessibility, transcription factor co-binding and gene regulation have yet to be performed for the majority of these putative pioneer factors.

Several modes of pioneer action have been suggested by which pioneer transcription factors can displace nucleosomes or alter chromatin structure at *de novo* binding sites during reprogramming or development (Voss and Hager, 2014). Firstly, ATP-dependent chromatin remodellers or chaperones can assist pioneers to reposition or evict nucleosomes, or to exchange histone subunits for more unstable histone variants in order to facilitate chromatin accessibility and transcription factor binding (Henikoff, 2008, Clapier and Cairns, 2009, Hargreaves and Crabtree, 2011). This can be part of an "assisted loading" model, whereby transcription factors recruit remodellers in a transient manner, resulting in short-lived chromatin conformations to provide transient opportunities for other transcription factor molecules to recognise and bind DNA (Voss et al., 2011, Voss and Hager, 2014). Another proposed model is that of "nucleosome attack", where a combination of transcription factors cooperatively bind to nucleosomal DNA to generate sufficient free energy to counteract and destabilise histone-DNA contacts within the nucleosome in a non-specific manner (Miller and Widom, 2003, Mirny, 2010). Alternatively, pioneer factors may have unique features which interact specifically with nucleosomes to alter their properties. For example, FOXA and FOXO proteins structurally resemble histone H1, which may lead to displacement of H1 and destabilisation of neighbouring nucleosomes (Cirillo et al., 2002, Hatta and Cirillo, 2007). In a similar manner, the trimeric NF-Y transcription factor contains two subunits with histone fold domains which could compete with nucleosomal histones during NF-Y binding to facilitate an altered chromatin structure and the binding of other transcription factors (Nardini et al., 2013, Oldfield et al., 2014). However, there is still an extremely limited understanding of the mechanisms by which specific pioneers achieve their regulation of chromatin structure.

1.6 Key aims and questions

Chromatin acts as a barrier to transcription factors and polymerases, limiting their ability to transcribe DNA and regulate transcriptional networks. CGIs represent a major feature of vertebrate gene promoters, and their ability to recruit a wide variety of chromatin modifying activities appears to be essential for normal chromatin-based regulation of transcription. In fitting with these two points, it has been proposed that the elevated chromatin accessibility of CGIs may represent an important characteristic of these genomic elements as part of a transcriptionally permissive chromatin state. Here I attempt to address the contribution of different molecular features of CGIs towards this elevated chromatin accessibility, using mouse embryonic stem cells (ESCs) as a model system.

After initially establishing a suitable protocol to assay chromatin accessibility genome-wide in ESCs (Chapter 3), I aim to examine the relationship between non-methylated DNA and chromatin accessibility, and to interrogate the contribution of transcriptional activity and DNA methylation towards CGI chromatin accessibility (Chapter 4). An important question regarding CGIs is whether non-methylated CpG-rich sequences are sufficient to establish an open chromatin structure. I aim to address this question using a recombination system to target CpG-rich sequences into the mouse genome (Chapter 4). Furthermore, I set out to examine whether specific chromatin-modifying proteins targeted to CGIs by their ZF-CxxC domains could contribute to the regulation of chromatin structure and accessibility either positively or negatively, as predicted for TrxG- and PcG-associated factors respectively (Chapter 5).

While CGIs are important features of the vertebrate promoters, I also wanted to examine factors responsible for chromatin structure and accessibility at other regulatory elements in mouse ESCs. Specifically, I chose to address how the pioneer transcription factor OCT4 is able to access previously inaccessible regions of the genome to establish transcription

factor binding and chromatin accessibility (Chapter 6). Together, this thesis aims to provide some insight towards an improved understanding of chromatin accessibility at regulatory elements in ESCs, with a focus on CGIs. This may be important for future studies to interpret how chromatin accessibility may or may not be linked to transcriptional activity at gene promoters and other regulatory elements.

CHAPTER 2 - MATERIALS AND METHODS

2.1 Cell lines and culture conditions

2.1.1 Mouse embryonic stem cell culture

Mouse embryonic stem cells (ESCs) were grown on gelatin-coated plates in DMEM medium supplemented with 15% fetal bovine serum (FBS), 10ng/mL leukemia-inhibitory factor, penicillin/streptomycin, beta-mercaptoethanol, L-glutamine and non-essential amino-acids, with the exception of the *Brg1*$^{fl/fl}$ ESCs, which were maintained in DMEM KnockOut medium supplemented with 10% FBS and 5% serum replacement, plus additional factors described above. Cells were fed every 24-48h, and passaged using TryPLE disassociation reagent. For drug treatments, cells were counted using a haemocytometer and plated at specific cell densities 4-24h prior to treatment. See Table 1 for treatment and cell line details.

Table 1 Genotype and treatment conditions for mouse embryonic stem cell lines.
Description of genotype, deletion strategy and drug treatments for studying chromatin-modifying proteins in conditional or constitutive knockout ESCs. TAM = tamoxifen; DOX = doxycycyline.

Name	Genotype	Strategy	Drug treatment	Reference
E14	Wild type	n/a	n/a	
J1	Wild type	n/a	n/a	
DNMT TKO	*Dnmt1*$^{-/-}$;*Dnmt3a*$^{-/-}$;*Dnmt3b*$^{-/-}$	Constitutive KO	n/a	(Tsumura et al., 2006)
C*fp1*$^{fl/fl}$	*Cfp1*$^{fl/f}$*;Cre-ERT*	Conditional KO	96h TAM	Provided by H. Koseki, RIKEN
Fbxl19$^{CxxC\text{-}flox}$	*Fbxl19*$^{CxxC\text{-}flox}$;*Cre-ERT (CxxC-domain floxed)*	Conditional KO	96h TAM	Provided by H. Koseki, RIKEN
Kdm2.LF$^{fl/fl}$	*Kdm2a.LF*$^{fl/fl}$;*Kdm2b.LF*$^{fl/fl}$; *Cre-ERT*	Conditional KO	96h TAM	Provided by H. Koseki, RIKEN
Kdm2.LF$^{fl/fl}$*;SF*$^{-/-}$	*Kdm2a.LF*$^{fl/fl}$;*Kdm2b.LF*$^{fl/fl}$; *Kdm2a.SF*$^{-/-}$;*Kdm2b.SF*$^{-/-}$;*Cre-ERT*	Conditional KO (LF isoforms) Constitutive KO (SF isoforms)	96h TAM	Generated by Anne Turberfield, Oxford
Ring1a$^{-/-}$ *;Ring1b*$^{fl/fl}$	*Ring1a*$^{-/-}$*;Ring1b*$^{fl/f}$; *Cre-ERT*	Conditional KO	96h TAM	(Endoh et al., 2008)
Eed cKO	*Eed*$^{-/-}$*; Eed4*^{DOXsensitive}	Conditional KO	14d DOX	(Ura et al., 2008)
ZHBTC4	*hCMV-Oct-3;tTA*	Conditional KO	24h DOX	(Niwa et al., 2000)
Brg1$^{fl/fl}$	*Brg1*$^{fl/fl}$;*Cre-ERT*	Conditional KO	72h TAM	(Ho et al., 2011)
TOT2N	hBAC RP11 419E6 (chr8); 14x TetO	Stable	G418	(Blackledge et al., 2014)

2.1.2 Inhibition of RNA polymerase II

To inhibit RNA polymerase II (RNAPII) elongation or engagement, E14 ESCs were pre-plated at a 2.5×10^6 cells/10cm plate and allowed to grow for 24h prior to treatment with 300 nM flavopiridol (FLAV) or 500 nM triptolide (TRP) for 50min, as previously described (Jonkers et al., 2014). To limit re-activation of RNAPII, cells were immediately washed with ice-cold PBS and harvested by cell scraping prior to nuclei isolation for RNA and ATAC analysis.

2.1.3 Alkaline phosphatase staining

Alkaline phosphatase staining was performed by washing cells with PBS, followed by air drying for 15min. Cells were then incubated with freshly prepared AP buffer (0.4 mg/mL naphthol phosphate N-5000, 1 mg/mL Fast Violet B Salt F-161, 100mM Tris-HCl (pH 9.0), 100mM NaCl, 5mM $MgCl_2$) for 10min protected from light, prior to washing with PBS and imaging using a Leica M125 microscope.

2.1.4 Nuclei isolation and optimisation

Nuclei isolation for DNase I or Tn5 digestion was performed as previously described (Hosseini et al., 2013), using 1mL HS lysis buffer (50mM KCl, 10mM $MgSO_4.7H_20$, 5mM HEPES, 1mM PMSF, 3mM DTT) per 5×10^6 cells with varying concentrations of NP40 (IGEPAL CA630) (0 – 0.1%). After incubation with HS lysis buffer for 1min at room temperature, nuclei were pelleted by centrifugation at 1000×g for 5min at 4°C, and then washed three times with ice-cold RSB buffer (10mM NaCl, 10mM Tris (pH 7.4), 3mM $MgCl_2$), and mixed 1:1 with 2X TrypanBlue. Nuclei were then imaged using differential interference contrast (DIC) microscopy with an inverted Olympus microscope connected to a spinning disk confocal system (UltraView Perkin Elmer) using 60X objective.

2.2 Molecular Biology

2.2.1 Nuclear extracts

Harvested cell pellets were resuspended in 10X volumes of Buffer A (10mM Hepes (pH 7.9), 1.5mM $MgCl_2$, 10mM KCl, 0.5mM DTT, 0.5mM PMSF, and cOmplete protease inhibitors) followed by incubation of ice for 10min and centrifugation at 1500×*g* for 5min at 4°C. The cell pellet was then resuspended in 3X volumes of buffer A supplemented with 0.1% NP-40, mixed by inversion and centrifuged at 1500×*g* for 5min at 4°C. Nuclei were then resuspended in 1X volume of buffer B (5mM Hepes (pH 7.9), 26% glycerol, 1.5mM $MgCl_2$, 0.2mM EDTA, and cOmplete protease inhibitors, 0.5 mM DTT), and then NaCl was added to a final concentration of 400mM. Extraction was allowed to proceed at 4°C for 1h. Nuclei were then pelleted at 16,000×*g* for 20min at 4°C, and the supernatant was taken as the nuclear extract and stored at -80°C.

2.2.2 Acid extraction of histones

Cell pellets were washed with PBS and centrifuged at 1500×*g* for 5min at 4°C, prior to being resuspended in 1 mL RSB (10mM TrisHCl (pH 7.4), 10mM NaCl, 3mM $MgCl_2$). Cells were then pelleted at 1500×*g* for 5min at 4°C, and resuspended in 1mL RSB supplemented with 0.5% NP-40 followed by incubation on ice for 10min. After spinning at 1500×*g* for 5min at 4°C, pellets were resuspended in 0.5mL 5mM $MgCl_2$ followed by the addition of 0.5mL 0.8M HCl. After incubation of ice for 20min, samples were centrifuged at 16,000×*g* for 20min at 4°C, and the supernatant was transferred to fresh tube prior to precipitation of histone extract by the addition of trichloroacetic acid (TCA) to a final concentration of 25%. After incubation of ice for 30min, samples were centrifuged at 16,000×*g* for 15min at 4°C, and pellets were washed with 1mL ice-cold acetone, centrifuged at 16,000×*g* for 15min at 4°C, for a total of two times. Histone extract pellets were then air-dryed for 10-15min, and resuspended in 1X SDS loading

buffer (2% SDS, 100mM TrisHCl (pH 6.8) 100mM DTT, 10% glycerol, bromophenol blue), and incubated at 95°C for 5min, prior to storage at -20°C.

2.2.3 SDS-PAGE and Western blotting

Nuclear or histone extracts were diluted to final concentration of 1X SDS loading buffer, and run on SDS-PAGE gels, with between 6-15% acrylamide, depending on molecular weight of protein of interest. Electrophoresis was performed in 25mM TrisHCl, 192mM glycine, 0.1% SDS for 1h at 200V, prior to semi-dry transfer in 39mM glycine, 48mM TrisHCl, 0.037% SDS, 20% methanol at constant current of 200mA (limited voltage of 23V). Transfers for nuclear extracts were performed using 0.45µm nitrocellulose membrane for 1h, while histone extracts required 0.2µm membrane and a transfer time of 20min. Membranes were stained with Ponceau S to ensure efficient and even transfers, prior to blocking with 5% milk PBS-T (1X PBS with 0.1% Tween-20) for 1h at room temperature. Antibody details are detailed in Table 2. Membranes were typically probed with primary antibody overnight at 4°C, prior to washing 3× 5min with PBS-T the following day, and incubation with secondary antibody in 1% BSA PBS-T. Secondary antibodies either for standard ECL detection (anti-rabbit or anti-mouse IgG, HRP-LINKED (ThermoScientific)), or fluorescent-based detection using the LiCOR (goat anti-rabbit or anti-mouse IRDye 800CW or 680RD) were both used.

Table 2 Antibody details for western blotting

	Details	Manufacturer	Dilution range
CFP1	Rabbit monoclonal	David Brown, Klose lab	1:1000
HDAC1	Rabbit monoclonal	Abcam (ab109411)	1:2000
KDM2A	Rabbit monoclonal	Anca Farcas, Klose lab (Farcas *et al*. 2012)	1:1000
KDM2B	Rabbit monoclonal	Anca Farcas, Klose lab (Farcas *et al*. 2012)	1:1000
BRG1	Rabbit monoclonal	Abcam (ab110641)	1:5000
OCT4	Rabbit monoclonal	CST (#5677)	1:1000
SOX2	Rabbit monoclonal	CST (#23064S)	1:1000
NANOG	Rabbit monoclonal	CST (#8822S)	1:1000
RING1B	Rabbit monoclonal	CST (#5694S)	
EED	Rabbit polyclonal	Millipore (09-774)	1:500
TBP	Mouse monoclonal	Abcam (ab818)	1:2000
CTCF	Rabbit polyclonal	Millipore (07-729)	1:1000
H3K4me3	Rabbit polyclonal	Klose lab	1:2000
H2A	Rabbit monoclonal	CST (#12349s)	1:1000
H2AK119ub1	Rabbit monoclonal	CST (#8240S)	1:1000
H3K27me3	Rabbit polyclonal	Diagenode (C15410195)	1:1000

2.3 Expression and purification of Tn5 transposase

2.3.1 Preparation of Tn5 MEDS oligonucleotides

Tn5 mosaic end oligonucleotides A and B (ME-A 5' GTCTCGTGGGCTCGGAGATGTGT ATAAGAGACAG 3'; ME-B 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG 3') were hybridized with equal amounts of the partially complementary reverse oligonucleotide (ME-REV 5' $_{P}$CTGTCTCTTATACACATCT 3') by incubation at 95°C for 10min, followed by incubation at room temperature to allow for gradual annealing to produce mosaic end double stranded oligonucleotides (MEDS)-A and MEDS-B.

2.3.2 *E.coli* expression of Tn5 transposase

The pTBX-Tn5 plasmid containing the hyperactive Tn5 allele (E54K and L372P mutations) with intein-chitin binding domain (CBD) tag was transformed into C3013 *Escherichia coli* cells, and used to inoculate a 5mL starter culture overnight in LB and 100µg/mL ampicillin. The starter culture was then used to inoculate 1L culture of LB (100µg/mL ampicillin), which was cultured at 37°C until A600 reached approximately 0.7 to 0.9, and then chilled

to 10°C. IPTG was added to final concentration of 0.25mM and cells were incubated at 23°C for 4h. Cells were then pelleted at 12,500 rpm for 20min, and cell pellet frozen at -70°C to await purification.

2.3.3 Purification of Tn5 by thiol-mediated cleavage of intein-CBD

After thawing on ice, the cell pellet was resuspended in 20mL ice-cold HEGX buffer (20mM HEPES-KOH (pH 7.2), 0.8M NaCl, 1mM EDTA, 10% glycerol, 0.2% Triton X-100), with cOmplete protease inhibitors (Roche) and then lysed by sonication (BioRuptor; 9× 30sec on/off at 80% power) on ice. The lysate was then centrifuged at 12,500rpm for 30min at 4°C, and the supernatant transferred to a fresh beaker prior to the dropwise addition of 2.1mL of 10% neutralized PEI to precipitate DNA. Lysate was then centrifuged at 12,500rpm for 20min at 4°C, and the soluble cell extract was run through a pre-washed 10mL chitin column (washed with 100mL HEGX buffer) at 4°C. The column was then washed with a total of 200mL cold HEGX buffer. 250nmoles of prepared MEDS-A and MEDS-B oligonucleotides were diluted in 12mL of degassed HEGX buffer with protease inhibitors, added to the chitin column and incubated overnight at room temperature. The following day, the column was returned to 4°C and washed with a total of 200mL cold HEGX buffer. To elute Tn5, the column was incubated with 20mL HEGX buffer containing 50mM DTT for 48h at 4°C, prior to elution in 1mL fractions. The column was then washed three times with 5mL cold HEGX containing 50mM DTT to maximise recovery of Tn5.

Eluted fractions were run on 10% SDS-PAGE and Coomassie stained to confirm the presence of the 55kD Tn5 transposase protein and purity. All eluted fractions containing the Tn5 product were pooled together and concentrated to approximately 4mL using 10kD MW VivaSpin column by spinning at 4,500×*g*, prior to dialysis overnight at 4°C in 2L of 2X Tn5 dialysis buffer (100mM HEPES-KOH (pH 7.2), 200mM NaCl, 0.2mM EDTA, 0.2% Triton X-100, 20% glycerol) containing fresh 2mM DTT using Spectra/Por 6-8kD dialysis membrane. Dialyzed sample was collected and Tn5 yield and molar concentration was

determined by Bradford colormetric assay. Purified Tn5 was diluted to 24µM using 2X Tn5 dialysis buffer, and diluted 1:2 with 100% glycerol to give a final concentration of 12µM Tn5 in 60% glycerol, before aliquoting and storing at -20°C.

2.3.4 Testing *in vitro* activity of purified Tn5 transposase

500 ng of genomic DNA was incubated with varying amounts of purified Tn5 in 10mM TAPS, 5mM $MgCl_2$, 10% dimethylformamide at 55°C for 30min. Digestion of genomic DNA was confirmed by agarose gel (1%) electrophoresis.

2.4 Measuring chromatin accessibility

2.4.1 DNase I digestion and DNase-seq library generation

DNase I digestions were performed as previous described (Hosseini et al., 2013), with some modifications. Briefly, 5×10^7 cells were harvested and gently resuspended in 10mL freshly prepared HS Lysis buffer (0.05% NP-40). After 1min incubation at room temperature, nuclei were pelleted by centrifugation at 1,000×*g* for 5min at 4°C, and washed twice with 10mL ice-cold RSB buffer. Isolated nuclei were then resuspended at 2.5×10^7 nuclei/mL and 5×10^6 nuclei were then incubated with varying amounts of DNase I (Roche) (0 – 320U) for exactly 3min at 37°C. Digestions were stopped with an equal volume of 10mM Tris-HCl (pH 8.0), 10mM NaCl, 0.5% SDS, 10mM EDTA, followed by incubation with RNase and proteinase K at 55°C for 4h. Genomic DNA was then purified by phenol-chloroform extraction with PhaseLock tubes, followed by precipitation with 95% ethanol and sodium chloride. Purified genomic DNA was quantified by NanoDrop, assessed by 1% agarose gel electrophoresis for digestion efficiency and used for DNase I quantitative real-time PCR (DNase-qPCR) (see 2.4.3).

DNA from the DNase I digestions was subjected to size selection for 50-200bp fragments using several different methods. 20µg DNA was run on low-melting temperature agarose gel, and fragments were manually excised and purified using the QIAquick Gel Extraction kit. Alternatively, 20 to 40µg of DNA was run on 2% EGel® SizeSelect Agarose Gel Electrophoresis or the 3% SizeSelect PippinPrep systems according to manufacturers' protocols. Size-selected DNA was then used to make DNase-seq libraries with the NEBNext® Fast DNA Library Prep Set for Ion Torrent™ kit. Briefly, DNA fragments were end-repaired followed by ligation of the IonTorrent barcodes. Excess barcodes and adaptors were removed by size-selection on the EGel®, followed by PCR amplification of 4-6 cycles if necessary. Alternatively, 2 to 5µg of DNA from the DNase I digestions was used without preliminary size-selection and was directly subjected to end repair and adaptor ligation, followed by size selection of adaptor-ligated fragments corresponding to insert sizes of between 75 and 200bp, and PCR amplification. Libraries were quantified by qPCR using SensiMix SYBR (Bioline) and KAPA Library Quantification DNA standards (KAPA Biosystems) and multiplexed at 10pM for sequencing using the Ion Proton™ sequencing system.

2.4.2 ATAC-seq sample and library preparation

Chromatin accessibility was assayed using an adaptation of the assay for transposase accessible-chromatin (ATAC)-seq (Buenrostro et al., 2013). Briefly, 5×10^6 cells were harvested, washed with PBS and nuclei were isolated using 1mL freshly prepared HS Lysis buffer for 1min at room temperature. Nuclei were centrifuged at 1,000×*g* for 5min at 4°C, followed by a total of three washes with ice-cold RSB buffer (10mM NaCl, 10mM Tris (pH 7.4), 3mM $MgCl_2$), to remove as much of contaminating cytoplasmic and mitochondrial material as possible. Nuclei were then counted, and 5×10^4 nuclei were resuspended in Tn5 reaction buffer (10mM TAPS, 5mM $MgCl_2$, 10% dimethylformamide) and 2 µl of Tn5 transposase (25µM) made in house (see Section 2.3). Nuclei were then incubated for 30min at 37°C, before isolation and purification of tagmented DNA using

QiaQuick MinElute columns (Qiagen). To control for sequence bias of the Tn5 transposase, a Tn5 control sample was generated by tagmenting genomic DNA from ESCs with Tn5 for 30min at 55°C. ATAC-seq libraries were prepared by PCR amplification using custom made Illumina barcodes previously described (Buenrostro et al., 2013) and the NEBNext® High-Fidelity 2X PCR Master Mix with 8-10 cycles. Libraries were purified with two rounds of AMPure XP bead cleanup (1.5X beads:sample), followed by quality control of libraries using the Bioanalyser (High Sensitivity DNA). Libraries were quantified by qPCR using SensiMix SYBR (Bioline) and KAPA Library Quantification DNA standards (KAPA Biosystems). An initial ATAC-seq experiment was sequenced on the Ion Proton™ sequencing system, but the majority of ATAC-seq libraries were sequenced on Illumina NextSeq500 using 80bp paired-end reads in biological duplicate or triplicate.

For the calibrated ATAC-seq of *Kdm2.LF*$^{fl/fl}$ and *Kdm2.LF*$^{fl/fl}$*;SF*$^{-/-}$ ESCs, 2.5×10^6 *Drosophila melanogaster* S2 cells were spiked into 1×10^7 ESCs with or without tamoxifen treatment, and processed as normal ATAC-seq experiments. To control for variation in spike-in between samples and biological replicates, a genomic DNA preparation was made from each sample and tagmented directly with Tn5 for 30min at 55°C. Both ATAC and Tn5 control libraries were then sequenced as normal.

2.4.3 DNase- and ATAC-qPCR

To assess digestion efficiency of DNase I or Tn5 transposase at specific loci, 200bp amplicons were designed across DNase I hypersensitive sites identified by ENCODE, as well as control intergenic regions that did not demonstrate any DNase I hypersensitivity. Where possible, 5ng of DNA was used for each qPCR reaction to assay the amplification of target and control amplicons with the Quantace SYBR Green master mix (Bioline) using a Rotor-Gene 6000 (Corbett). The cycle threshold (Ct) values were calculated using the

Rotor-Gene software, and were normalised for each primer pair according to the amplification efficiency of the control primer pair.

$$Normalied\ Ct = \frac{10^{\left(-\frac{1}{slope\ of\ target\ amplicon}\right)}}{10^{\left(-\frac{1}{slope\ of\ control\ amplicon}\right)}} \times Ct$$

The proportion of copies remaining after nuclease digestion were then calculated relative to the undigested control using the ΔΔCt method (McArthur et al., 2001).

$$\%\ copies\ remaining = 2^{-\left(DHS\ Ct_{digested} - Control\ Ct_{\ digested}\right) - \left(DHS\ Ct_{undigested} - Control\ Ct_{undigested}\right)} \times 100$$

Table 3 Primers for DNase- and ATAC-qPCR

Locus	Gene/Name	F/R	Sequence (5' – 3')
chr11:100921731-100921913	Intergenic control region	F	TTTGCCAGGTGCTCTGCTAA
		R	GCTGCCATTCCATCAAAACA
chr1:64121635-64121817	Klf7 DHS	F	AGGAGGGCGTCCATTAGG
		R	AAATAGCTTGCCTCGTACCG
chr5:110770641-110770851	Ep400 DHS	F	GCGAGGGACTAACGGTACCTA
		R	CTCGTACGGATGAGGCCTTT
chr3:51510937-51511123	CTCF-binding site DHS	F	TCCTGATTTTCCAGCCTCTG
		R	ATTGCCTTCACCCCTAGCTT
RMCE-HygroTK	HCG1 (recombinant product)	F	GGTAGAACCAGGTGCCAAGA
		R	CTGCTGATAACGGCTGTCAA
RMCE-HygroTK	HCG2 (recombinant product)	F	GGTAGAACCAGGTGCCAAGA
		R	CTTGCGAGAGTGCCCTAAAG

2.5 Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed as described previously (Farcas et al., 2012), with minor modifications. Cells were fixed for 1h in 2mM disuccinimidyl glutarate (DSG) and 12.5min in 1% formaldehyde. Reactions were quenched by the addition of glycine to a final concentration of 125µM. After cell lysis and chromatin extraction, chromatin was sonicated using a BioRuptor sonicator (Diagenode), followed by centrifugation at 16,000×*g* for 20min at 4°C, and used fresh or stored at -80°C. Chromatin was quantified by diluting chromatin ten-fold with 0.1M NaOH, and measuring DNA concentration by NanoDrop. 150µg chromatin/IP was diluted ten-fold in ChIP dilution

buffer (1% Triton-X100, 1mM EDTA, 20mM Tris-HCl (pH 8), 150mM NaCl) prior to pre-clearing with prepared protein A magnetic Dynabeads (Invitrogen) blocked for 1h with 0.2mg/mL BSA and 50µg/mL yeast tRNA. Antibodies used for ChIP experiments were anti-OCT-4A (Cell Signalling, #5677), anti-SOX2 (Cell Signalling, #23064), anti-Nanog (Cell Signalling, #8822), and anti-BRG1 (abcam, ab110641). Antibody-bound chromatin was isolated on protein A magnetic Dynabeads. ChIP washes were performed with low salt buffer (0.1% SDS, 1% Triton, 2mM EDTA, 20mM Tris-HCl (pH 8.1), 150mM NaCl), high salt buffer (0.1% SDS, 1% Triton, 2mM EDTA, 20mM Tris-HCl (pH 8.1), 500mM NaCl), LiCl buffer (0.25M LiCl, 1% NP40, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris-HCl (pH 8.1)) and TE buffer (x2 washes) (10mM Tris-HCl (pH 8.0), 1mM EDTA). To prepare ChIP-seq material, ChIP DNA was eluted using 1% SDS 100mM $NaHCO_3$, and cross-links reversed at 65^{o}C in the presence of 200mM NaCl. Samples were then treated with RNase and proteinase K before being purified with ChIP DNA Clean & Concentrator™ kit (Zymo). ChIP-seq libraries were prepared using the NEBNext Ultra DNA Library Prep Kit, and sequenced as 40bp paired-end reads on Illumina NextSeq500 platform. All ChIP-seq experiments were carried out in biological triplicate.

2.6 Nuclear RNA isolation and sequencing

To isolate nuclear RNA, cells were subjected to nuclei isolation described for ATAC-seq protocol. Nuclei were then resuspended in 1mL of TriZOL reagent (ThermoScientific) and processed according to the manufacturer's protocol. Nuclear RNA was treated with the TURBO DNA-free Kit (ThermoScientific) and used for cDNA generation using the SuperScript II Reverse Transcriptase kit for qRT-PCR analysis. After quality control, RNA samples were depleted for rRNA using the NEBNext rRNA Depletion Kit and protocol. RNA-seq libraries were prepared using the NEBNext Ultra Directional RNA-seq kit and

libraries were sequenced on the Illumina NextSeq500 with 80bp paired-end reads. All RNA-seq experiments were carried out in biological triplicate.

2.7 Recombinase-mediated cassette exchange

2.7.1 Generating RMCE-HygroTK ESCs

The hygromycin phosphotransferase-thymidine kinase (HygroTK) fusion gene was amplified from a plasmid provided by Dr Neil Blackledge (pTAG-PCGF1) using Phusion HF polymerase. LoxP/LoxP' and XbaI digestion sites were introduced during PCR amplification, and the PCR product was gel extracted and cloned into the pBSII-CAG vector between a CAG promoter and poly-adenylation sequence, flanked on either side by 1kb homology arms for the human gene desert in the TOT2N mouse ESC (Blackledge et al., 2014). CRISPR/Cas9 guide RNAs were designed by Dr Neil Blackledge to target the TetO array in the TOT2N ESCs, and were co-transfected with the pBSII-CAG-HygroTK vector using Lipofectamine 3000. The following day, cells were trypsinised and re-plated at varying densities, before selection with 250µg/mL hygromycin for 10 days. Positive clones were identified by genomic DNA PCR using TruStart polymerase, and confirmed by sequencing of PCR products.

2.7.2 Generation of targeting construct and cloning *S. enterica* DNA fragments

To introduce exogenous sequences into these RMCE-HygroTK ESCs, I generated a plasmid expressing Turbo-Cre recombinase under the CAG expression promoter (pCAG-CRE). In order to ensure that every cell transfected with Cre also had the required donor sequence, I generated a new ligation-independent cloning (LIC) site into the pCAG-CRE vector which was flanked by loxP and loxP' recombination sites. This fragment (loxP-LIC-

loxP') was synthesized by Invitrogen GeneArt and cloned into the SalI restriction site within the pCAG-CRE vector (pCAG-CRE-Exchange Cassette (-EC)). To optimise the system, I cloned *eGFP* into the exchange cassette plasmid to generate pCAG-CRE-GFP. Ten ~1kb sequences were amplified using Phusion HF polymerase from *Salmonella enterica* genomic DNA (ATCC® 700720D-5™) with overhanging LIC adaptor sequences, and LIC cloned into the pCAG-CRE-EC vector. *S. enterica* was chosen after bioinformatic fragmentation and comparison of various bacterial genomes with the CpG density and GC percentage of mammalian CGIs (Supplementary Figure 2).

Table 4 Primers for cloning *S. enterica* fragments.
Primers used for amplification of fragments from *S. enterica* genomic DNA. Ligation independent cloning (LIC) adaptors are indicated in *italics*.

	Forward (5'-3')	Reverse (5'-3')
LCG1	*TACTTCCAATCCATG*TCGGGTGGAAATTCTGGTAG	*TATCCACCTTTACTG*CTGAACTGGCAGCGATAACA
LCG2	*TACTTCCAATCCATG*CTATCCGTCGCGATAATGGTG	*TATCCACCTTTACTG*CATAAGCCGTCGGTGAAACT
LCG3	*TACTTCCAATCCATG*GGATTGCGTCTATGCCTGATC	*TATCCACCTTTACTG*GGTCGGCAACTTTGAAACAT
HCG1	*TACTTCCAATCCATG*CAACGGCCTCTTTAAGGAAA	*TATCCACCTTTACTG*CGCGATATCAACCCAGTTTT
HCG2	*TACTTCCAATCCATG*TGGACAGCGAGAATCAACTG	*TATCCACCTTTACTG*CCTCGCCTTCCAGGATAACT
HCG3	*TACTTCCAATCCATG*TGGACGAAAACGTGTGACAT	*TATCCACCTTTACTG*ACGCGTACCGTGGTTTACTC
HCG4	*TACTTCCAATCCATG*CCCATCACTTCCAACGACTTG	*TATCCACCTTTACTG*TTCCTGGTGGCGAATAAAAC
HCG5	*TACTTCCAATCCATG*ATATGGCAGGACTGGCTGAAC	*TATCCACCTTTACTG*ATAACGTCCCTGCATCTTCG
HCG6	*TACTTCCAATCCATG*TACGCAACAGGCAGAAGATG	*TATCCACCTTTACTG*CAACACTTCGCGCAGATAAAC
HCG7	*TACTTCCAATCCATG*AGTACACCCTGCCAGCATCTG	*TATCCACCTTTACTG*GAAGCGGATTTCACCTTTGA

2.7.3 Insertion of bacterial sequence into RMCE locus

To target DNA sequences into the RMCE-HygroTK ESCs, 4×10^6 cells were pre-plated 24h in advance and 3µg of pCAG-CRE-GFP or *S. enterica* fragment-containing plasmids were transfected using Lipofectamine 3000. The following day, cells were trypsinised and re-plated at varying densities, before selection with either 3 or 6µM ganciclovir for 10 days.

Ganciclovir-resistant clones were isolated and expanded in 96-well plates for genomic DNA extraction and PCR screening using TruStart polymerase.

2.7.4 Methylation-sensitive qPCR

Methylation of recombinant sequences was determined by digesting genomic DNA for 4h at 37°C with either MspI (methylation-insensitive) or HpaII (methylation-sensitive), similar to that described previously (Hashimoto et al., 2007). Following heat inactivation of 65°C for 20min, SYBR Green qPCR was performed on digested genomic DNA as well as an undigested control using primers specific to recombinant or endogenous sequences. Comparison of the amplification efficiencies of digested and undigested samples using the ΔΔCt method allows direct measurement of digestion efficiency of the methylation sensitive HpaII, allowing inference of DNA methylation levels when compared to known methylated or non-methylated endogenous sequences. See 2.4.3 for more details on ΔΔCt calculations.

2.8 Massively parallel sequencing, data processing and normalisation

2.8.1 Sequencing using IonProton and Illumina technologies

The majority of next generation sequencing data included in the present thesis was generated on the Illumina NextSeq500 sequencing platform located at the Botnar Research Centre, Institute of Musculoskeletal Sciences, University of Oxford, with support from Ed Hookway and Udo Oppermann, using paired-end reads. Alternatively, DNase-seq data generated during early stages of the thesis were sequenced on the IonProton located in the Department of Biochemistry, University of Oxford, with single-end reads.

2.8.2 Data alignment, normalisation and visualisation

Single-end DNase-seq reads were aligned to the mouse mm10 genome using bowtie2 (Langmead and Salzberg, 2012), after automated read quality trimming by IonProton software. For ATAC-seq experiments (NextSeq500), paired-end reads were aligned to the mouse mm10 genome using bowtie2 with the "--no-mixed" and "--no-discordant" options. Non-uniquely mapping reads and reads mapping to a custom “blacklist” of regions of the genome with artificially high counts, such as mitochondrial DNA sequences, were discarded, with the exception for the DNMT TKO repetitive element analysis, for which all aligned reads were kept. For ATAC-seq experiments which contained a spike-in genome (*D. melanogaster*; genome version dm6), reads were aligned against a concatenated genome of the two genome sequences (mm10+dm6), and reads which mapped more than once were discarded, prior to separation of mm10- or dm6-specific reads.

For RNA expression analysis in wild type ESCs, 4-thiouridine (4SU)-seq, generated by Dr Vincenzo di Cerbo, was used. 4SU-seq provides a more nascent measure of transcriptional activity than conventional RNA-seq protocols (Friedel and Dolken, 2009, Rabani et al., 2011, Schwanhausser et al., 2011). RNA-seq data were initially aligned against the rRNA genomic sequence (GenBank: BK000964.3) using bowtie2 to filter out rRNA fragments, prior to alignment against the mm10 genome using the STAR RNA-seq aligner (Dobin et al., 2012). To improve mapping of nascent, intronic sequences, reads that failed to map using STAR were aligned against the genome using bowtie2. Total RNA-seq for J1 and DNMT TKO ESCs was generated and processed by Martin Cusack, Kriaucionis lab, Oxford, using TopHat2 (Kim et al., 2013b).

Whole genome bisulfite sequencing data for mouse ESCs (Stadler et al., 2011) was aligned using bismark with default settings and was used to calculate % methylation for CpGs genome-wide (Krueger and Andrews, 2011). % CpG methylation scores for bisulfite sequencing data from mouse tissues (Hon et al., 2013) were converted from mm9 to

mm10 using liftOver (Hinrichs et al., 2006), and annotated to genomic intervals using HOMER (Heinz et al., 2010).

PCR duplicates were removed using SAMtools (Li et al., 2009b). Biological replicates were randomly downsampled to contain the same number of reads for each individual replicate. Following quality control and confirmation of reproducibility between replicates through visual inspection of the data and genome-wide correlations, high quality biological replicates were merged to create a representative genome track using DANPOS2 for ATAC-seq or transcription factor ChIP-seq (Chen et al., 2013) or MACS2 pileup function for other ChIP-seq or DNase-seq datasets. CpG density coverage track was generated using a 200bp window shifted by 25bp. Bigwig tracks were visualised using the UCSC Genome Browser (Kent et al., 2002, Raney et al., 2014).

2.8.2.1 Calibrating ATAC-seq with Drosophila control

Spiking an identical concentration of *D. melanogaster* cells into ATAC-seq experiments allows for calibration of each sample for digestion efficiency and total mouse cell number. When comparing ATAC-seq for untreated and tamoxifen-treated cell lines using an internal calibration, the number of mm10 reads were randomly downsampled to reflect the total number of dm6 reads in that sample. Furthermore, to adjust for variation in the dm6/mm10 ratio between biological replicates, each sample was adjusted using the percentage of dm6 reads relative to mm10 reads in each sample's genomic DNA control.

2.9 Annotation and analysis of regulatory elements

2.9.1 Peak calling

For ATAC-seq experiments, peakcalling was performed using the DANPOS2 dpeak function using untreated and treated samples in biological duplicate or triplicate with matched input. Peaks for BioCAP-seq, RING1B and SUZ12 ChIP-seq were identified using MACS2 (Zhang et al., 2008). Intersections with ATAC peaks or TSS intervals were performed using BEDtools. Computationally predicted CpG islands were obtained from UCSC genome browser (mm10).

2.9.2 Annotation of putative promoters and enhancers

ATAC or OCT4 peaks were annotated as putative promoters or putative distal regulatory elements in a manner similar to that described previously (Hay et al., 2016), using the relative and absolute coverage of H3K4me3 (a promoter-associated modification) and H3K4me1 (associated with distal regulatory elements), as calculated by reads per kilobase per million (RPKM). More precisely, intervals were annotated as promoters if H3K4me3 RPKM > 10, H3K4me1 RPKM < 5 and H3K4me1 RPKM – H3K4me3 RPKM < 0. Alternatively, for distal regulatory elements, H3K4me3 RPKM < 10, H3K4me1 RPKM > 0.75 and H3K4me1 RPKM – H3K4me3 RPKM > 0. Regions that could not be confidently assigned to either group were not considered for further analysis.

2.9.3 Genome-wide correlations

Genome-wide correlations between ChIP-seq datasets were performed using deepTools2 (Ramírez et al., 2016) with Pearson correlations using 1 kb windows. In contrast, comparison of DNase-seq with bisulfite sequencing was performed by manually annotating mappable 1 kb windows with DNase-seq RPKM and % methylation, and correlations were performed using the Spearman (non-parametric) correlation method. Comparison of different remodelling complexes was performed by calculating RPKM for

chromatin remodeller ChIP-seq in mouse ESCs (de Dieuleveult et al., 2016) across all ATAC peaks (n=76,642) and correlating with the log2 fold change of ZHBTC4 ATAC-seq.

2.9.4 Gene ontology and motif analysis

Transcription factor peaks were analysed using the GREAT package (McLean et al., 2010), in particular to extract information regarding developmental expression timing from MGI Gene eXpression Database. Gene ontology analysis for differentially regulated genes was performed using HOMER2. Vertebrate transcription factor motifs were identified in HCG1 and HCG2 sequences using the MATCH algorithm (Goessling et al., 2001).

2.10 Differential accessibility, factor binding and gene expression

To compare chromatin accessibility at gene promoters, a custom-made non-redundant mm10 gene set was used (±500bp). Briefly, refGene intervals were filtered on size (>200bp), gene body and TSS mappability, unique TSS and TTS, in order to remove poorly mappable and highly similar transcripts. ESC-active TSS were identified using an RPKM threshold of 0.7 for two independent Cap Analysis of Gene Expression (CAGE) ESCs datasets from FANTOM (FANTOM et al., 2014). Read counts were generated using custom scripts. GC content and CpG nucleotide frequencies were calculated using BEDtools (Quinlan and Hall, 2010).

Significant changes for ATAC-seq or ChIP-seq datasets were identified using the DiffBind package (Stark and Brown, 2011) at TSS or ATAC peak intervals. Alternatively, calibrated ATAC-seq datasets were compared using custom normalisation factors generated from

the *Drosophila* internal control, and analysed using DESeq2 (Love et al., 2014). In both cases, *padj* < 0.05 and a fold change greater than 1.5 fold was deemed significant.

Changes in ATAC-seq were visualised using heatmaps or metaplots produced using HOMER2 (Heinz et al., 2010), with heatmaps visualised using Java TreeView (Saldanha, 2004). Log2 fold change values were visualised using R (v 3.2.1), with scatterplots coloured by density using stat_density2d. Regression and correlation analyses were also performed in R using standard linear models and Pearson correlation respectively.

For differential RNA-seq analysis, DESeq2 was used with the custom-built, non-redundant mm10 gene set. A *padj* < 0.05 and a fold change greater than 1.5 fold was deemed significant. When considering distal regulatory elements, the gene expression changes for the nearest TSS were considered.

2.11 Repetitive elements analysis

The mm10 repeat masker (rmsk) intervals were downloaded from the UCSC genome browser. Genomic intervals for different repeat elements (RepName) were extracted and RPKM and CpG densities were calculated for the J1 and DNMT TKO ATAC-seq and RNA-seq using merged files (representing biological triplicate) with custom scripts. Repeat age was calculated for individual repeats, and averaged for each repeat class (RepName) as previously described (Ward et al., 2013, Long et al., 2016) using the number of substitutions from consensus sequence (milliDiv), divided by the estimated mutation rate for mammalian species (2.2×10^{-9}/year) (Kumar and Subramanian, 2002). Normalised interval plots were generated using HOMER.

2.12 Accession codes and genomic datasets

Table 5 Published genomic datasets included in this study.

Description of datasets used in this thesis, including assay, cell types, publication and NCBI Gene Expression Omnibus (GEO) accession codes.

Assay/Datasets	Cell Type	Reference	GEO Accession Code
H3K4me1, H3K4me3 ChIP-seq	ESC	(Yue et al., 2014)	GSE49847
BioCAP	ESC	(Long et al., 2013b)	GSE43512
Whole genome bisulfite sequencing	ESC	(Stadler et al., 2011)	GSE30206
Whole genome bisulfite sequencing	Mouse tissues	(Hon et al., 2013)	GSE42836
DNase-seq	Mouse cell lines and tissues	(Yue et al., 2014, Vierstra et al., 2014a)	GSE51336
GRO-seq (untreated, flavopiridol, triptolide)	ESC	(Jonkers et al., 2014)	GSE48895
KDM2A, KDM2B ChIP-seq	ESC	(Farcas et al., 2012)	GSE41267
H3K4me3 ChIP-seq	MLL1/2-null ESCs	(Denissov et al., 2014)	GSE52071
CHD1, CHD2, CHD4, CHD6, CHD8, CHD9, EP400 ChIP-seq	ESC	(de Dieuleveult et al., 2016)	GSE64825
INO80 ChIP-seq	ESC	(Wang et al., 2014)	GSE49137
RNA-seq	iPSC timeocourse	(Cieply et al., 2016)	GSE70022
ATAC-seq	Early mouse embryo	(Wu et al., 2016)	GSE66390
Polycomb ChIP-seq	ESC	(Rose et al., 2016)	GSE83135

Table 6 Genomic datasets generated in this study.

Summary of sequencing statistics for next-generation sequencing experiments, including the number of replicates, % mappability and final read depth per replicate.

Experiment	Assay	Sample	Rep.	Read Number	Mapped		Final Read Depth
DNase-seq Digestion Curve (0.03% NP-40)	DNase-seq	10U	1	7481451	5404183	72.2%	3227676
		20U	1	8440544	6090811	72.2%	3738858
		40U	1	7820975	5451362	69.7%	2976584
		60U	1	8026300	5552531	69.2%	2889096
		80U	1	6880295	5009604	72.8%	2386655
		120U	1	8985968	6673942	74.3%	3775582
		160U	1	7004030	5249426	74.9%	2879596
		320U	1	5383435	4048500	75.2%	2617017
DNase-seq Size Selection (0.03% NP-40)	DNase-seq	Fraction A	1	4360883	2744392	62.9%	1387543
		Fraction B	1	4925834	3134232	63.6%	1583418
		Fraction C	1	3776858	2456869	65.1%	1385934
		Fraction D	1	4106906	2809551	68.4%	1558107
		Fraction E	1	3083006	1938596	62.9%	1148242
DNase-seq Digestion Curve (0.05% NP-40)	DNase-seq	60U	1	20459659	14809554	72.4%	10721552
		75U	1	24193080	17580292	72.7%	12796585
		90U	1	17223300	12570347	73.0%	9189388
		120U	1	21350791	15560721	72.9%	12094882
DNase-seq Digestion Curve Biological Duplicate (0.05% NP-40)	DNase-seq	40U	1	4206718	3193631	75.9%	2875126
			2	3464412	2598799	75.0%	2495896
		60U	1	6135436	4608675	75.1%	4083816
			2	4282252	3105991	72.5%	2981188
		80U	1	7533765	5574984	74.0%	5250699
			2	4099112	3006290	73.3%	2872567
		100U	1	8000258	5916277	74.0%	5519608
			2	4828953	3512530	72.7%	3345115
		120U	1	6575631	4716688	71.7%	4415719
			2	6730792	4898993	72.8%	4654465

Table 6 continued….

Experiment	Assay	Sample	Rep.	Read Number	Mapped		Final Read Depth
DNase-seq gDNA controls	DNase-seq	gDNA control (IonProton)	1	15215297	9162220	60.2%	7815100
	DNase-seq	gDNA control (HiSeq)	1	31154914	21591997	69.3%	10557350
ESC ATAC-seq Optimisation (IonProton Sequencing)	ATAC-seq	1µl Tn5 ATAC	1	8177236	445805	5.5%	428290
			2	7410254	429839	5.8%	409638
		2µl Tn5 ATAC	1	8430204	494548	5.9%	474346
			2	12868760	653320	5.1%	625329
		5µl Tn5 ATAC	1	9679298	487392	5.0%	470772
			2	8327317	510737	6.1%	494389
		gDNA control (IonProton)	1	9658472	430250	4.5%	423290
E14 ESC ATAC-seq (Illumina)	ATAC-seq	E14 ATAC	1	58122186	34520418	59.4%	16941795
			2	75348457	45696492	60.6%	22522647
			3	74558876	45424193	60.9%	22260722
		gDNA control (Illumina)	1	67698833	38861360	57.4%	19633387
Transcriptional Inhibitors (ESCs)	ATAC-seq	Untreated	1	35633473	18067728	50.7%	17408045
			2	38894668	19450222	50.0%	18745458
			3	43724007	20244085	46.3%	18985159
		Flavopiridol	1	44435199	23081495	51.9%	22174225
			2	41308788	21239586	51.4%	20324432
			3	31665426	17182144	54.3%	16648548
		Triptolode	1	32217375	17113250	53.1%	16549120
			2	26268293	13890441	52.9%	13441811
			3	42883361	22735788	53.0%	22033027
DNA Methylation-null ESCs	ATAC-seq	J1 (wild type)	1	42059818	29736292	70.7%	22662530
			2	40802365	29649574	72.7%	22924431
			3	46695157	31882555	68.3%	24759302
			4	35410344	24748708	69.9%	18956677
		DNMT TKO	1	46392280	32955190	71.0%	24356765
			2	36432184	25675537	70.5%	19700704
			3	41513659	28176784	67.9%	21677214
			4	44012434	31108765	70.7%	24214483
		J1 gDNA control	1	28923379	12871603	44.5%	9743551
		DNMT TKO gDNA control	1	31144961	15762255	50.6%	12017967
Fbxl19$^{CxxC\text{-}flox}$ ESCs	ATAC-seq	Untreated	1	87334943	56660418	64.88%	54342235
			2	47028896	30738754	65.36%	29901417
			3	32955915	19935457	60.49%	19102552
		Tamoxifen	1	32531337	20845646	64.08%	20082540
			2	57931095	37855990	65.35%	36786432
			3	38392077	24119269	62.82%	23167854
Cfp1$^{fl/fl}$ ESCs	ATAC-seq	Untreated	1	29634303	21283382	71.8%	20803653
			2	29154519	21153818	72.6%	20787573
			3	30232868	21654834	71.6%	21153756
			4	33196962	23687361	71.4%	23197534
		Tamoxifen	1	39810030	28477920	71.5%	27743041
			2	30061718	21629865	72.0%	21307454
			3	29218533	20612135	70.5%	20114810
			4	35827702	24921612	69.6%	24281192
Ring1a$^{-/-}$;Ring1b$^{fl/fl}$ ESCs	ATAC-seq	Untreated	1	40136333	24963593	62.2%	23967850
			2	44085880	26356338	59.8%	25078099
		Tamoxifen	1	39325188	24993742	63.6%	24001403
			2	36077688	24035210	66.6%	21405089
Eed* cKO *ESCs	ATAC-seq	Untreated	1	17016269	11460005	67.3%	11003054
			2	20313903	13771011	67.8%	13239588
		Doxycycline	1	20462880	13804271	67.5%	13141497
			2	18179857	12412466	68.3%	11910792

Table 6 continued….

Experiment	Assay	Sample	Rep.	Read Number	Mapped		Final Read Depth
ZHBTC4 (OCT4^{cond}) *ESCs*	ATAC-seq	Untreated	1	64779469	44549536	68.77%	42583471
			2	42800770	30601335	71.50%	29703385
			3	50560942	34478700	68.19%	33155661
		Doxycycline	1	53598021	37305979	69.60%	35708824
			2	47183790	33661830	71.34%	32351468
			3	64747247	45216011	69.83%	43279376
	OCT4 ChIP-seq	Untreated	1	23283443	14719990	63.22%	14320863
			2	26718474	17416442	65.19%	16992714
			3	22820152	14673000	64.30%	14208363
		Doxycycline	1	24477978	16137344	65.93%	15357569
			2	26973135	17842629	66.15%	17305842
			3	19513778	12620960	64.68%	12328375
	SOX2 ChIP-seq	Untreated	1	27958366	18639635	66.67%	17846180
			2	21596359	14313756	66.28%	13843049
			3	23325172	15900847	68.17%	15371737
		Doxycycline	1	18785372	13006165	69.24%	12673007
			2	25128867	17390037	69.20%	16464012
			3	22596056	15145965	67.03%	14418250
	NANOG ChIP-seq	Untreated	1	23870443	15255641	63.91%	14743761
			2	23383799	14676848	62.77%	14184705
			3	20658117	13122815	63.52%	12697300
		Doxycycline	1	22197976	14705713	66.25%	14127679
			2	27089309	15065366	55.61%	14393230
			3	26579864	16420068	61.78%	15695760
	ChIP Input	Untreated		28330985	18337136	64.72%	17845438
	Nuclear RNA-seq	Untreated	1	62955752	43122979	68.50%	34120238
			2	57456184	42213418	73.47%	34660057
			3	49046413	35981994	73.36%	28244942
		Doxycycline	1	48034562	35018203	72.90%	27382938
			2	46835948	34637361	73.95%	27318280
			3	51222331	33441346	65.29%	25487682
Brg1$^{fl/fl}$ ESCs	ATAC-seq	Untreated	1	27331438	17284030	63.24%	16946288
			2	28086165	16816386	59.87%	16287655
			3	20278277	12085389	59.60%	11752748
		Tamoxifen	1	23137532	13978446	60.41%	13579893
			2	12484092	7533943	60.35%	7289985
			3	18398537	11169126	60.71%	10798092
	OCT4 ChIP-seq	Untreated	1	19526130	12957214	66.36%	12664258
			2	18506763	12162854	65.72%	11865956
			3	19733956	11745225	59.52%	11306863
		Tamoxifen	1	20108099	12696319	63.14%	12415346
			2	20065413	12303247	61.32%	11991638
			3	24357798	13928201	57.18%	13471843
	SOX2 ChIP-seq	Untreated	1	22127599	15254285	68.94%	14931636
			2	20879063	14252712	68.26%	13929111
			3	18974734	11749502	61.92%	11422860
		Tamoxifen	1	21537686	13367922	62.07%	13005694
			2	16152069	10051163	62.23%	9755848
			3	18251129	10724175	58.76%	10432074
	ChIP Input	Untreated		27828227	18527234	66.58%	18091711
	Nuclear RNA-seq	Untreated	1	52217013	37175280	71.19%	31963445
			2	54194240	41100159	75.84%	36562061
			3	49509329	36431490	73.59%	31595013
		Tamoxifen	1	37670761	27671580	73.46%	23561695
			2	53699763	40091884	74.66%	32473392
			3	47202120	34400253	72.88%	30084606

Table 7 Calibrated ATAC-seq samples and sequencing depth.

Summary of sequencing statistics for calibrated ATAC-seq experiments, detailing quantitation of mouse (mm10) and spike-in (dm6) genomes used for normalisation.

Experiment	Sample	Rep	Read Number	mm10 Aligned Reads		dm6 Aligned Reads		Normalised Read Count
KDM2A/B-SF and -LF deletion	*Kdm2.LF*$^{fl/fl}$ Untreated	1	24416850	34175630	98.85%	396471	1.15%	17087815
		2	26535006	37624423	98.63%	523830	1.37%	15673224
		3	28071246	38180936	99.64%	136363	0.36%	13251575
	Kdm2.LF$^{fl/fl}$ Tamoxifen	1	25221234	35016311	98.89%	389916	1.10%	15534761
		2	26676522	37380277	98.60%	529553	1.40%	12576946
		3	23688648	32311471	99.68%	100628	0.31%	13382668
	Kdm2.LF$^{fl/fl}$*;SF*$^{-/-}$ Untreated	1	32585347	45581396	98.94%	485849	1.05%	13048197
		2	24841554	33874255	97.59%	835177	2.41%	10531001
		3	27514374	37738459	99.57%	161452	0.43%	11375187
	Kdm2.LF$^{fl/fl}$*;SF*$^{-/-}$ Tamoxifen	1	27119670	37280270	98.43%	596405	1.57%	12107436
		2	28281138	38599018	97.40%	1030754	2.60%	10878375
		3	25717345	35317023	99.59%	143927	0.41%	12352880
	Kdm2.LF$^{fl/fl}$ Untreated gDNA control	1	5442630	6863811	98.18%	127140	1.82%	2819012
		2	7011920	8404172	98.00%	171704	2.00%	2813170
		3	6902476	8379747	99.57%	36536	0.43%	2857911
	Kdm2.LF$^{fl/fl}$ Tamoxifen gDNA control	1	4535437	5649575	98.41%	91110	1.59%	2824788
		2	6082580	7449912	98.36%	123786	1.63%	2823575
		3	4983715	6062763	99.62%	23269	0.38%	2859254
	Kdm2.LF$^{fl/fl}$*;SF*$^{-/-}$ Untreated gDNA control	1	5856914	7309807	98.72%	94488	1.28%	2834086
		2	5590088	6874500	97.62%	167770	2.38%	2802002
		3	5276295	6554292	99.55%	29394	0.45%	2857068
	Kdm2.LF$^{fl/fl}$*;SF*$^{-/-}$ Tamoxifen gDNA control	1	6909822	8440833	98.22%	152695	1.78%	2818702
		2	6714107	8286782	97.33%	226984	2.67%	2793279
		3	7434146	8736125	99.54%	40537	0.46%	2856700

CHAPTER 3 - EXAMINING CHROMATIN ACCESSIBILITY IN EMBRYONIC STEM CELLS

3.1 Introduction

CpG islands (CGI) are found at the majority of gene promoters in vertebrate systems (Saxonov et al., 2006, Long et al., 2013b), and their lack of DNA methylation and elevated CpG density appear to be important for the regulation of transcription (Blackledge and Klose, 2011, Deaton and Bird, 2011). One central characteristic of CGIs is that they exhibit high levels of chromatin accessibility (Antequera et al., 1990, Tazi and Bird, 1990, Sabo et al., 2004, Crawford et al., 2006) and nucleosome depletion (Li et al., 2011b, Fenouil et al., 2012). This accessibility to underlying DNA sequences in the context of chromatin may be important for facilitating sequence-specific transcription factors or the transcriptional machinery to engage with CGIs as platforms for the regulation of gene expression (Cockerill, 2011). However, the molecular features of CGI chromatin that are responsible for this elevated accessibility have yet to be carefully examined.

There are several approaches to measuring chromatin structure and accessibility in living cells, and many of these have been adapted recently to use massive parallel sequencing technologies to profile chromatin accessibility at the genome-scale (Sabo et al., 2004, Crawford et al., 2006, Li et al., 2011b, Simon et al., 2012). The majority of these approaches rely upon incubation of nuclei with nucleases, in order to measure the nuclease sensitivity of regions of the genome, such as for the widely used deoxyribonuclease I (DNase I) (Song and Crawford, 2010, John et al., 2013) or micrococcal nuclease (MNase) (Li et al., 2011b). Importantly, while DNase I is able to cleave both intra- and internucleosomal DNA, MNase only cleaves DNA between nucleosomes. MNase followed by massive parallel sequencing (MNase-seq) therefore provides useful information for nucleosome occupancy and positioning, while DNase I

digestion followed by massive parallel sequencing (DNase-seq) can provide a more general measure of DNA accessibility to nucleases and other *trans*-acting factors (Cockerill, 2011). DNase-seq also seemed advantageous to other chromatin accessibility assays, such as formaldehyde assisted isolation of regulatory elements (FAIRE), which, following formaldehyde fixation and sonication, enriches for DNA fragments lacking protein-DNA crosslinks (Giresi et al., 2007, Simon et al., 2012). Although this removes all nucleosome-bound DNA and enriches for nucleosome-free, accessible DNA regions (Song et al., 2011), regions highly bound by other proteins such as transcription factors can be excluded, even though they may display nuclease hypersensitivity. A further advantage for DNase-seq was that it had been successfully applied to measure differential chromatin accessibility during development or in response to hormone treatments (Ling et al., 2010, Biddie et al., 2011a, He et al., 2012, Grøntved et al., 2013), indicating that it would be suitable for quantitative measurement and comparison of chromatin accessibility at CGIs during the genetic ablation experiments that I had planned. Therefore, I set out to optimise and apply DNase-seq to a cell culture system that is often used in the study of CGIs, mouse embryonic stem cells (ESCs).

3.2 Results

3.2.1 Application of DNase-seq to mouse embryonic stem cells

Although DNase I hypersensitivity was one of the earliest measures of chromatin accessibility to be established (Wu, 1980), only a handful of groups have managed to implement its application for measuring chromatin accessibility genome-wide (Song and Crawford, 2010, John et al., 2013). One of the reasons for this is that this technique relies upon a series of experimental variables that must be optimised for each cell type of interest. I therefore performed a series of experiments with the aim to optimise the DNase-seq protocol for the application to mouse ESCs.

3.2.1.1 Optimisation of nuclei isolation for DNase-seq

One of the essential steps required for DNase I profiling of chromatin accessibility is the efficient isolation of high quality nuclei prior to DNase I digestion, a process which can be highly cell type-dependent (John et al., 2013). As such, I set out to optimise the preparation of nuclei from mouse ESCs. One of the main determinants for the preparation of high quality nuclei is the concentration of the detergent agent used to permeabilise the membrane. Therefore, to determine the optimal concentration of detergent for the isolation of nuclei from mouse ESCs, a range of non-ionic detergent concentrations were tested (0 to 0.1% NP-40) (Figure 3.2.1). The resulting nuclei were examined by differential interference contrast (DIC) microscopy to visualise nuclear integrity and the efficiency of cytoplasm removal. By DIC, darker and denser nuclei were surrounded by bright cytoplasm, which could not be observed in pre-lysis samples. At low NP-40 concentrations (0.01 or 0.02%), the cytoplasm could still be observed surrounding the nuclei. At intermediate concentrations (0.03 to 0.05%), nearly all cytoplasm was removed, and nuclear integrity appeared to remain intact. At higher concentrations (0.06 to 0.1%), extensive cellular debris was apparent, and nuclei had a swollen morphology. Based on these qualitative assessments, a NP-40 concentration of 0.03% appeared to provide an efficient nuclei isolation, while still maintaining nuclei integrity. Therefore, this concentration of NP-40 was used for subsequent application to the DNase-seq protocol.

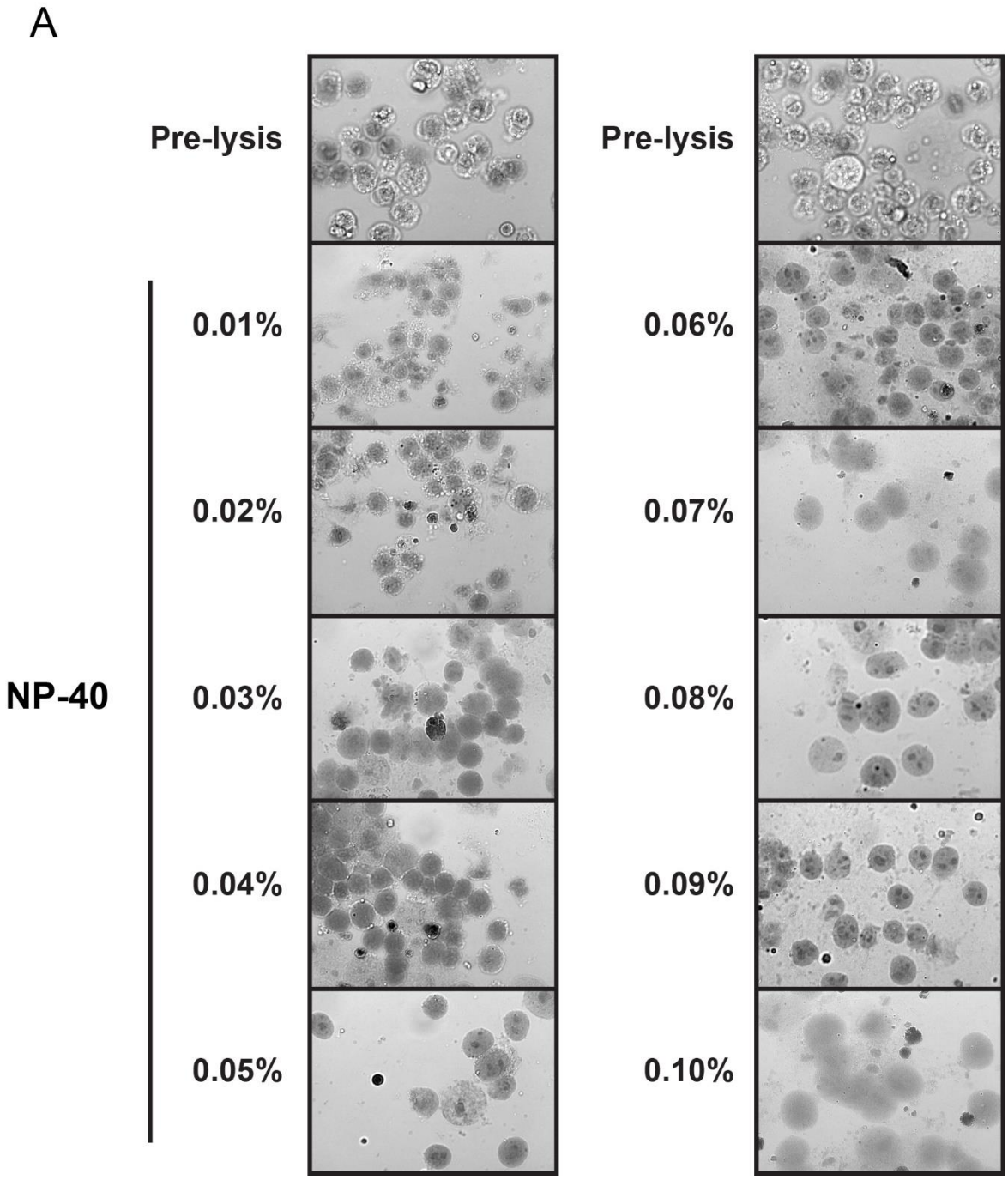


Figure 3.2.1 Optimisation of nuclei isolation from mouse embryonic stem cells.

A) E14 mouse ESCs were subjected to nuclei isolation, with increasing concentrations of NP-40, stained with Trypan blue and observed by DIC microscopy.

3.2.1.2 Qualitative and quantitative assessment of DNase I digestion

Having established a nuclei isolation protocol, I then performed a digestion curve with varying concentrations of DNase I on isolated mouse ESC nuclei to optimise the detection of DNase I hypersensitivity at regulatory elements in the genome (Figure 3.2.2). After incubation with DNase I, DNA was isolated and subjected to agarose gel electrophoresis. This revealed increased migration of bulk genomic DNA with increasing concentrations of DNase I, in addition to the release of short (0.1 to 0.5 kb), double-cut DNA fragments (Figure 3.2.2 A). Although this allowed for a qualitative assessment of DNase I digestion efficiency, a more quantitative approach would be preferable. Therefore, the DNase I-digested samples from the digestion curve was used for a quantitative real-time PCR (qPCR) assay known as DNase-qPCR. This type of analysis measures the amount of DNase I digestion at DNase I hypersensitive sites (DHS) relative to a control, DNase I-insensitive genomic region (McArthur et al., 2001, John et al., 2013). By measuring the amplification of DHS primer sets in DNase I-digested samples relative to an undigested control, the digestion efficiency of DNase I (ie DNase I sensitivity) can be calculated. As expected, DNase-qPCR of the ESC digestion curve at two gene promoters and one enhancer DHS revealed higher digestion efficiencies with the use of higher DNase I concentrations (Figure 3.2.2 B). The digestion efficiencies that I observed for the two gene promoters (40-90%) were comparable with digestion efficiencies recommended for DNase-seq experiments (John et al., 2013). In addition to confirming the successful digestion of nuclei with DNase I, the DNase-qPCR results indicated that my samples were potentially suitable for DNase-seq library generation.

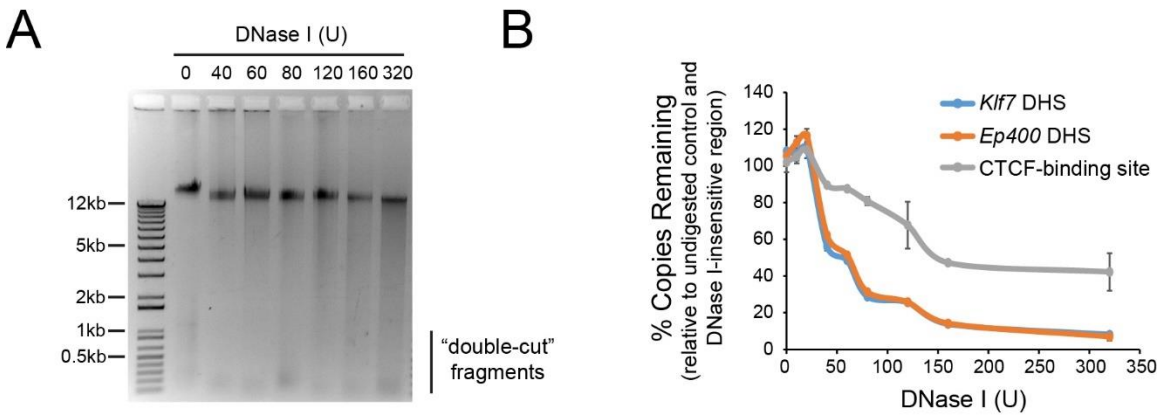


Figure 3.2.2 Electrophoresis and DNase-qPCR of ESC DNase I digestion curve.

A) Agarose gel (1%) electrophoresis of genomic DNA isolated from mouse ESC nuclei digested with varying amounts of DNase I.

B) DNase-qPCR of mouse ESC DNase digestion curve. Primers were designed within strong DHS at the *Klf7* and *Ep400* promoters, and a CTCF-binding site with a moderate level of accessibility, and normalised to a control genomic region.

3.2.1.3 Application of DNase-seq to mouse ESCs

To understand how the values from the DNase-qPCR experiment related to downstream DNase-seq data quality, DNase-seq libraries were prepared for eight different DNase I digestions by size-selecting DNase-I double-cut fragments prior to adaptor ligation and massively parallel sequencing (Figure 3.2.3), and compared with a previously published mouse ESC DNase-seq experiment (Thurman et al., 2012). This identified DNase I hypersensitivity at several gene promoters, as measured by an enrichment of DNase-seq reads at these loci (Figure 3.2.3 A). This enrichment was highly dependent on the DNase I digestion efficiency. For example, samples digested with 40 to 160U DNase I appeared to show reasonable enrichment of DNase-seq signal at DHS, with 80U generating the highest data quality. In contrast, samples with low digestion efficiencies (10U and 20U) showing little enrichment at DHS, as did the most digested sample (320U), suggesting under- and over-digestion of the nuclei respectively. Although I could detect DNase I hypersensitivity at gene promoters similar to the previously published mouse ESC DNase-seq experiment (Figure 3.2.3 A), visual examination of my DNase-seq data revealed that some DHS showed little or no enrichment at any of the DNase I concentrations that I

tested (Figure 3.2.3 B). These differences were particularly pronounced at distal regulatory elements, as demonstrated by a comparison of DNase-seq signal at promoters or enhancers, which revealed that my assay captured promoter accessibility very well, but performed poorly in detecting accessibility at enhancer DHS (Figure 3.2.3 C). This contrasted the ENCODE DNase-seq which captured these sites equally well (Figure 3.2.3 D). This failure to robustly capture chromatin accessibility at a significant number of DHS throughout the genome suggested a substantial problem with my DNase-seq protocol that would require further optimisation.

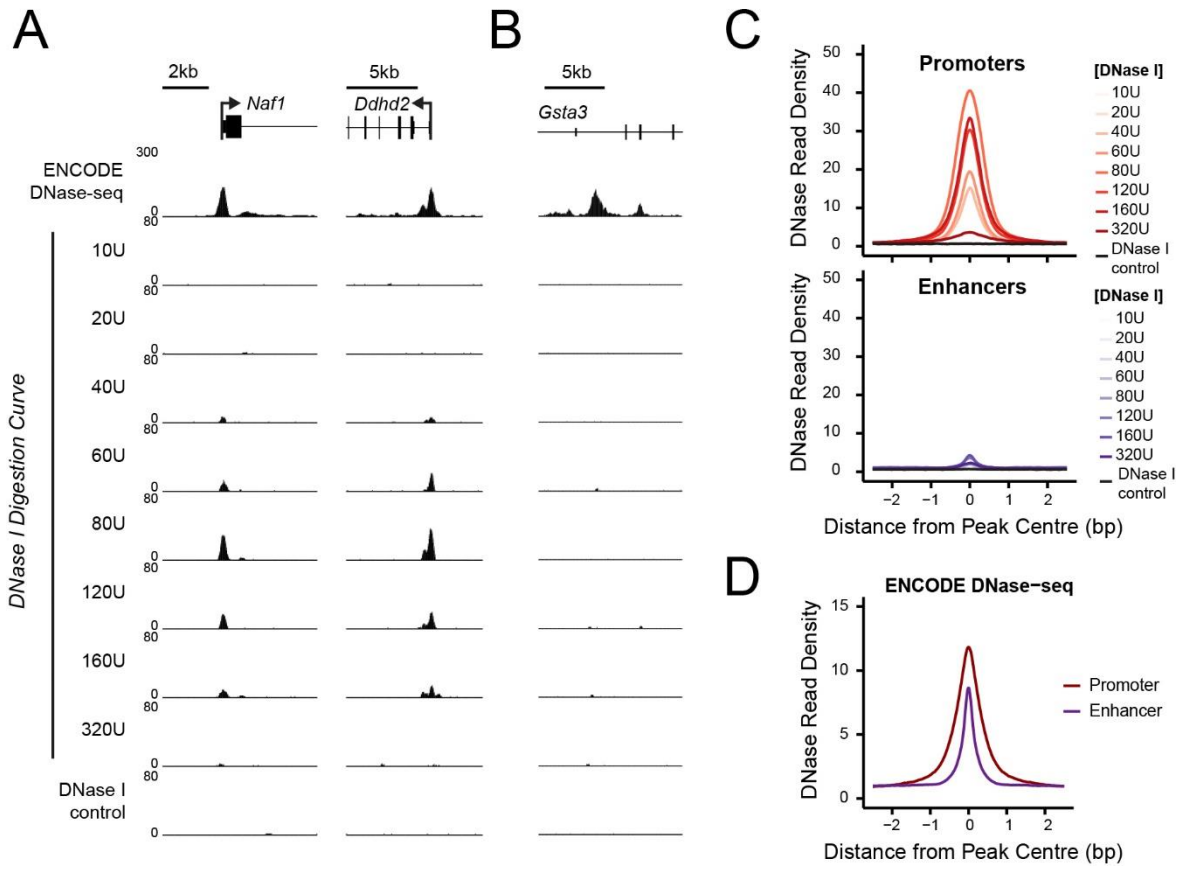


Figure 3.2.3 DNase I digestion curve of mouse ESCs followed by sequencing.

A) DNase-seq profiles of the DNase I digestion curve, compared to a previously published dataset by the ENCODE project. DNase I control represents genomic DNA digested with DNase I and provides measure of sequence bias.

B) Same as (A) but for a distal regulatory site which DNase I hypersensitivity is not detected in my digestion curve, but is present in the ENCODE DNase-seq dataset.

C) DNase I hypersensitive sites (DHS) identified in the mouse ESC ENCODE DNase-seq dataset were annotated as promoters or enhancers based on H3K4me3 and H3K4me1 (see Methods) and the average profiles of the ESC DNase-seq digestion curve were plotted.

D) Mouse ESC DNase-seq from ENCODE at promoter or enhancer DHS peaks.

3.2.1.4 An examination of DNase-seq library fragment sizes

The preparation of DNase-seq libraries requires the purification and size-selection of small (50-250bp) fragments released from DHS by DNase I digestion. Selection of different size fragments could potentially lead to differences in DNase-seq signal in quantitative (eg. signal-to-noise) and qualitative (eg. types of regulatory elements detected) terms (Buenrostro et al., 2013, Vierstra et al., 2014). To examine this possibility, the 80U digestion sample was used to make five DNase-seq libraries, with different median sizes (75, 100, 125, 150, or 175 bp) (Figure 3.2.4 A; Fractions A-E respectively). Sequencing and analysis of the five DNase-seq libraries revealed DNase I hypersensitivity at promoter sites in all five size ranges (Figure 3.2.4 B and D). Interestingly, at some sites the accessibility varied depending on library fragment size, with either a larger enrichment with smaller or larger fragments (Figure 3.2.4 B). However, there were still DHS which were not detected in any of the size fraction libraries (Figure 3.2.4 C), and similarly to the previous experiment, this was particularly evident at enhancer DHSs (Figure 3.2.4 D), regardless of the library fragment size. Therefore, the issues with detecting enhancer DHS could not be explained by fragment size biases, and further optimisation of the assay was required.

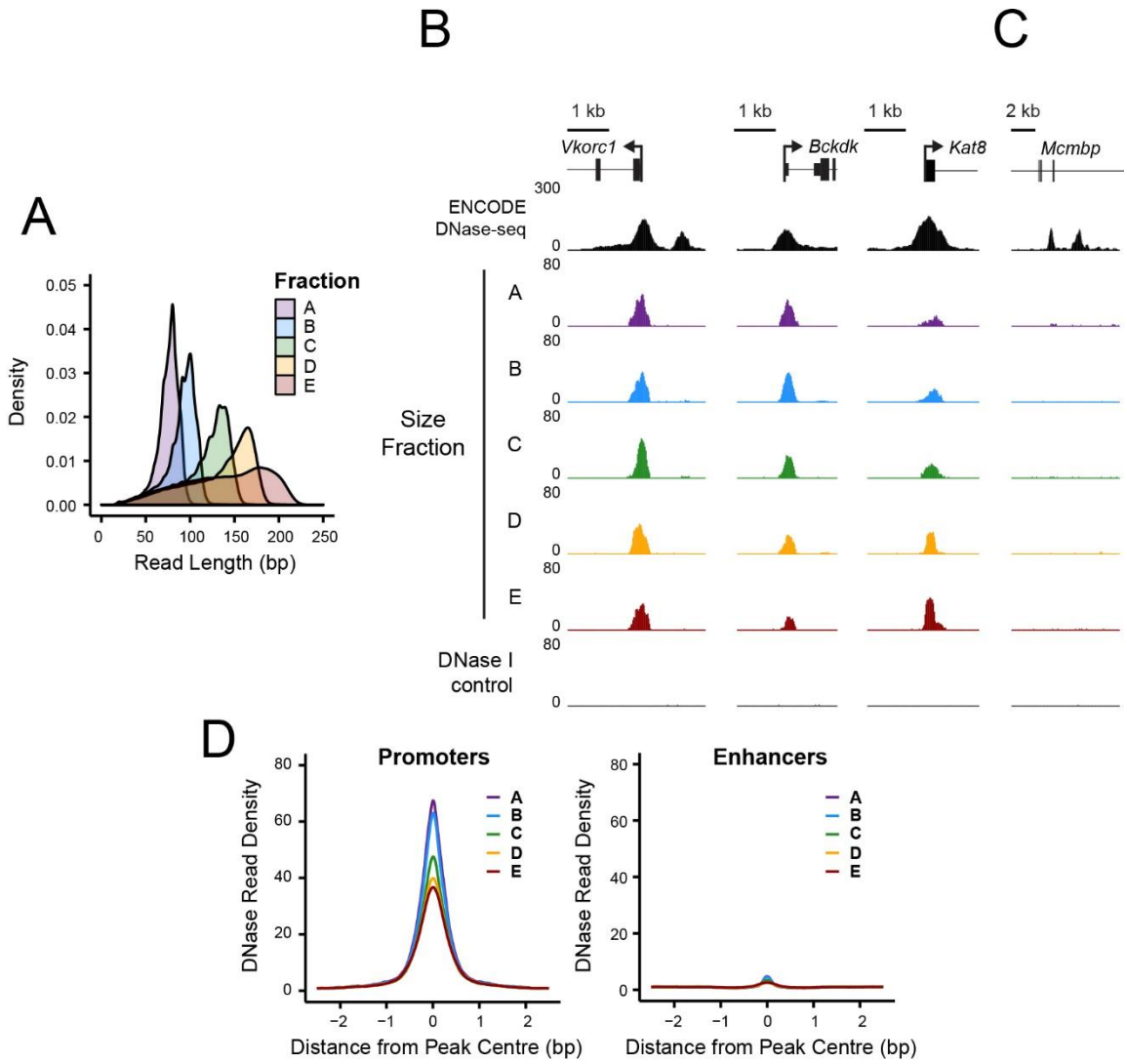


Figure 3.2.4 Variation of size selection for DNase-seq library preparation.

A) Read length distribution after sequencing of DNase-seq libraries subjected to different size selection parameters. Ion Proton sequencing platform has a maximum read length of 200bp, including adaptor sequences, and as such libraries with larger sizes demonstrate lower quality reads which are automatically trimmed as part of the Ion Proton basecalling and filtering algorithm.

B) Genomic snapshots of DNase-seq libraries with different fragment sizes.

C) Genomic snapshot of distal DHS which lack DNase I hypersensitivity in DNase-seq libraries.

D) Metaplots of different fragment size DNase-seq libraries profiled over promoter or enhancer DHS.

3.2.1.5 Altered nuclei isolation aids detection of enhancer DHS

The initial optimisation step that I performed to optimise the DNase-seq approach involved the comparison of different nuclei isolation protocols (Figure 3.2.1). As a result of the qualitative comparison performed, a NP-40 concentration of 0.03% was used for subsequent DNase-seq experiments. However, due to the purely subjective means of arriving at this decision, it was possible that nuclei isolation was not actually optimal for subsequent DNase I digestion and sequencing. To examine this possibility, nuclei were isolated from mouse ESCs with a slightly higher concentration of NP-40 (0.05%), followed by DNase I digestion with a range of DNase I concentrations, library preparation and sequencing. Examination of subsequent DNase-seq profiles identified that DNase I hypersensitivity was captured equally well at both promoters and enhancers (Figure 3.2.5 A,B), and that it was highly correlated with previously published mouse ESC DNase-seq (Figure 3.2.5 C). A direct comparison of the different digestion points showed that the 90U sample had a slightly higher enrichment at DHS than the other samples, although the differences appeared to be marginal (Figure 3.2.5 B). The alteration in NP-40 concentration for the nuclei isolation therefore resulted in a substantial difference in the DNase-seq profiles. Although it is difficult to draw definitive conclusions about the reason for this, the most likely explanation seems to be that the initial NP-40 concentration (0.03%) was insufficient to permeabilise cellular and nuclear membranes to allow necessary levels of DNase I to enter the nuclei, allowing only digestion of the most hypersensitive genomic regions (ie promoters). Now that I had established optimal conditions for DNase-seq, it was then important to establish whether it could be performed in a robust and reproducible manner.

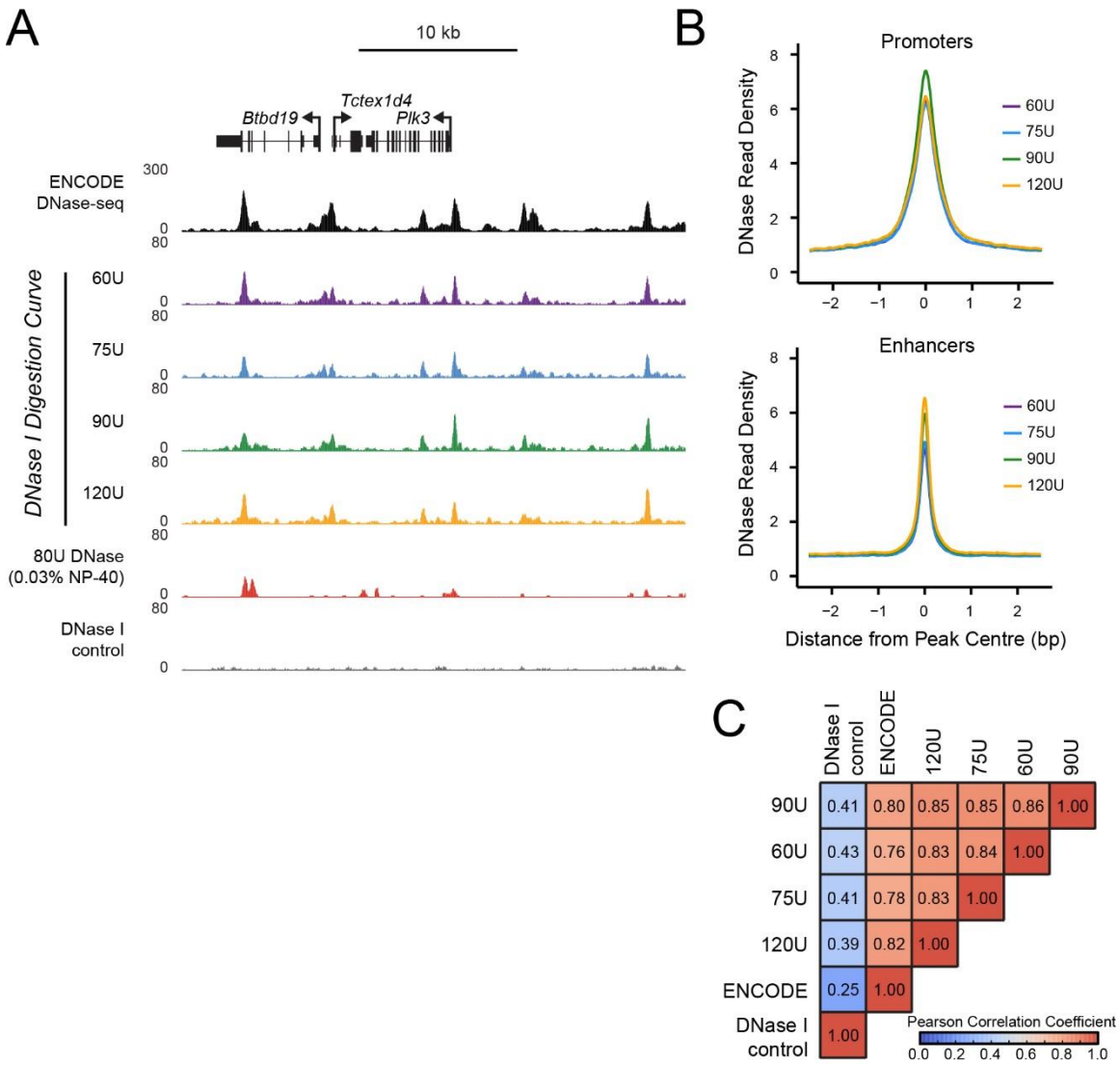


Figure 3.2.5 DNase I digestion of mouse ESCs with altered NP-40 concentration.

A) Genomic snapshots of mouse ESC DNase-seq digestion curve at various DHS. Previous DNase-seq experiment (80U DNase 0.03% NP-40) is included for comparison.

B) Metaplot of DNase-seq libraries profiled over promoter or enhancer DHS.

C) Pearson correlation matrix of DNase-seq libraries using 1kb windows across the mouse genome. Mouse ESC DNase-seq from ENCODE is provided as a reference.

3.2.1.6 DNase-seq profiles are not robust across biological replicates

In order to examine the molecular determinants of chromatin accessibility at CGIs it was essential to have a robust and quantitative chromatin accessibility assay. Therefore, I sought to test the reproducibility of the DNase-seq protocol by performing two biological replicates with mouse ESCs using a range of DNase I concentrations (Figure 3.2.6). Unfortunately, visualisation of the two repeats showed radically different profiles for the DNase-seq digestion curves (Figure 3.2.6 A), with the second biological replicate displaying very low enrichment at DHS. Higher concentrations of DNase I produced slightly better enrichment of DNase-seq signal for the second replicate (Figure 3.2.6 A,B), but these profiles were still of insufficient signal-to-noise compared to its replicate (Figure 3.2.6 A,B). In fact, I performed DNase-seq on several additional occasions, with similar variation between biological replicates (data not provided). Indeed, few laboratories have managed to perform quantitative and reproducible DNase-seq, and perhaps my inability to achieve this reflects the difficulty other groups have had in establishing this protocol. Therefore, despite having successfully generated high quality DNase-seq profiles (Figure 3.2.5), the fact that I was unable to reproduce this in a robust manner meant that this approach would be unsuitable for measuring quantitative differences in chromatin accessibility. Ultimately, I would need a robust and sensitive chromatin accessibility assay to examine the contribution of different chromatin-associated pathways towards chromatin accessibility at CGI promoters.

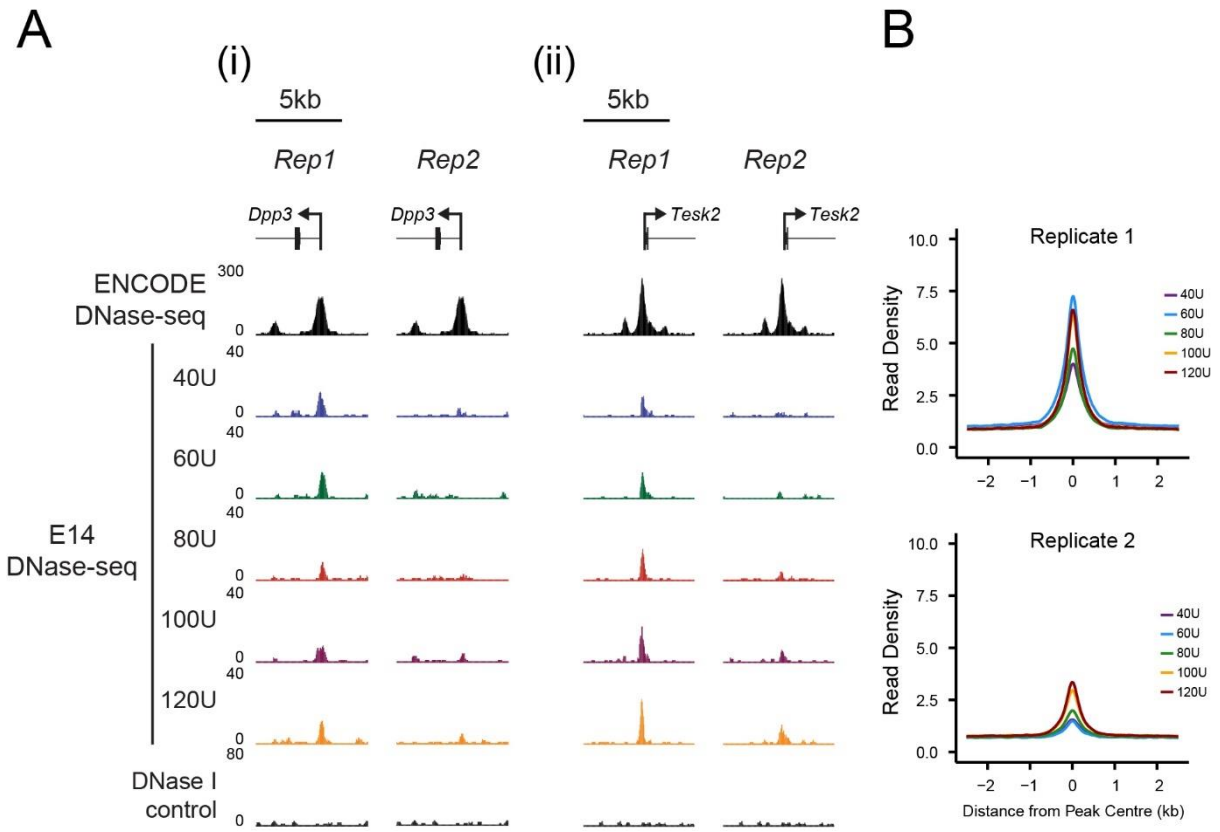


Figure 3.2.6 Failure to robustly reproduce DNase-seq profiles.

A) Genomic snapshots of biological replicates (Rep1 and Rep2) of mouse ESC DNase-seq at the *Dpp3* (i) and *Tesk2* (ii) promoters.

B) Metaplot analysis biological replicate DNase-seq libraries at DHS from the DNase ENCODE dataset.

3.2.2 Using ATAC-seq to measure chromatin accessibility

The difficulty in generating reproducible accessibility profiles with DNase-seq meant that it was necessary to investigate alternative measures of chromatin accessibility to study CGI chromatin structure and accessibility. While several alternatives to DNase-seq exist (eg. FAIRE, Sono-Seq, RED-seq), one of the most attractive was ATAC-seq (assay for transposase-accessible chromatin using sequencing). This protocol was first described while I was attempting to establish DNase-seq (Buenrostro et al., 2013), and similarly to DNase-seq, relies upon the hypersensitivity of genomic regions to nuclease digestion, in this case the Tn5 transposase. ATAC-seq also provides several additional advantages compared to DNase-seq, including requiring fewer cells, and a faster and simpler experimental protocol. Therefore, I sought to explore the use of Tn5 transposase and the ATAC-seq protocol in order to investigate the determinants of chromatin accessibility.

The Tn5 transposase has been widely applied to molecular biology applications in recent years (Reznikoff, 2008). The transposase typically used in such applications recognises and binds two highly specific 19bp insertion sequences, which have been modified from their endogenous forms to increase *in vitro* Tn5 activity (Reznikoff, 2008). A complex of Tn5 with these hyperactive “mosaic end” (ME) sequences is usually sufficient for transposition of other DNA species. During transposition, the Tn5 uses a “cut and paste” mechanism, whereby it binds and cleaves the target DNA, simultaneously joining the target DNA with the ME sequences. This process is also known as tagmentation, because it both fragments and adds an adaptor sequence (or tag) to DNA. Tagmentation is now commonly used for *in vitro* library generation (Adey et al., 2010). The same principle applies for ATAC-seq, only in the context of nuclei, whereby incubation of nuclei with the Tn5 transposase results in the preferential transposition of the Tn5 ME sequences into open, accessible chromatin (Buenrostro et al., 2013, Buenrostro et al., 2015). This tagmented DNA is then amplified by PCR using primers complementary to the ME sequences. It is at this stage that sequencing adaptor and barcode sequences can be

added, allowing massively parallel sequencing of the ATAC-seq library. ATAC-seq profiles therefore represent Tn5 hypersensitivity throughout the genome, and are highly comparable with DNase I hypersensitivity (Buenrostro et al., 2013). In order to examine whether ATAC-seq would indeed provide a useful experimental approach, I needed to test the ATAC-seq protocol using the Tn5 transposase. Although a commercially available form of the Tn5 transposase is available, I decided it would be more cost effective and eliminate batch-to-batch variability to produce the transposase myself.

3.2.2.1 Expression and purification of Tn5 transposase in *E. coli*

To express the Tn5 transposase, I used an expression construct containing a hyperactive Tn5 allele (E54K and L372P mutations) with a C-terminal intein tag and chitin-binding domain (CBD), which has been previously described (Picelli et al., 2014) (Figure 3.2.7 A). Following expression of the protein in *Escherichia coli* and lysis of the cells, the cell lysate was passed over a chitin column, allowing the intein-CBD-Tn5 fusion to bind. The column was then incubated with an excess of mosaic end double-stranded (MEDS) oligonucleotides, before rigorous washes to remove any unbound oligonucleotides which might contaminate subsequent purification steps. Thiol-mediated cleavage was used to elute the Tn5-MEDS complexes, and the subsequent elution fractions were analysed by SDS-PAGE following by Coomassie staining. This showed an enrichment of a ~55kD protein in the elution fractions (Figure 3.2.7 C), which corresponded to the expected size for the Tn5 transposase (Picelli et al., 2014). Having purified the Tn5 transposase, it was then essential to examine whether it was active. To test this, I incubated the purified Tn5 with genomic DNA. This resulted in the digestion of the genomic DNA to an average size of 100bp (Figure 3.2.7 D), consistent with previously described tagmentation experiments (Picelli et al., 2014), and confirmed that I had purified an active form of the Tn5 transposase. This meant that I could proceed to testing its application to ATAC-seq in ESCs.

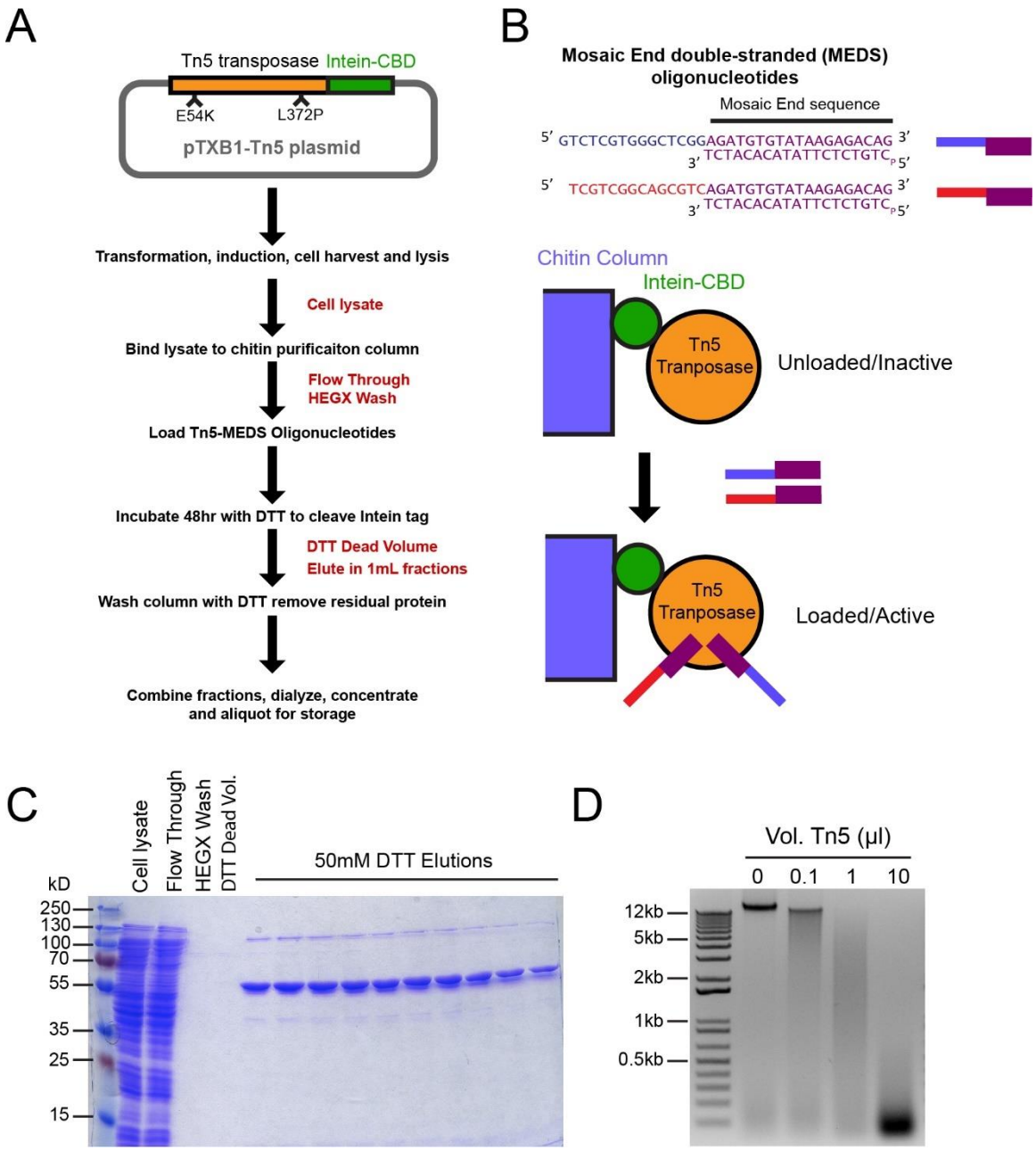


Figure 3.2.7 Expression and purification of Tn5 transposase.

A) Schematic for expression and purification of Tn5 transposase in *E. coli*. Those steps depicted in red correspond to fractions loaded on the Coomassie-stained gel in (C).

B) Schematic depicting the loading of Tn5 transposase with mosaic end oligonucleotides during column purification of Tn5.

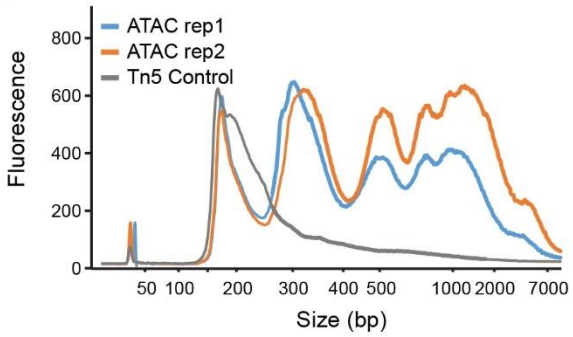
C) Coomassie-stained SDS-PAGE gel of Tn5 purification and elution fractions.

D) Agarose gel electrophoresis of mouse genomic DNA digested with varying amounts of custom-made Tn5 transposase.

3.2.2.2 Application of purified Tn5 transposase to ATAC-seq of mouse ESCs

Having purified the active Tn5 transposase, it was important to establish whether it could be used to measure chromatin accessibility with the ATAC-seq protocol. To this end, mouse ESC nuclei were incubated with varying amounts of the Tn5 transposase, in biological duplicate, followed by DNA purification and PCR amplification with primers specific to the Tn5 ME sequences to make ATAC-seq libraries. As a control, a Tn5 control library was made by tagmenting mouse genomic DNA to account for any sequencing or Tn5-specific biases. Examination of the DNA fragment sizes of ATAC-seq libraries (which represent tagmented DNA plus 132bp of transposase ME and adaptor sequences) revealed an apparent nucleosomal periodicity of approximately 180 bp (Buenrostro et al., 2015). This periodicity was not apparent in the Tn5 control library, confirming that the periodicity observed in the ESC ATAC samples was a digestion profile of nucleosomal chromatin (Figure 3.2.8 A). These ATAC-seq libraries were then sequenced on the IonProton platform to examine chromatin accessibility genome-wide. However, the alignment statistics from this sequencing experiment were very poor, with approximately 90% of reads failing to map to the mouse genome (Figure 3.2.8 B). Nevertheless, it was possible to observe enrichment for ATAC-seq signal at similar loci to that of DNase-seq (Figure 3.2.8 C), suggesting that ATAC-seq was measuring chromatin accessibility successfully. Although the low sequencing depth made further analysis difficult, quantitation of ATAC-seq enrichment in DHS intervals suggested that the signal-to-noise of the ATAC-seq libraries was similar or better than that of the DNase-seq library used for comparison (Figure 3.2.8 D). Furthermore, these enrichment values were consistent across Tn5 concentrations biological replicates, suggesting that ATAC-seq is a highly reproducible assay, even across different tagmentation conditions, something that was not observed for DNase I digestions.

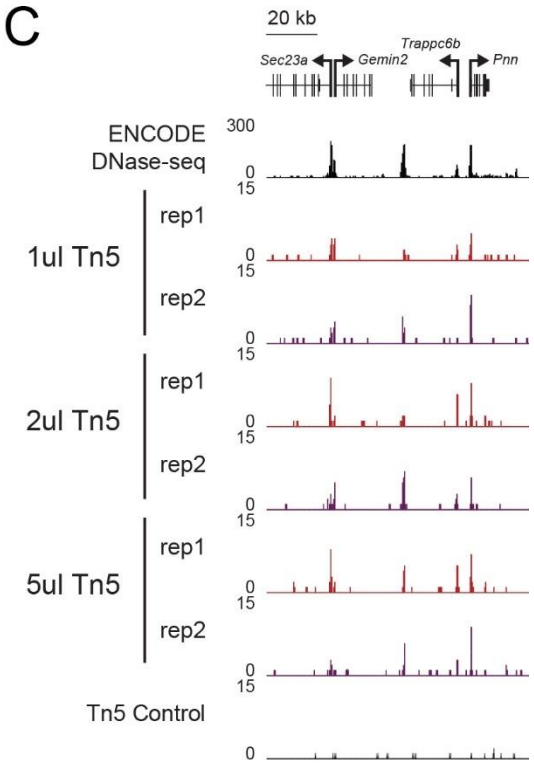
A



B

Name	Processed Reads	Failed To Align	
1ul Tn5 rep1	8177236	7471652	**91.37%**
1ul Tn5 rep2	7410254	6698645	**90.40%**
2ul Tn5 rep1	8430204	7648042	**90.72%**
2ul Tn5 rep2	12868760	11784820	**91.58%**
5ul Tn5 rep1	9679298	8899200	**91.94%**
5ul Tn5 rep2	8327317	7527700	**90.40%**
Tn5 Control	9658472	8862409	**91.76%**

C



D

Sample	% Reads within Peaks
1ul Tn5 rep1	30.26%
1ul Tn5 rep2	25.96%
2ul Tn5 rep1	30.38%
2ul Tn5 rep2	29.28%
5ul Tn5 rep1	30.72%
5ul Tn5 rep2	33.84%
Tn5 Control	3.35%
ENCODE Dnase	20.65%

Figure 3.2.8 Trial ATAC-seq experiment profiles chromatin accessibility, but with poor alignment statistics.

A) Bioanalyser fragment size profiles of two mouse ESC ATAC-seq libraries (rep1 and rep2) compared with a control ATAC-seq library made with Tn5-digested genomic DNA.

B) Alignment statistics for IonProton ATAC-seq.

C) Genomic snapshot of ATAC-seq libraries sequenced on the IonProton platform.

D) Signal-enrichment of ATAC-seq in ENCODE DHS intervals.

Despite several attempts to understand the poor alignment statistics of libraries sequenced with the IonProton, I was unable resolve this issue. However, I speculated that this may have been a compatibility issue between the Tn5 tagmented libraries and the IonProton sequencer. Therefore, I generated new ESC ATAC-seq libraries in biological triplicate that would be compatible with the Illumina sequencing platform, and sequenced them on the Illumina NextSeq500 (Figure 3.2.9). In contrast to those sequenced on the IonProton, ATAC-seq libraries sequenced on the Illumina platform had an alignment rate of approximately ~70%. Furthermore, the quality of the chromatin accessibility profiles, and the reproducibility between biological replicates, was excellent (Figure 3.2.9 A,B), allowing for the identification and annotation of ATAC hypersensitive sites (AHS) (Figure 3.2.9 C). As previously described, ATAC-seq was well correlated with DNase-seq, verifying its use as a measure of chromatin accessibility (Figure 3.2.9 B). The high quality and robustness of the mouse ESC ATAC-seq gave confidence that this approach could be taken forward for further experiments examining the determinants of CGI chromatin accessibility in mouse ESCs.

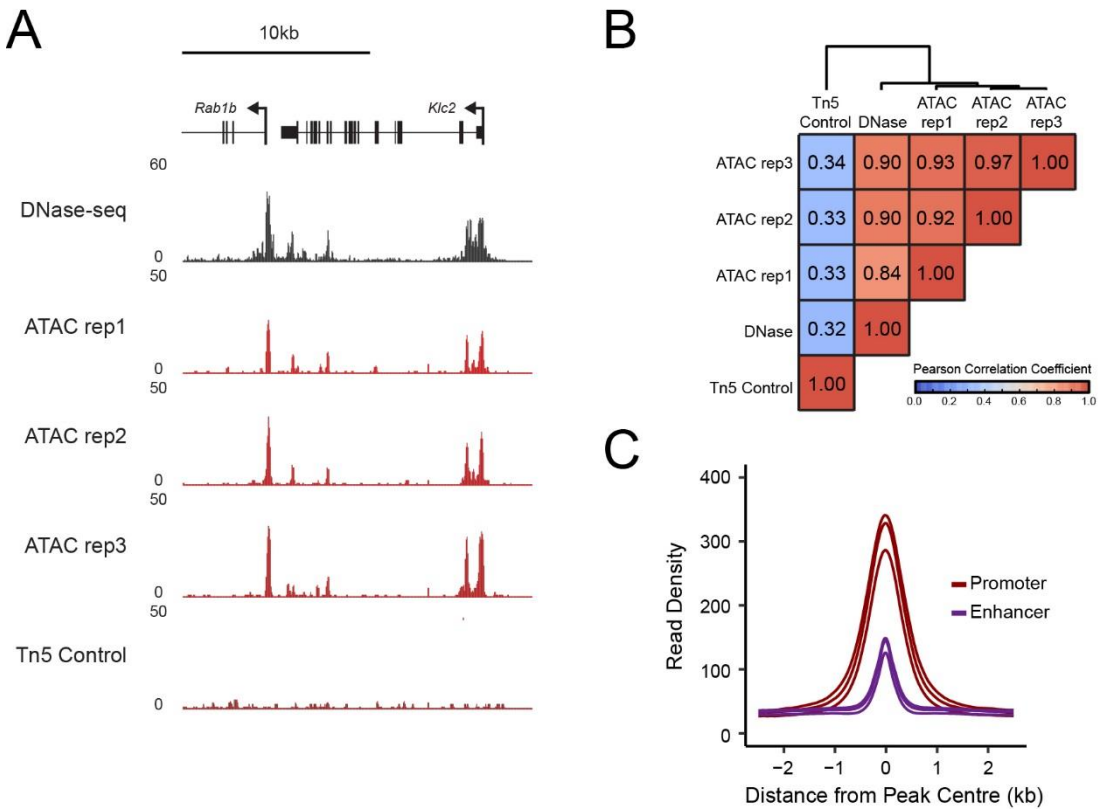


Figure 3.2.9 ATAC-seq is a robust measure of chromatin accessibility.

A) Genomic snapshot of mouse ESC ATAC-seq libraries in biological triplicate.

B) Pearson correlation matrix for 1kb windows of the mouse genome. Biological triplicate for ATAC-seq libraries are compared with ENCODE DNase-seq from mouse ESCs.

C) Metaplot analysis of ATAC-seq at ATAC hypersensitive sites (AHS; n = 100,457) annotated as either promoters or enhancers. Data is merged biological triplicate.

3.3 Discussion

Chromatin accessibility is an interesting and potentially important feature of CGIs throughout the genome. In order to study the causes and consequences of this accessibility I have attempted to optimise and apply two different genome-wide measurements of chromatin accessibility in mouse ESCs. Initial optimisation of ESC nuclei isolation followed by DNase I digestion and DNase-seq appeared to only partially capture chromatin accessibility at a subset of regulatory elements, mainly promoters (Figure 3.2.1, Figure 3.2.3 & Figure 3.2.4). Alteration of the nuclei isolation resolved this problem (Figure 3.2.5), but I was unable to robustly reproduce these DNase-seq profiles (Figure 3.2.6). As an alternative to DNase-seq, I employed a relatively new assay, ATAC-seq, for which I successfully expressed and purified the Tn5 transposase (Figure 3.2.7). In contrast to DNase-seq, I was able to generate robustly reproducible and high quality ATAC-seq profiles in ESCs (Figure 3.2.9), although there were some technical difficulties associated with using the IonProton sequencing platform (Figure 3.2.8). ATAC-seq was therefore deemed to be suitable to take forward and apply to different cell culture systems to examine the determinants of elevated chromatin accessibility at CGIs.

3.3.1 Technical difficulties of genome-wide chromatin accessibility assays

DNase I hypersensitivity has been widely employed by different labs using either Southern blotting (Wu, 1980, Mucha et al., 2000) or DNase-qPCR (McArthur et al., 2001, Ho et al., 2011) to measure chromatin accessibility. However, there are a limited number of labs who have managed to establish robust and reproducible genome-wide applications of this assay (Song and Crawford, 2010, Hosseini et al., 2013, John et al., 2013). Nevertheless, I set out to try and establish this protocol within our laboratory. It quickly became apparent that standardising DNase I concentration and reaction time across biological samples was the main challenge behind this protocol, and it was difficult to reproducibly generate DNase-seq digestion profiles. This was consistent with observations from several other

groups who perform DNase-seq (Maria Suciu personal communications; John et al., 2013), and there seemed to be two main approaches to handle this problem. Firstly, a range of DNase I digestions should be performed with variable DNase I concentrations for a fixed time, and then quantitatively and qualitatively compared using gel electrophoresis and DNase-qPCR to judge the most suitable digestion points of each sample to be used for library preparation (John et al., 2013). This means that two samples performed at precisely the same time may require different enzymatic conditions to be selected in order to achieve consistent data quality, raising concerns about the robustness of this approach for detecting subtle differences between samples. Alternatively, the reactions can be monitored in real time, subjectively judging the viscosity of lysed nuclei at intermittent periods during the reaction time. Once the sample reaches a water-like viscosity, the reactions are stopped and carried forward for qualitative and quantitative assessment (M. Suciu personal communications), as described above. The requirement of these rather extensive quality control measures highlights the inherent variability in DNase I digestion, which likely arises from variation in nuclei quality and quantity between biological samples. Ultimately, these approaches aim to capture the enzymatic condition at which sufficient digestion has occurred to capture digestion of hypersensitive chromatin, but prior to over-digestion, as this will degrade not only bulk “inaccessible” DNA, but also the fragments already cleaved by DNase I in hypersensitive chromatin. While I attempted to empirically optimise conditions to capture the optimal digestion point, it appears as if the inherent variability of digestion efficiencies between biological samples meant that it was going to be challenging to apply this protocol in an suitably robust manner.

In contrast to DNase-seq, ATAC-seq appeared to be inherently robust and reproducible. While optimal DNase I digestion requires a “limited” or incomplete digestion of nuclei, the Tn5 digestion appears to reach saturation, due to the lower number of cells, longer incubation time, and inability of the Tn5 transposase to continually cleave tagmented DNA

(Reznikoff, 2008). Certainly, after many experiments performed using ATAC-seq, it appears to be highly reproducible between biological replicates and different ESC cultures. In general, the majority of variation in chromatin accessibility profiles using ATAC-seq appears to be related to problems with nuclei isolation or number. Unlike other studies which have described high concentrations of mitochondrial DNA in ATAC-seq libraries (Buenrostro et al., 2015), my adapted protocol, which involves washing the nuclei several times prior to Tn5 digestion, typically results in ATAC-seq libraries with less than 10% mitochondrial DNA. Furthermore, in contrast to DNase-seq, ATAC-seq can be performed on 10- to 100-fold fewer cells, requires significantly less time, and can therefore be performed on many more samples at once. The only challenge with the ATAC-seq libraries proved to be the incompatability with the IonProton sequencing platforms, most likely due to the formation of hairpin structures by tagmented DNA libraries during the sequencing reaction (Alex Kurze personal communications). Fortunately, this was not a problem using the Illumina sequencing platforms. Ultimately, ATAC-seq proved to be the most easily applied and reproducible genome-wide chromatin accessibility assay, and I then proceeded to apply this to a range of mouse ESC systems for the study of CGI chromatin accessibility, particularly with respect to DNA methylation and CpG density (Chapter 4) and ZF-CxxC domain-containing proteins (Chapter 5).

CHAPTER 4 - CHROMATIN ACCESSIBILITY AT CPG ISLANDS

4.1 Introduction

In the context of chromatinised DNA, accessibility to the genome is important for facilitating the recognition and binding of sequence-specific transcription factors (Voss and Hager, 2014) and may assist the transcriptional machinery to engage with gene promoters (Cockerill, 2011). Early single gene and biochemical studies had suggested that transcriptionally active gene promoters and CGIs had elevated accessibility as typically measured by nuclease sensitivity (Weintraub and Groudine, 1976, Stalder et al., 1980, Antequera et al., 1989, Tazi and Bird, 1990). More recently, genome-scale measurements of chromatin accessibility using DNase-seq (Sabo et al., 2004) or nucleosome positioning by MNase-seq (Li et al., 2011b, Fenouil et al., 2012, Teif et al., 2012) have validated and extended these observations. Furthermore, they have revealed that chromatin accessibility and nucleosome depletion are characteristic features of CGI promoters, whereas CpG-poor promoters are less accessible and more nucleosome-enriched. However, the mechanisms that drive elevated chromatin accessibility at these sites are poorly understood.

Several lines of evidence suggest that the chromatin structure and accessibility of CGI promoters may in part be influenced by DNA sequence features, specifically an enrichment of non-methylated CpG dinucleotides, and that this may be independent of chromatin-based transcriptional regulation. For example, inhibition of RNA polymerase II (RNAPII) in mammalian cells did not affect the nucleosome-depleted state of CGI promoters (Fenouil et al., 2012), suggesting that other features inherent to CGIs may dictate their chromatin structure. Intriguingly, several *in vitro* studies have suggested that CpG density or DNA methylation might influence the stability of the DNA-nucleosome

interactions, with CpG-rich sequences (Ramirez-Carrozzi et al., 2009) and non-methylated sequences (Lee and Lee, 2012, Collings et al., 2013) resulting in nucleosomes with lower stability. Furthermore, CpG methylation is coincident with reduced chromatin accessibility *in vivo* (Razin and Cedar, 1977, Ball et al., 1983, Antequera et al., 1989, Han et al., 2011, Thurman et al., 2012), and the recombination of non-methylated DNA sequences into mammalian cells results in more accessible chromatin than when the same sequences are pre-methylated (Keshet et al., 1986, Palacios et al., 2010). Intriguingly, when DNA containing a human CGI was inserted into the yeast genome it also exhibited elevated nuclease sensitivity (Mucha et al., 2000), suggesting that formation of accessible chromatin at CGIs could result directly from conserved features inherent to CGI DNA sequences. In spite of this long established relationship between CGIs and chromatin accessibility, it still remains unclear how CpG density and the lack of DNA methylation might contribute towards chromatin accessibility at these important regulatory elements. Here I set out to examine the relationship between CGIs and chromatin accessibility and begin to explore the molecular principles that underpin chromatin accessibility at CGI sequences in vertebrates.

4.2 Results

4.2.1 CpG density, DNA methylation and chromatin accessibility

Previous studies providing evidence that CGIs are associated with accessible chromatin typically examined a limited number of cell lines or tissues. I therefore wanted to explore the breadth of this relationship across different cell types and tissues to understand if accessible chromatin was indeed an inherent feature of CGI chromatin. Additionally, I wanted to characterise in detail the relationship between chromatin accessibility and DNA

methylation in mouse embryonic stem cells (ESCs), the model cell system that I intended to use to experimentally study CGI chromatin structure and function.

4.2.1.1 CGIs exhibit chromatin accessibility across many cell types

To understand if CGIs had elevated chromatin accessibility across different tissues, I used DNase-seq data from the ENCODE project for 35 different mouse cell types or tissues and examined the chromatin accessibility at CGIs. The results of this analysis are easily visualized by observing a gene-rich region of the genome for a subset of these tissues. What was evident was that CGIs were generally coincident with sites of DNase I hypersensitivity (DHS) across all tissues examined (Figure 4.2.1 A). In contrast, tissue-specific DHS tended to occur away from CGIs (Figure 4.2.1 A). To quantify this effect, DHS intervals were identified in all cell types and merged to create a union DHS set, representing all possible DNase I hypersensitive regions across these cell types. The number of cell types that non-CGI and CGI transcription start sites (TSS) were coincident with a DHS interval were then scored. This revealed that CGI promoters were strongly associated with a ubiquitous accessibility across cells types, whereas non-CGI TSS accessibility was much more restricted to specific tissues (Figure 4.2.1 B). Furthermore, when DHS were sorted by their breadth of accessibility, there was a very clear positive relationship between the number of cell types a DHS was present in and an elevated CpG density and GC content (Figure 4.2.1 C,D). These observed relationships have also been supported in publications emerging during the course of my studies (Natarajan et al., 2012, Wang et al., 2012, Sheffield et al., 2013). The fact that CGIs appear to be ubiquitously accessible suggests that CGIs might contain sequence-specific features that predispose them to exist in this DNase hypersensitive and accessible chromatin state.

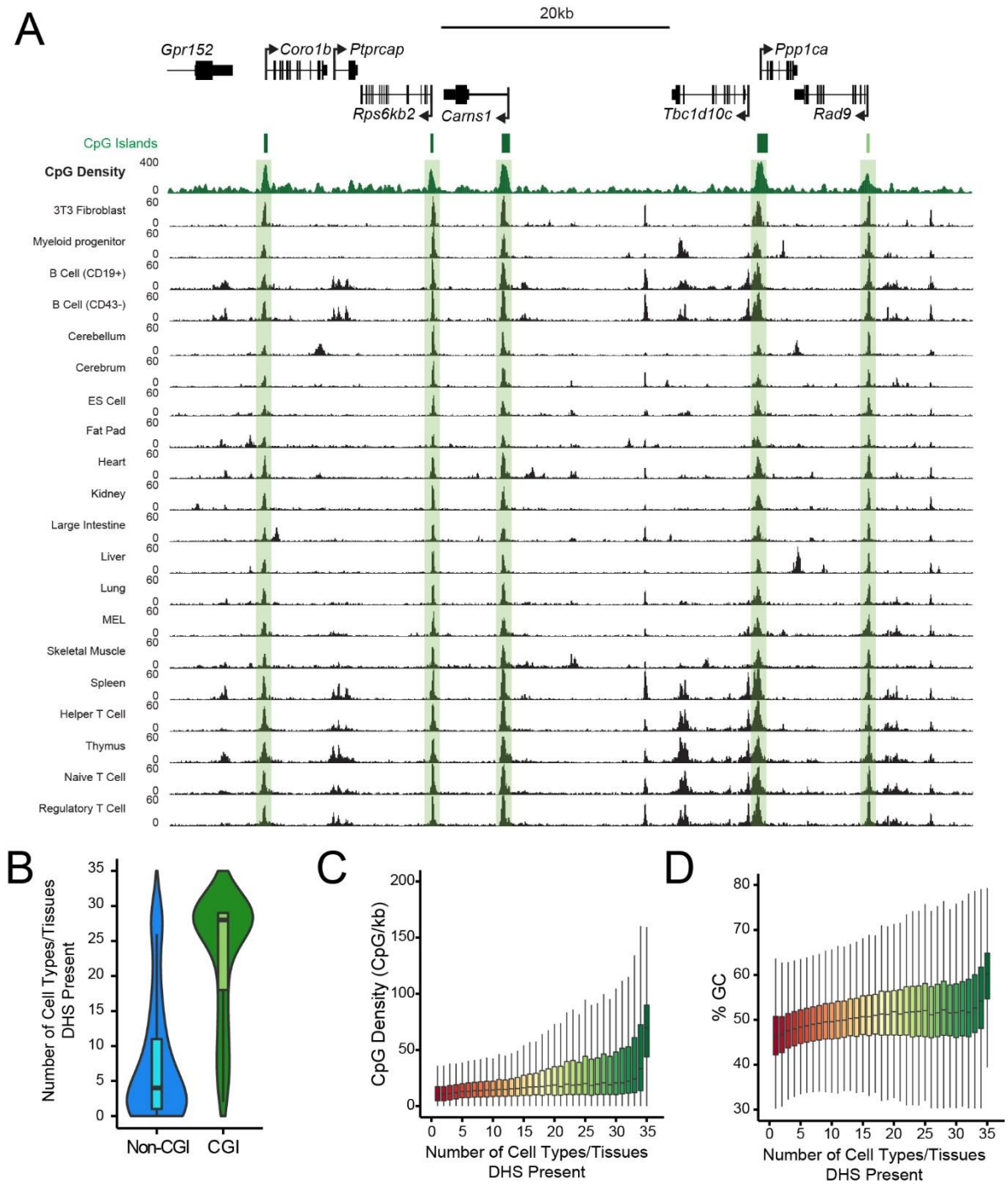


Figure 4.2.1 CpG-rich DNA exhibits chromatin accessibility across many cell types.

A) Genomic snapshot of DNase-seq data from 20 different cell types or tissues from the ENCODE project. CpG islands are highlighted in green boxes.

B) Non-redundant TSS (n=20634) which do (CGI) or do not (Non-CGI) overlap with a CGI, compared against the number of different cell types (out of 35 total) in which it is coincident with a DNase I hypersensitive site (DHS).

C) CpG density of DHS compared against the number of different cell types or tissues that DHS is present in.

D) Same as in (B), but for GC percentage.

4.2.1.2 Non-methylated DNA is associated with chromatin accessibility

CGIs displayed widespread chromatin accessibility across a wide-range of tissues (Figure 4.2.1). However this analysis relied on the use of computationally predicted CGIs. Therefore it does not take into consideration the methylation state of the DNA sequence, which can vary between tissues, albeit for a limited subset of CGIs. To explore in more detail the relationship between chromatin accessibility at CGIs it was important to integrate DNA methylation levels into this analysis. I chose to do this in mouse ESCs where extensive DNA methylation profiling has been carried out, and as the model system which I would further use to experimentally study CGI chromatin structure and function. To assess DNA methylation levels I used biotinylated CxxC affinity purification (BioCAP) sequencing data (Blackledge et al., 2012, Long et al., 2013b). BioCAP involves incubation of fragmented genomic DNA with a recombinant CxxC DNA-binding domain to specifically bind and enrich for non-methylated CpG dinucleotides (Blackledge et al., 2012). When combined with massively parallel sequencing, regions with BioCAP enrichment (non-methylated islands; NMIs) provide a means to identify non-methylated CpG-rich regions of DNA, and exclude methylated CGIs that might otherwise be considered if only computation CGI prediction was used (Blackledge et al., 2012, Long et al., 2013b). To this end, I focused subsequent analysis of CGIs on the experimentally determined NMI interval set from mouse ESCs.

To examine the relationship between non-methylated DNA and chromatin accessibility I therefore compared BioCAP enrichment with chromatin accessibility measurements generated by the assay for transposase accessible chromatin (ATAC)-seq (Chapter 3). Visualisation of BioCAP signal across regions of the genome revealed that loci with elevated CpG density (eg CGIs) and low levels of DNA methylation (ie high BioCAP signal) also had elevated chromatin accessibility (Figure 4.2.2 A). A quantitative comparison of DNA methylation and chromatin accessibility throughout the genome

revealed that these two features correlated well (r = 0.76) (Figure 4.2.2 B). Indeed, nearly all NMIs identified by BioCAP overlapped with accessible chromatin (94%) (Figure 4.2.2 C). Intriguingly, chromatin accessibility across NMIs appeared to be related to the CpG density (Figure 4.2.2 D), suggesting that CpG density might have the potential to influence or shape chromatin accessibility at CGIs. In addition to using BioCAP as a measure of DNA methylation, I also used whole genome-bisulfite sequencing datasets to examine the relationship between DNA methylation and chromatin accessibility in different mouse tissues. Bisulfite sequencing relies upon the conversion of non-modified cytosines to uracil, while methylated cytosines are unaffected, and therefore allows one to calculate percentage methylation at individual CpG sites. By comparing bisulfite sequencing methylation scores with DNase-seq signal at 1kb windows across the genome, I confirmed that there was a negative correlation between DNA methylation and chromatin accessibility across all tissues (Figure 4.2.2 E), in agreement with my observation in ESCs using BioCAP. Together, these observations are in fitting with previous work that linked non-methylated DNA to chromatin accessibility (Choi, 2010, Stadler et al., 2011, Thurman et al., 2012). Having explored and validated this relationship in ESCs and other cell types, I then wanted to begin to examine the potential mechanisms that might regulate the chromatin accessibility at non-methylated CGIs.

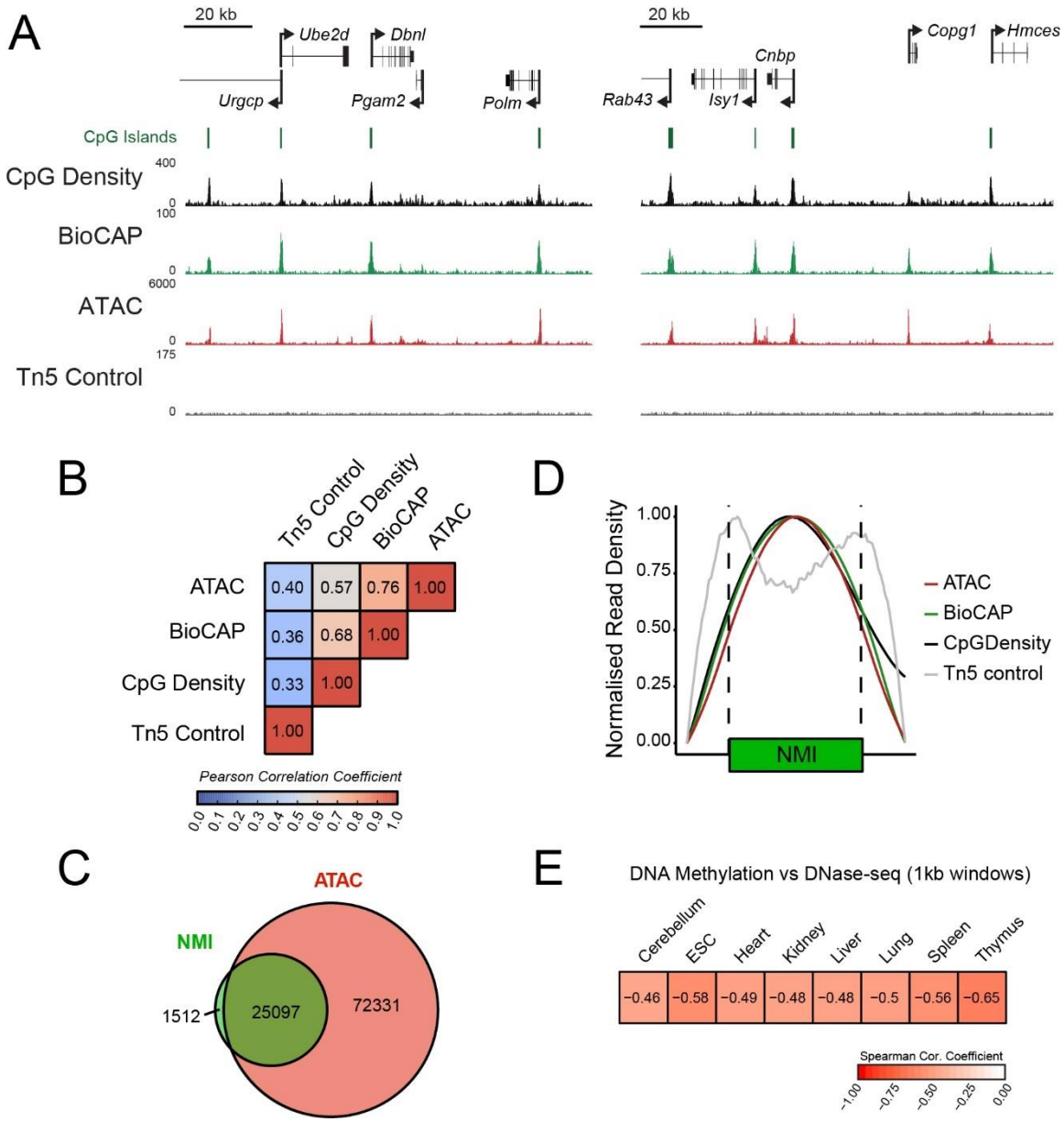


Figure 4.2.2 Examining chromatin accessibility at non-methylated DNA genome-wide.

A) Genomic snapshots for two gene-rich regions in mouse embryonic stem cells, depicting CpG density, BioCAP (ie non-methylated DNA), and ATAC-seq.

B) Genome-wide correlation between chromatin accessibility profiles (ATAC) and DNA methylation (BioCAP) in 1 kb windows. Control input track is included for reference.

C) Venn diagram of the overlap between ATAC hypersensitive regions with non-methylated islands (NMI) identified by BioCAP.

D) Metaplot depicting NMI intervals normalised by size to demonstrate the average chromatin accessibility profile across a NMI region.

E) Correlation of whole-genome bisulfite sequencing data from eight different mouse tissues compared with DNase-seq profiles in 1 kb windows across the genome.

4.2.2 Chromatin accessibility at NMI promoters is independent of transcription

There was a clear relationship between non-methylated, CpG-rich DNA and elevated chromatin accessibility throughout the ESC genome. However, CGI-associated gene promoters are more frequently associated with transcriptional activity than non-CGI promoters (Carninci et al., 2006, Core et al., 2008). A previous study which used MNase-based nucleosome positioning and occupancy measurements had suggested that CGIs appeared to maintain their nucleosome-depleted state in the absence of RNAPII activity (Fenouil et al., 2012), suggesting that the accessible nature of CGIs might be independent of their transcriptional activity. However, it was unclear how this measurement of nucleosome occupancy by MNase-seq compared to other nuclease sensitivity assays (eg ATAC-seq, DNase-seq), and if the open and accessible nature of NMIs measured by these assays might simply be the result of higher levels of transcription at these loci. I therefore wanted to examine this question more carefully. To achieve this, I set out to examine chromatin accessibility at NMIs in ESCs following perturbation of RNAPII function.

4.2.2.1 Promoter accessibility correlates with CpG density rather than transcriptional activity

To examine the determinants of chromatin accessibility at gene promoters, I initially identified transcriptional start sites (TSS) that were active in mouse ESCs based upon cap analysis of gene expression (CAGE) (FANTOM et al., 2014). I subsequently segregated these active TSSs based upon the presence or absence of an NMI. I chose to use NMIs rather than computationally predicted CGIs to ensure that only CpG-rich promoters that were non-methylated in ESCs were considered. When I compared ATAC-seq signal across all active TSS, it was evident that NMI promoters had significantly higher chromatin

accessibility than non-NMI promoters (Figure 4.2.3 A), similar to conclusions drawn from recent studies (Deng et al., 2013, FANTOM et al., 2014). To consider if this difference was related to transcriptional activity, I considered gene expression using 4-thiouridine (4SU) pulse-labelled RNA followed by highthroughput sequencing (4SU-seq) RNA-seq data (courtesy of Dr Vincenzo di Cerbo), which provides a nascent profile of expression within a defined time period (Friedel and Dolken, 2009, Rabani et al., 2011, Schwanhausser et al., 2011). Examination of 4SU-seq expression levels revealed that NMI promoters had significantly higher transcriptional activity than non-NMI promoters (Figure 4.2.3 B), as previously reported (Carninci et al., 2006, Core et al., 2008, FANTOM et al., 2014).

Given that NMI promoters seemed to be transcribed at substantially higher levels than non-NMI promoters, it was important to clarify whether the elevated chromatin accessibility observed at these sites was related to this difference in transcriptional activity. Examination of individual NMI or non-NMI genes of comparable expression suggested that the chromatin accessibility at these promoters correlated with the presence of an NMI, not transcriptional activity (Figure 4.2.3 C). To test this hypothesis, I separated active TSS based upon their enrichment for 4SU-seq in their gene bodies, and then further separated them based upon the presence or absence of a NMI at their TSS. A comparison of NMI and non-NMI TSS within each expression quartile revealed that NMI promoters were always significantly more accessible than non-NMI promoters, regardless of their transcriptional activity (Figure 4.2.3 D) (FANTOM et al., 2014). In fact, even the most highly expressed non-NMI genes were significantly less accessible that the least expressed NMI promoters ($p < 6\times10^{-11}$). Although these observations indicated that the main determinant of promoter accessibility appeared to be the presence or absence of an NMI, there was a slight correlation between gene expression and chromatin accessibility within the NMI and non-NMI groups (Pearson correlation coefficient of 0.42 and 0.37 respectively). This suggests that transcription may partially contribute to chromatin accessibility at gene promoters, although it could also be explained by other factors, such

as an enrichment of polycomb repressive complexes at inactive TSS (Section 5.2.2). Nevertheless, the difference in accessibility at NMI compared to non-NMI promoters, regardless of expression, strongly supported the conclusion that CpG density and non-methylated DNA are associated with the chromatin accessibility at gene promoters, and that this appears to be largely independent of transcription.

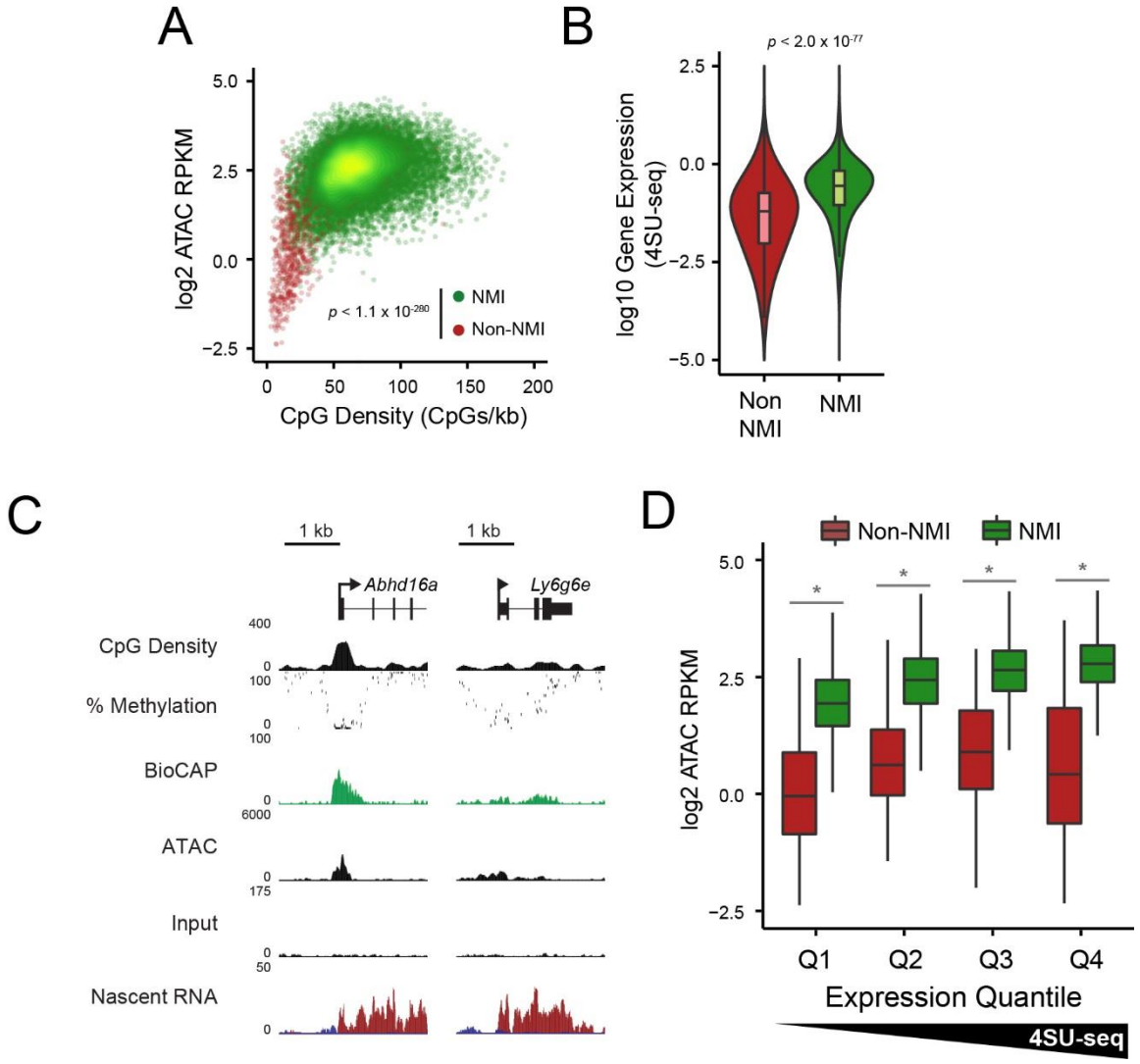


Figure 4.2.3 Relationship between gene expression and promoter accessibility.

A) TSS active in mouse embryonic stem cells (n = 13,464) ± 500 bp comparing CpG density and chromatin accessibility measured by ATAC-seq for NMI and non-NMI TSS.

B) Violin plot comparing transcriptional activity of active NMI or non-NMI TSS as measured by nascent 4SU-seq gene body reads per kilobase per million (RPKM).

C) Genomic screenshot of one NMI (*Abhd16a*) and one non-NMI promoter (*Ly6g6e*) of similar transcriptional activity.

D) Comparison of chromatin accessibility (ATAC RPKM) at active TSS ± 500 bp for NMI or non-NMI promoters ranked from low (Q1) to high (Q4) gene expression quartiles. * denotes p value < 1×10^{-20} as measured by Wilcoxon rank test.

4.2.2.2 NMI accessibility is maintained after RNAPII inhibition

Comparison of NMI and non-NMI TSS in mouse ESCs revealed that elevated chromatin accessibility associated with CpG-rich promoters appeared to be largely unrelated to transcriptional activity. To test this possibility, I treated mouse ESCs with two chemical inhibitors of RNAPII, flavopiridol and triptolide, in order to disrupt transcription. Flavopiridol inhibits P-TEFb kinase activity, subsequently preventing elongation of engaged or paused RNAPII (Chao and Price, 2001, Henriques et al., 2013, Jonkers et al., 2014). In contrast, triptolide interferes with RNAPII engagement with DNA by inhibiting the XPB component of the TFIIH helicase (Titov et al., 2011, Jonkers et al., 2014). The use of these inhibitors would therefore allow me to separately examine if there was any contribution of RNAPII elongation (flavopiridol) or engagement (triptolide) to promoter accessibility (Figure 4.2.4 A). As demonstrated previously using global run-on sequencing (GRO-seq) to measure RNAPII activity, the acute treatment of ESCs with flavopiridol and triptolide (50 minutes) is sufficient to block RNAPII elongation and engagement respectively (Jonkers et al., 2014). This short treatment should theoretically avoid many of the issues that might result from the chronic disruption to RNAPII function. Using a similar protocol to Jonkers *et al.*, I was able to successfully inhibit RNAPII function, with pre-mRNA levels reduced to 5-20% of untreated cells (Figure 4.2.4 B). This therefore allowed me to proceed and examine the contribution of RNAPII to chromatin accessibility.

To test whether promoter chromatin accessibility was affected by the inhibition of transcriptional activity, I performed ATAC-seq in untreated ESCs and cells treated with flavopiridol or triptolide. Initial comparison of the ATAC-seq signal at several promoters in flavopiridol- and triptolide-treated cells revealed that chromatin accessibility appeared to be largely unchanged after inhibition of either RNAPII elongation or engagement (Figure 4.2.4 C). Indeed, there were no apparent reductions in total chromatin accessibility at TSS genome-wide with either flavopiridol or triptolide (Figure 4.2.4 D). Intriguingly, there

appeared to be a slight increase in ATAC-seq signal for flavopiridol-treated cells at some TSS (Figure 4.2.4 C,D), consistent with elevated RNAPII occupancy as measured by GRO-seq. These increases in chromatin accessibility were significant for 9 % of TSS (Figure 4.2.4 E). These increases indicated that elevated RNAPII occupancy might contribute to chromatin accessibility. However, inhibition of RNAPII engagement by triptolide resulted in very few significant changes in promoter chromatin accessibility. Flavopiridol, in addition to CDK9, is also known to efficiently inhibit other proteins within the nucleus, such as CDK1, CDK4, CDK6, CDK8, and CDK12 (Conroy et al., 2009, Schmerwitz et al., 2011, Bösken et al., 2014), suggesting that the observed effect could also potentially result from the dysregulation of RNAPII-independent factors. Ultimately, the fact that triptolide had little to no effect on promoter chromatin accessibility, and that flavopiridol did not result in reduced chromatin accessibility, is good evidence that RNAPII binding and transcriptional elongation is not essential to maintain chromatin accessibility at NMI-associated gene promoters. The experiments presented here are supported by previous work using an alternative RNAPII chemical inhibitor, α-amanitin, which found only a minor contribution of RNAPII occupancy towards the nucleosome-depleted state of CGI promoters (Fenouil et al., 2012). These observations therefore indicate that other sequence- or chromatin-associated factors are likely to be responsible for NMI chromatin accessibility independent of transcription.

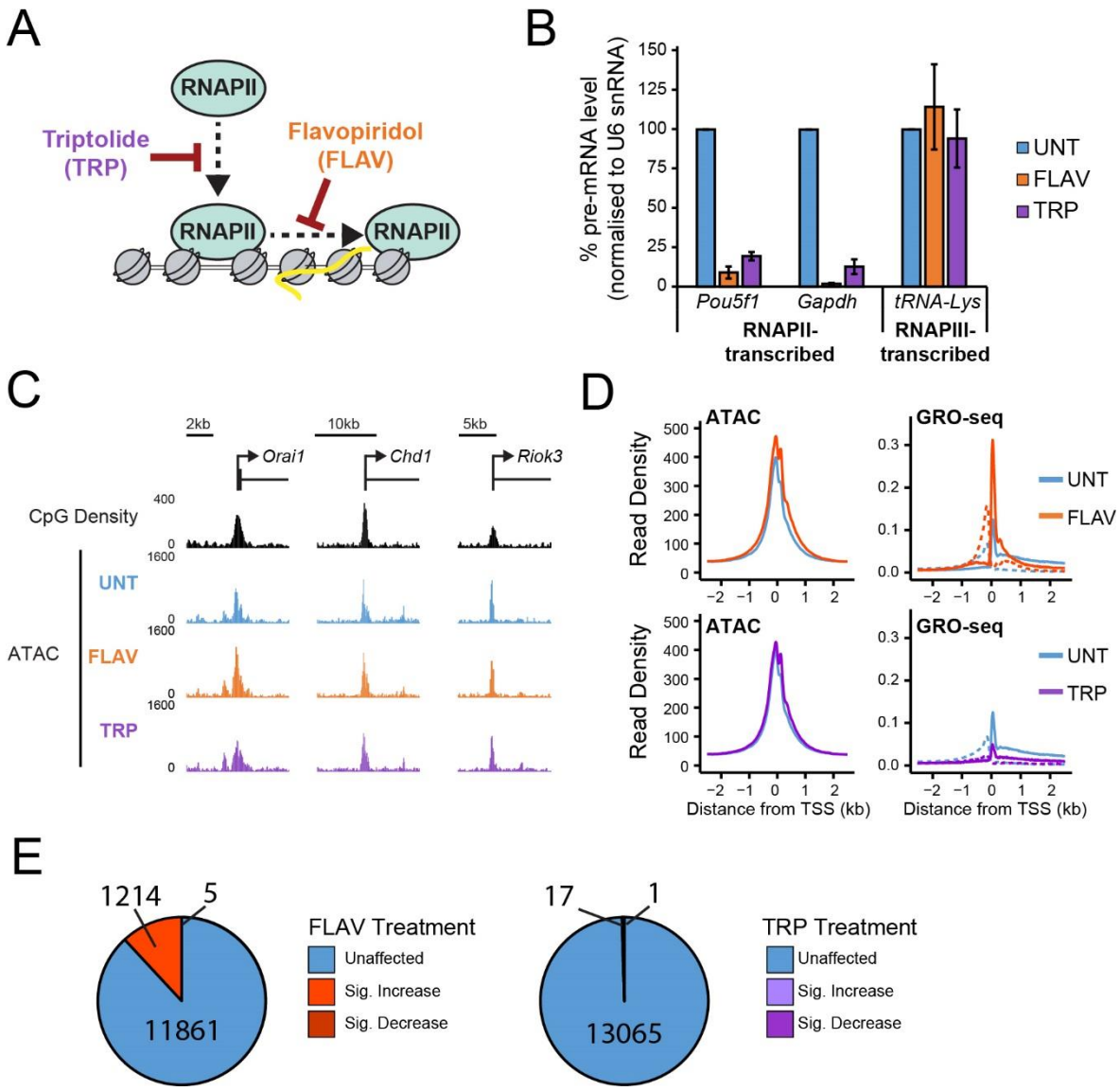


Figure 4.2.4 RNAPII has a minor influence on chromatin accessibility at promoters.

A) Schematic depicting inhibition of RNA polymerase II (RNAPII) activity, either by flavopiridol (FLAV) through inhibition of elongation or triptolide (TRP) which inhibits RNAPII engagement.

B) Quantitative real-time reverse transcriptase PCR validation of FLAV and TRP treatment using nuclear RNA and pre-mRNA-specific primers. Data is normalised to the RNA polymerase III-transcribed tRNA-Lys, and presented relative to untreated control. *n*=4 and error bars denote ±standard deviation.

C) Genomic snapshot for ATAC-seq at CGI promoters after treatment of ESCs with FLAV or TRP.

D) Metaplot analysis at active TSS for ATAC-seq and GRO-seq (Jonkers et al., 2014) after treatment of mouse ESCs with FLAV or TRP.

E) Piechart depicting significant changes in ATAC-seq (padj < 0.05; > 1.5 fold change) at active TSS after treatment of mouse ESCs with FLAV or TRP.

4.2.3 Loss of DNA methylation results in chromatin accessibility changes

In most cell types, the majority of CGIs are non-methylated. In contrast, some alternative genomic regions with an elevated CpG density are hypermethylated, particularly repetitive DNA elements and imprinted regions (Bird, 2002, Meissner et al., 2008), although a small subset of CGIs can be methylated as well (Yamada et al., 2004, Weber et al., 2005, Shen et al., 2007). Given the relationship that I observed between non-methylated, CpG-rich DNA and chromatin accessibility, I hypothesized that if the non-methylated state alone drives accessibility then the removal of DNA methylation from methylated CpG-rich sequences would lead to increases in chromatin accessibility. To test this hypothesis, I therefore profiled chromatin accessibility using ATAC-seq in a constitutive knockout ESC line lacking the three central DNA methyltransferases DNMT1, DNMT3A, and DNMT3B (DNMT TKO) that subsequently lack nearly all DNA methylation (Tsumura et al., 2006). After having performed this experiment in biological quadruplicate, I initially proceeded to try and understand whether there were any significant changes in chromatin accessibility throughout the ESC genome, and whether any such changes might be related to underlying CpG density and/or DNA methylation at these loci.

4.2.3.1 DNA methylation limits chromatin accessibility of many distal regulatory elements

To understand if DNA methylation may act to regulate chromatin accessibility, I began by visually examining sequences which were partially or completely methylated in wild type mouse ESCs, as measured by BioCAP and whole-genome bisulfite sequencing (Figure 4.2.5 A). For example, the imprinted *H19* locus has one allele methylated in ESCs and shows low levels of ATAC-seq signal. As predicted, and observed previously (Ito et al., 2013), loss of DNA methylation resulted in increased accessibility at this locus. Many changes were also observed at sites distal to TSS, with both gains and losses of ATAC-

seq signal (Figure 4.2.5 B). However, quantitation of accessibility changes at highly methylated regions (HMR; >50% methylation), lowly methylated regions (LMR; 20-50% methylation) and non-methylated regions (NMR; <20%) throughout the genome failed to identify any appreciable trends between normal DNA methylation and changes in chromatin accessibility (Figure 4.2.5 C). To explore the relationship between DNA methylation and chromatin accessibility in a more statistically robust manner, ATAC-seq peaks were identified in both wild type and DNMT TKO cells, and regions that were reproducibly and significantly different in their chromatin accessibility identified using DiffBind. This revealed a large number of sites which significantly increased in chromatin accessibility (n=10,753), compared to those that decreased (n=4,930) (Figure 4.2.5 D). This appeared to be consistent with DNA methylation tending to limit or constrain chromatin accessibility (Yang et al., 2012, Taberlay et al., 2014). Methylation-sensitive regions (ie. sites with increased accessibility in DNMT TKO) tended to be distal to gene promoters (Figure 4.2.5 E), suggesting that they might be distal regulatory elements like enhancers. Intriguingly, these changes appeared to be at sites that already displayed some previous accessibility (Figure 4.2.5 F). This is in fitting with DNA methylation limiting, but not preventing completely, the establishment of chromatin accessibility, as previously suggested (Palacios et al., 2010). However, whether or not any of the changes observed were related to the loss of DNA methylation and/or their underlying CpG density was unclear. To test this, I compared the wild type DNA methylation levels and CpG densities at methylation-sensitive peaks with sites that were unaffected or decreased in their chromatin accessibility. Reassuringly, the sites that gained chromatin accessibility in the DNMT TKO cells had high levels of DNA methylation in wild type ESCs (median 76%) (Figure 4.2.5 G) as also reported in a recent DNase-seq experiment (Domcke et al., 2015). However, these sites did not exhibit high CpG densities when compared with unchanged or decreased ATAC peaks (Figure 4.2.5 G). This suggested that although methylation may constrain chromatin accessibility at some sites, the underlying CpG

density was not a central feature associated with increased accessibility in the absence of DNA methylation.

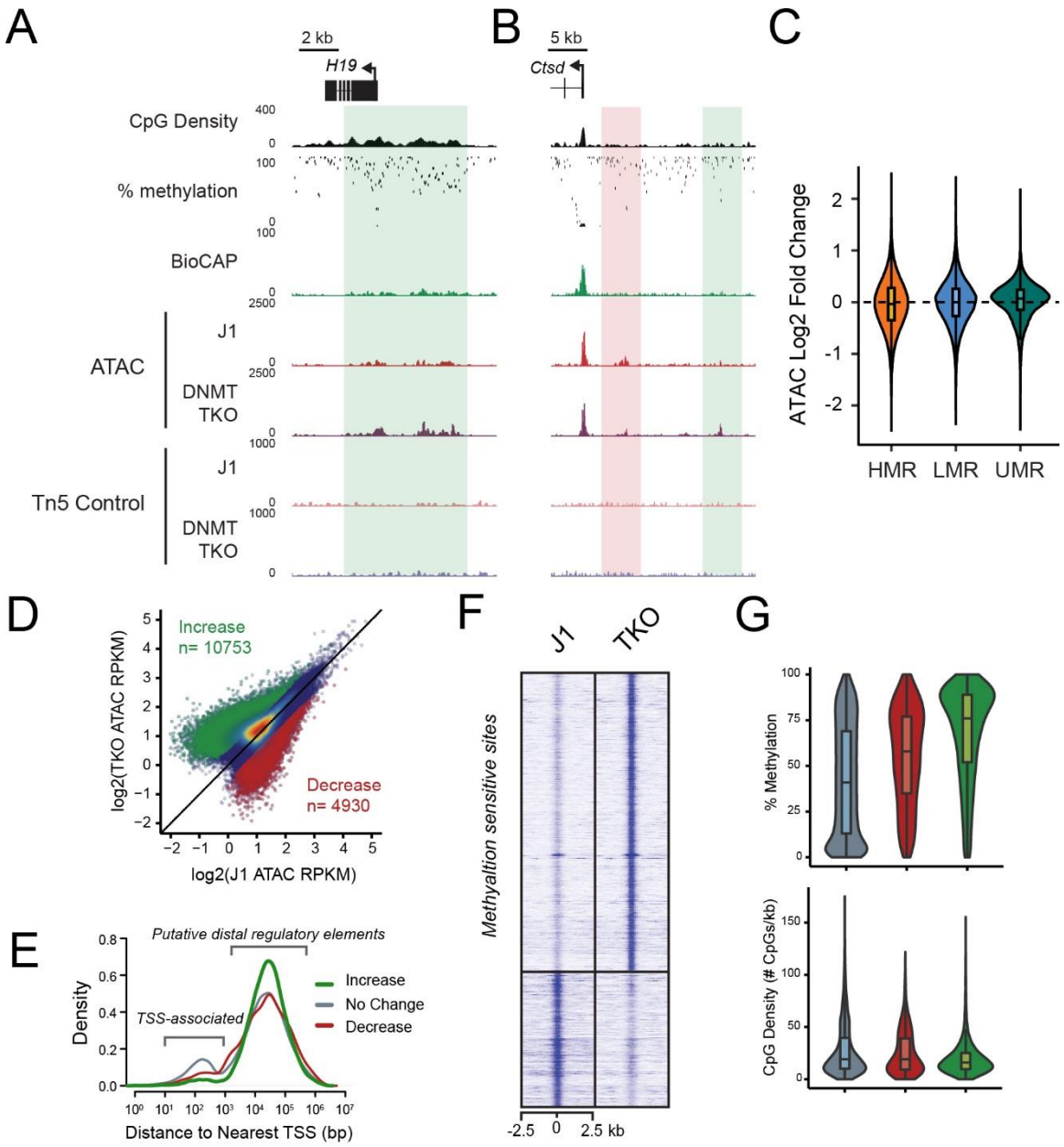


Figure 4.2.5 Loss of DNA methylation results in changes in chromatin accessibility at regulatory elements.

A) Genomic snapshots of ATAC-seq in the DNMT triple-knockout (TKO) and wild type J1 mouse ESCs at the imprinted H19 locus.

B) Same as A), illustrating examples of distal increases (green boxes) or decreases (red boxes).

C) Violin plot quantifying log2 fold change in DNMT TKO ATAC-seq compared to wild type J1 at mappable regions of the genome (1kb) segregated by DNA methylation levels in wild type ESCs, as determined by bisulfite sequencing. Highly methylated regions (HMR; >50% methylation; *n*=1,579,617), lowly methylated regions (LMR; 20-50% methylation; *n*=36,123) and non-methylated regions (NMR; <20%; *n*=28,293).

…continued on following page…

D) Scatterplot of wild type J1 and DNMT TKO ATAC-seq reads per kilobase per million (RPKM) at ATAC hypersensitive sites. Significant differences are highlighted in green.
E) Frequency distributions of distance to nearest TSS for ATAC hypersensitive sites which increase, decrease or remain unchanged in the DNMT TKO ESCs.
F) Heatmap analysis of J1- and TKO-specific ATAC hypersensitive sites.
G) Comparison of DNA methylation (upper panel) and CpG density (lower panel) for J1- and TKO-specific ATAC hypersensitive sites.

4.2.3.2 Methylated CGIs do not gain accessibility after loss of DNA methylation

An unbiased identification of ATAC hypersensitive regions in the DNMT TKO cells revealed many gains in chromatin accessibility related to the loss of DNA methylation, but this did not appear to be related to CpG density. To more directly address whether methylated CpG-rich sequences were subject to any changes in chromatin accessibility in the absence of DNA methylation, I separated computationally predicted CGIs into three groups, corresponding to their methylation state in wild type ESCs (non-methylated (0-20%), lowly methylated (20-50%) and highly methylated (50-100%)) using whole genome bisulfite sequencing data. Although some individual examples of CpG-rich sequences gaining small amounts of accessibility were observed (Figure 4.2.6 A), comparison of J1 and DNMT TKO ATAC-seq profiles at highly-methylated CGIs revealed that there was no apparent change in their average chromatin accessibility in cells lacking DNA methylation (Figure 4.2.6 B) and statistical analysis failed to identify any significant changes within these groups (data not shown). Surprisingly, non-methylated CGIs demonstrated a slight increase in chromatin accessibility, perhaps consistent with reports in human cancer cell lines lacking DNA methylation where the breadth of chromatin accessibility at non-methylated CGIs appeared to be slightly increased (Lay et al., 2015). Regardless, it was clear that CGIs with intermediate or high levels of methylation in wild type cells failed to gain chromatin accessibility in the absence of DNA methylation.

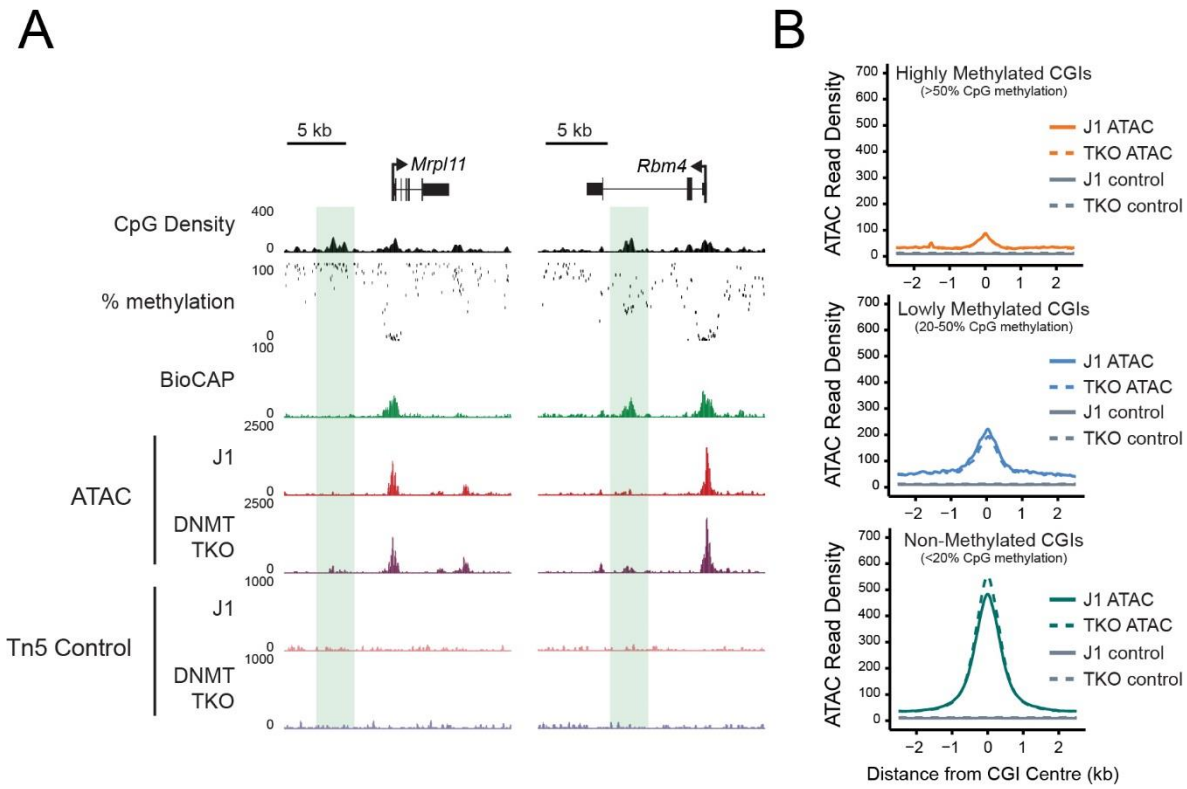


Figure 4.2.6 Methylated CGIs do not gain chromatin accessibility upon DNA methylation loss.

A) Genomic snapshot for two methylated CpG-rich sequences which exhibit small increases in chromatin accessibility in the DNMT TKO ESCs.

B) Metaplot analysis of DNMT TKO ATAC-seq at highly (50-100%; n = 1013), lowly (20-50%; n = 800) or non-methylated (0-20%; n = 13902) CGIs.

4.2.3.3 CpG-rich repeats gain chromatin accessibility upon DNA methylation loss

In addition to CGIs, the other major class of CpG-rich sequences in mammalian genomes are repetitive elements, which although normally silenced, in some cases become active in the DNMT TKO ESCs. Given their repetitive nature, these elements would not have been considered in my previous analysis. I therefore directly examined whether the chromatin accessibility of different repetitive classes might be regulated by DNA methylation, and if this was possibly related to their CpG composition. Average ATAC-seq enrichment was quantified at different classes of repetitive elements in the wild type J1

and DNMT TKO ESCs, and their change in chromatin accessibility was calculated. When repetitive elements were ranked by their CpG density, it was clear that there was a tendency for CpG-rich repetitive elements, particularly long terminal repeats (LTRs), to gain accessibility in the absence of DNA methylation (Figure 4.2.7 A), with a positive correlation between change in ATAC-seq signal and CpG density (Pearson correlation coefficient = 0.41). The LTRs which tended to be most dramatically affected were intra-cisternal A-type particle (IAP) elements (Figure 4.2.7 B). In general, the repetitive elements which gained accessibility upon loss of DNA methylation had younger evolutionary ages than repetitive elements which remained inaccessible (Figure 4.2.7 C). This is fitting with their elevated CpG density, as they have had a more limited time for evolutionary mechanisms to drive down CpG dinucleotide frequency compared to the remainder of the genome, while older repetitive element sequences demonstrate CpG-poor densities similar to the bulk genome (Coulondre et al., 1978, Duncan and Miller, 1980, Shen et al., 1994). Furthermore, the lack of re-activation for older repetitive elements is consistent with the presence of DNA methylation-independent repressive mechanisms such as KRAB zinc-finger proteins (Thomas and Schneider, 2011, Rowe et al., 2013).

To consider the relationship with CpG density with the changes in chromatin accessibility, I profiled the ATAC-seq read densities over two of the most affected IAP repetitive elements, IAP-LTR1 and IAP-LTR2a (Figure 4.2.7 D). This revealed that the bulk of increases in chromatin accessibility at these repetitive elements was at their TSS, and was not a direct reflection of the elevated CpG density across the entire repetitive element. This suggested that CpG density might not be the driving feature causing increases in accessibility at these sites and instead that other features located at the TSS may be responsible. However, transcriptional re-activation, as examined by RNA-seq, was only slightly correlated with increased chromatin accessibility of repetitive elements (*cor* = 0.15) (Supplementary Figure 1), and did not seem to be sufficient to explain the enhanced

accessibility at repetitive elements. Ultimately, these observations support a role for DNA methylation in limiting the chromatin accessibility of CpG-rich repetitive elements, although it is unclear whether this is due to the absence of DNA methylation, or the presence of other signals. Certainly further work is needed to examine the contribution of CpG density towards chromatin accessibility in the absence of DNA methylation, and to understand how hypomethylated DNA may help to establish chromatin accessibility.

4.2.4 Is non-methylated, CpG-rich DNA sufficient to create chromatin accessibility?

Correlative analyses in wild type cells (Figure 4.2.1, Figure 4.2.2 & Figure 4.2.3), and analysis of cells lacking DNA methylation (Figure 4.2.5, Figure 4.2.6, & Figure 4.2.7), had suggested that chromatin accessibility was at least partially related to DNA methylation levels and CpG density. This was supported by previous studies describing a role for the CpG content or methylation status of CpG-rich sequences in influencing nucleosome occupancy and chromatin compaction *in vitro* (Ramirez-Carrozzi et al., 2009, Lee and Lee, 2012, Collings et al., 2013), or upon recombination into cells (Mucha et al., 2000, Palacios et al., 2010). However, CGI sequences have co-evolved with and frequently coincide with promoters throughout the genome (Saxonov et al., 2006, Deaton and Bird, 2011). It can therefore be difficult to untangle CGI-dependent processes from features inherent to promoters and transcriptional activity. Although I have addressed one such confounding influence, RNAPII and transcription (Figure 4.2.3 & Figure 4.2.4), the complexity of endogenous promoter elements will always make it challenging to understand the precise contribution of CpG density towards chromatin structure and accessibility. To circumvent some of these issues, an alternative approach was employed to target exogenous DNA sequences into an inert region of a mammalian genome. By varying the CpG density and DNA methylation state of such sequences it would allow me to examine the contribution of CpG-rich sequences towards chromatin accessibility, while

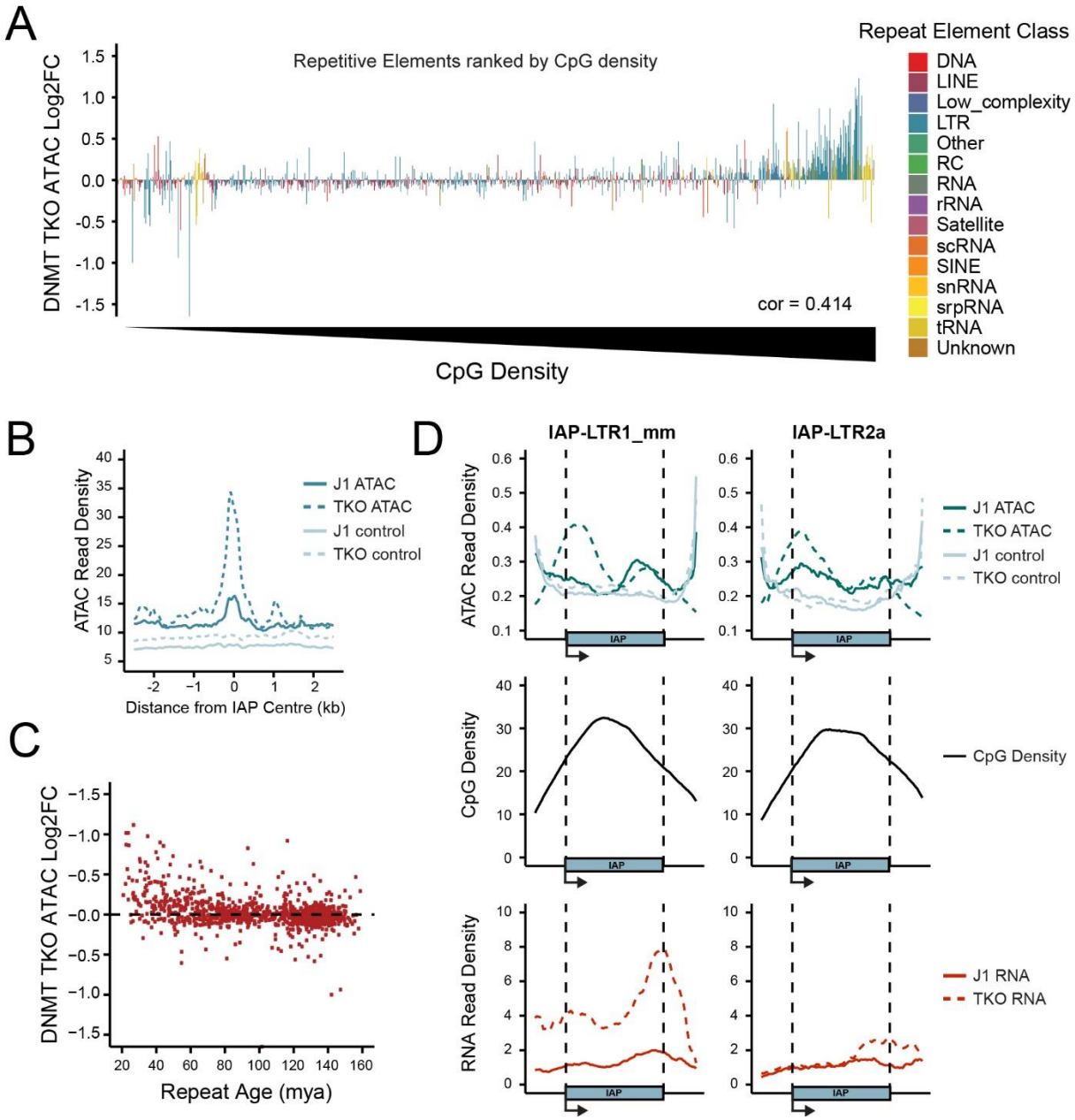


Figure 4.2.7 Chromatin accessibility at CpG-rich repetitive elements is limited by DNA methylation.

A) Quantification of the average log2 fold change of ATAC-seq reads (DNMT TKO/J1) at repetitive elements grouped by the RepName feature in RepMasker (n = 1200), ranked by CpG density and coloured by RepClass feature.

B) Metaplot analysis of ATAC-seq at IAP elements in DNMT TKO ESCs.

C) Scatterplot comparing the log2 fold change of DNMT TKO ATAC-seq at repetitive element types (RepType) with their repeat age (substitution rate).

D) Metaplot analysis of two different IAP elements, IAP-LTR1 and IAP-LTR2a, profiling ATAC-seq, CpG density, and RNA-seq. Interval represents normalised IAP element, ± 20 %, with strand indicated by arrow.

theoretically avoiding many confounding influences that exist at eukaryotic CGI sequences. To address this, I chose to employ a strategy of targeting bacterial DNA sequences with elevated CpG densities into a single, defined locus in mouse ESCs. In order to interrogate a wide range of different DNA sequences, I sought to develop a system that would allow me to achieve this in a semi-high throughput manner. The strategy I selected was a recombinase-mediated cassette-exchange (RMCE) approach.

4.2.4.1 Development and optimisation of RMCE targeting system

I set out to establish a RMCE targeting system in mouse ESCs, as depicted in Figure 4.2.8 A. In an attempt to avoid confounding influences at transcriptionally active genomic regions, I exploited a mouse ESC line containing a single copy of a human gene desert sequence (170kb) integrated into chromosome 8 (Blackledge et al., 2014), into which I introduced a copy of the hygromycin phosphotransferase-thymidine kinase (HygroTK) fusion gene flanked on either side by a loxP or variant loxP (loxP') recombination site (Figure 4.2.8 A). Transient over-expression of Cre recombinase, in the presence of a donor DNA sequence with compatible loxP/loxP' sites, would allow Cre-mediated recombination of the donor sequence with the HygroTK locus. In this system, HygroTK acts as a useful counter-selectable marker, whereby the presence of HygroTK conveys resistance of cells to hygromycin treatment, but sensitivity to ganciclovir. This therefore allows for both efficient targeting of the loxP-HygroTK-loxP' sequence into the gene desert, and selection for successful Cre-mediated recombinant clones with ganciclovir treatment (loss of HygroTK provides ganciclovir resistance). Importantly, loxP and loxP' sites cannot recombine with each other, and are only able to recombine with their identical loxP sequence (ie loxP to loxP, or loxP' to loxP'). This limits the potential for non-specific recombination and excision of the HygroTK gene without an exogenous donor DNA sequence. After successfully targeting the loxP-HygroTK-loxP' cassette into the human gene desert-containing mouse ESCs (data not shown), hereafter referred to as RMCE-

HygroTK ESCs, I confirmed they were ganciclovir sensitive (Figure 4.2.8 B), indicating the presence of the HygroTK cassette. To test the efficiency of recombination at this locus, cells were transfected with a Cre expression plasmid which also contained a donor sequence flanked by loxP/loxP´sites, and treated with ganciclovir for 10 days. Clones were then screened for recombination events by PCR, and 7/16 clones were confirmed to have undergone insertion of the donor sequence into the HygroTK locus (Figure 4.2.8 C). This indicated that my RMCE system allowed for targeted integration of specific DNA sequences into the mouse genome, and I therefore proceeded to design and target bacterial sequences using this approach in order to interrogate the influence of CpG density on establishing chromatin accessibility.

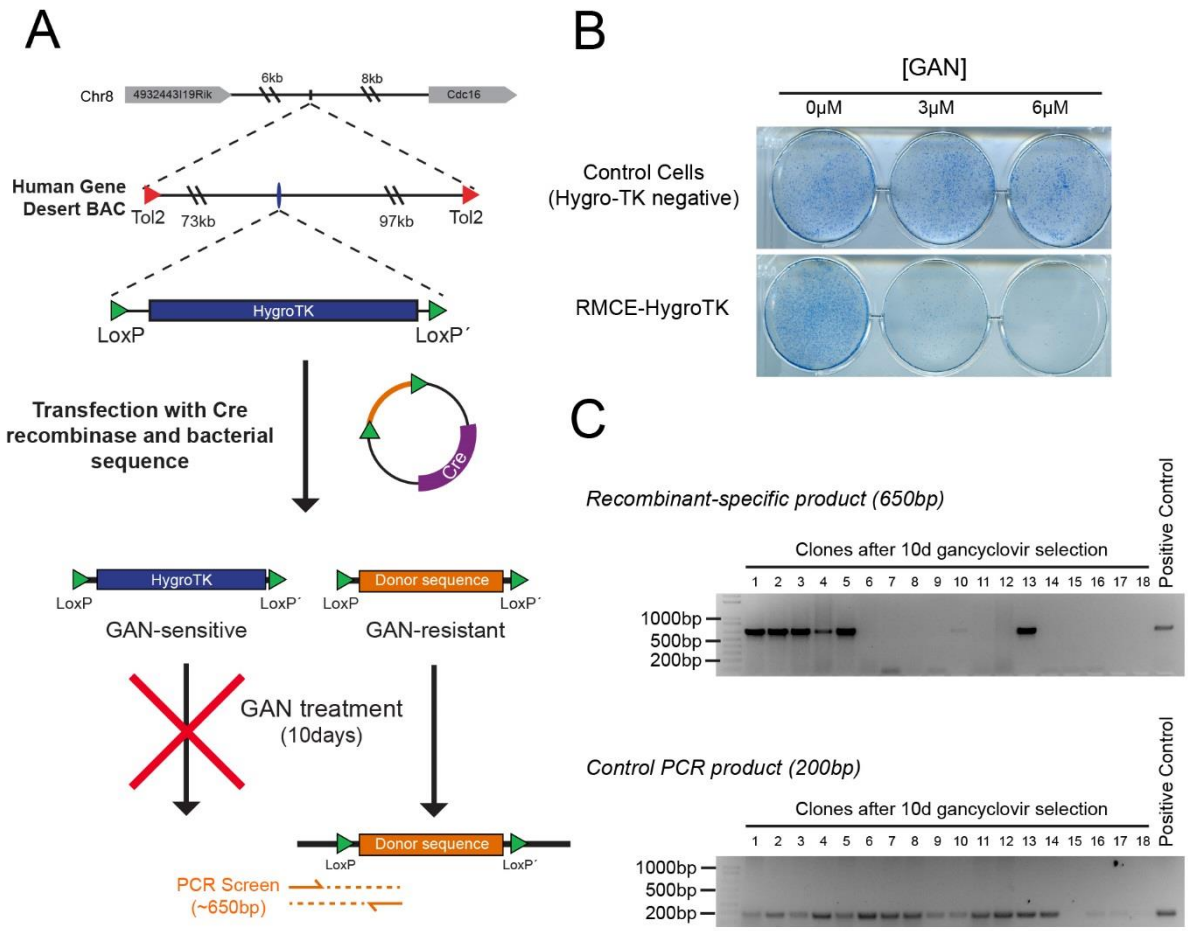


Figure 4.2.8 Establishing recombinase-mediated cassette exchange system in ESCs.

A) Schematic detailing RMCE system for targeting bacterial DNA sequences into a human gene desert in mouse ESCs. After transfection with an expression plasmid for Cre recombinase and

… continued on following page…

a donor DNA sequence, the HygroTK gene is excised and replaced by the donor sequence. Treatment with ganciclovir (GAN) then selects for clones which have lost the HygroTK gene.

B) Methylene blue staining of RMCE-HygroTK ESCs after treatment with GAN.

C) Genomic DNA PCR screen for recombinant clones after transfection with Cre-recombinase and donor sequence and 10 d GAN treatment. Upper panel shows recombinant-specific product, while lower panel depicts a control PCR to confirm the presence of DNA.

4.2.4.2 Targeting bacterial CpG-rich sequences into the mouse genome

Insertion of bacterial or other exogenous sequences into the genomes of living cells has revealed interesting principles underlying why CpG-rich sequences are refractory to DNA methylation (Krebs et al., 2014, Wachter et al., 2014, Long et al., 2016). However, the contribution of different CpG densities, GC content and DNA methylation towards chromatin accessibility has only been examined in a very preliminary manner (Mucha et al., 2000, Palacios et al., 2010, Hashimoto et al., 2016). The use of bacterial sequences as source for CpG-rich sequences is particularly useful in order to limit the potential for confounding effects of vertebrate sequence-specific factors. After a comparison of the genomic CpG density and GC content in several bacterial species, *Salmonella enterica* was selected, as it provided a similar range of CpG and GC content to mammalian NMIs (Supplementary Figure 2A). Bioinformatic fragmentation of the *S. enterica* genome allowed the selection of ten different 1kb DNA sequences to be cloned for subsequent targeting into the mouse genome, representing varying CpG and GC contents (Supplementary Figure 2B). These sequences were then cloned into a plasmid so that they were flanked by loxP and loxP' recombination sites, prior to transfection into the RMCE-HygroTK ESCs. Initially, a small scale experiment was performed with five of the ten sequences, two with low (LCG1-2) and three with high (HCG1-3) CpG densities. After transfection and ganciclovir selection, very few clones survived, and I was only able to recover between two and four clones for each sequence. Fortunately, I was able to confirm that 3/4 and 2/3 clones were positive for successful targeting and recombination of the HCG1 and HCG2 sequences, respectively (Figure 4.2.9 A,B). However, I was unable to

verify recombinant clones for LCG1, LCG2 or HCG3 (data not shown), even with multiple attempts at optimisation of the PCR screen. Regardless, I proceeded to test the DNA methylation and chromatin accessibility of the recombinant HCG1 and HCG2 sequences, which had CpG densities comparable with mammalian CGIs.

Insertion of non-methylated CpG-rich sequences into mammalian genomes typically results in low levels of DNA methylation (Jermann et al., 2014, Krebs et al., 2014, Wachter et al., 2014), and I was able to confirm that HCG1 and HCG2 had low or intermediate DNA methylation levels, as measured by methylation-sensitive restriction enzyme digest of genomic DNA (similar to Hashimoto et al., 2007) (Figure 4.2.9 C). To test the chromatin accessibility at these newly inserted exogenous sequences I adapted the ATAC protocol for qPCR quantitation of Tn5 cleavage frequency, similar to the DNase-qPCR protocol described earlier (McArthur et al., 2001). I then measured the Tn5 hypersensitivity of the recombinant HCG1 and HCG2 loci (Figure 4.2.9 D). Surprisingly, this revealed that the HCG1 and HCG2 CpG-rich sequences exhibited chromatin accessibility of comparable levels to an endogenous CGI hypersensitive site. This strongly suggested that non-methylated CpG-rich DNA might indeed be sufficient to establish accessible chromatin. However, several sequence motifs for transcription factors expressed in mouse ESCs were identified in the HCG1 and HCG2 sequences (Supplementary Figure 3). It is not possible to know whether these sites are bound *in vivo*, or whether they contribute to elevated chromatin accessibility, without experimental validation by ChIP and subsequent mutation of the binding sites. As such, it was difficult to conclude whether the underlying determinant in creating accessibility was elevated CpG density, the lack of DNA methylation or whether other factors might be implicated.

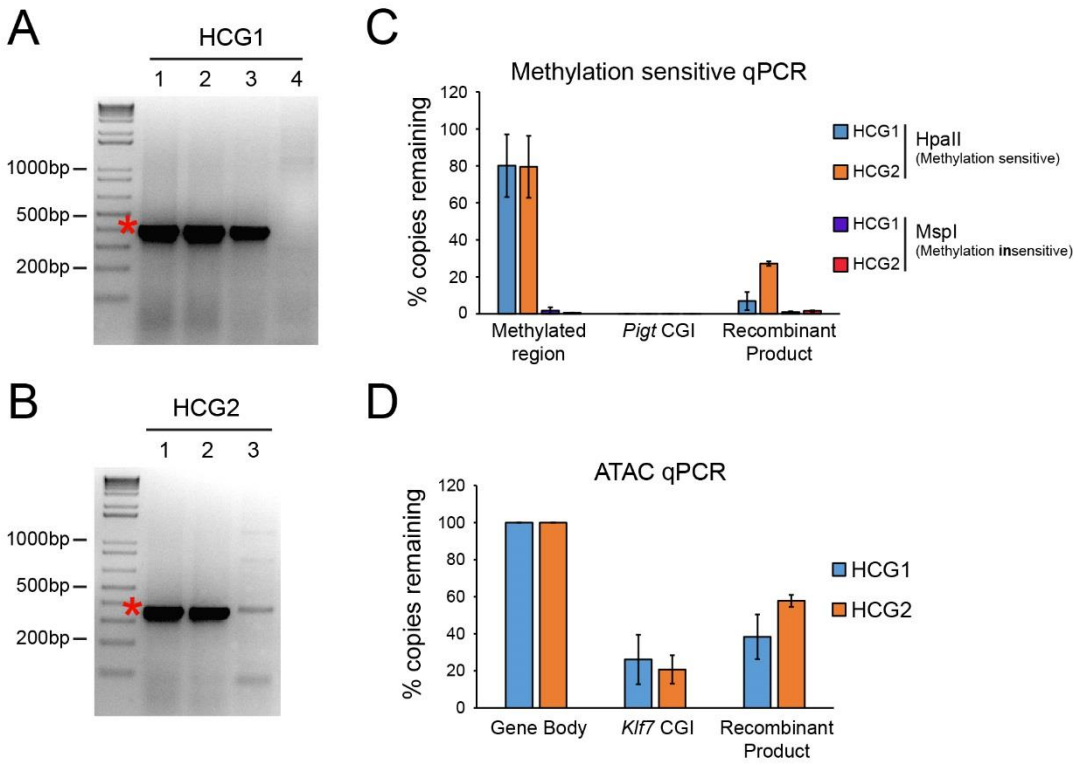


Figure 4.2.9 Integration of non-methylated, CpG-rich bacterial DNA into mouse ESCs.

A) PCR screen for the *S. enterica* DNA sequence HCG1 inserted into the RMCE-HygroTK locus. * denotes correct PCR product.

B) Same as in A), but for *S. enterica* DNA sequence HCG2.

C) Genomic DNA from multiple clones (HCG1 *n*= 3; HCG2 *n*=2) was digested with HpaII (methylation-sensitive) or MspI (methylation-insensitive) restriction enzymes, and compared with the amplification efficiency of undigested genomic DNA for a control, methylated region, an endogenous non-methylated CGI promoter (*Pigt*). Error bars represent standard deviation across biological clones.

D) ATAC-qPCR measuring nuclease digestion efficiency (ie nuclease hypersensitivity) at recombinant HCG1 and HCG2 sequences. The hypersensitive *Klf7* CGI is included for comparison. Error bars represent standard deviation across biological clones.

The insertion of two bacterial CpG-rich sequences was able to establish a new region of chromatin accessibility in a gene desert in mouse ESCs. To test the breadth of this observation, and to examine the influence of DNA methylation, I set out to generate recombinant lines for more *S. enterica* sequences, both with and without DNA methylation. Examining the effect of DNA methylation in this system would likely be critical to understand how CpG density could exert any effect on chromatin accessibility, and

whether this was potentially through recognition of non-methylated CpG dinucleotides by zinc-finger CxxC-domain proteins. I therefore methylated the *S. enterica* DNA sequences *in vitro*, prior to transfection into the RMCE-HygroTK mouse ESCs. However, even with non-methylated DNA fragments, I was unable to successfully identify any recombinant clones for the eight additional DNA sequences, across multiple independent attempts (data not shown). Indeed, after the initial optimisation experiments and successful targeting of HCG1 and HCG2, I was unable to generate any further recombinant lines. The problem seemed to be related to the ganciclovir selection, as the background resistance of the RMCE-HygroTK seemed to increase with cell passage number. I attempted multiple experiments to subsequently optimise the ganciclovir concentration for selection by varying both ganciclovir concentration and cell number, but was unable to solve the problem. The RMCE-HygroTK ESCs were maintained under hygromycin selection to retain the presence of the HygroTK cassette, and I was able to confirm the presence of this cassette during each attempted experiment by genomic DNA PCR, suggesting that the failure to select for recombinant clones was not due to spontaneous loss of the HygroTK cassette. I also attempted to use an alternative selectable marker to HygroTK, hypoxanthine-guanine phosphoribosyltransferase (HPRT), but was also unsuccessful with this method (data not shown). The reasons for the difficulty in maintaining the functionality of this system remain unclear, as a similar approach has been employed before (Krebs et al., 2014). The most likely explanation seems to be that the presence of the loxP exchange cassette in the large gene desert, which initially seemed advantageous in limiting confounding influences of active regulatory regions, might facilitate the silencing of the HygroTK gene in the presence of ganciclovir, resulting in loss of HygroTK expression and subsequently resistance to ganciclovir.

Ultimately, at this stage, I have been unable to answer the key question regarding the interdependency between chromatin accessibility, DNA methylation and CpG density. While the creation of chromatin accessibility at the HCG1 and HCG2 bacterial sequences

is encouraging, without methylated versions of these sequences, mutations in putative transcription factor binding sites, or a more thorough panel of CpG-rich or CpG-poor sequences, it is not possible to draw strong conclusions from this experiment. Future work will therefore focus on an alternative, CRISPR/CAS9-mediated strategy to target similar bacterial sequences to explore this important aspect of chromatin biology.

4.3 Discussion

Elevated CpG density and hypomethylated DNA are essential characteristics of the majority of mammalian promoters and are required for normal gene expression and regulation. However, how these two features influence chromatin structure and accessibility has not been fully characterised. To address this gap in our knowledge, I have attempted to address the possibility of a causal relationship between CpG density and chromatin accessibility. In addition to characterising a close relationship between CpG density and chromatin accessibility across tissues and ESCs (Figure 4.2.1 & Figure 4.2.2), I was able to establish that this was independent of transcriptional activity at CGI promoters (Figure 4.2.3 & Figure 4.2.4). This suggested that the underlying CpG density and DNA methylation state might be an important determinant in shaping chromatin accessibility, either directly or indirectly. Examination of chromatin accessibility in cells lacking DNA methylation supported a inhibitory effect of DNA methylation on accessibility at many elements distal to gene promoters (Figure 4.2.5). However, even in the absence of DNA methylation, few CpG-rich sequences displayed increased chromatin accessibility as would be predicted if CpG density was sufficient to mediate accessible chromatin (Figure 4.2.6 & Figure 4.2.7). I therefore attempted to experimentally address this by targeting bacterial CpG-rich sequences into the mammalian genome (Figure 4.2.8), and although the results suggested that these CpG-rich sequences gained chromatin accessibility *in vivo* (Figure 4.2.9), I was unable to sufficiently expand these experiments

to arrive at a robustly supported conclusion. Ultimately, it is therefore not possible at this stage to confirm or reject the hypothesis that elevated chromatin accessibility at CGI promoters is an inherent feature of CpG-rich, non-methylated DNA.

4.3.1 Evidence for and against CpG density directly influencing chromatin structure.

Biophysical and biological assays have previously supported a causal link between CpG density and chromatin structure or accessibility, but the conclusions and consequences of these studies still remain highly speculative and, in some cases, contradictory. Examination of DNA sequence determinants of nucleosome positioning have indicated that CpG-rich DNA sequences have a preference to form nucleosomes *in vitro* (Kaplan et al., 2009). This is in direct contrast with observations presented here in ESCs (Figure 4.2.1), and many other *in vivo* studies were CGIs exhibit nucleosome-free or accessible chromatin structures (Ramirez-Carrozzi et al., 2009, Choi, 2010, Deng et al., 2013). Indeed, such *in vitro* conclusions have been called into question by reconstitution experiments which identified that *in vitro* nucleosome occupancy only recapitulated *in vivo* positioning in the presence of ATP-dependent processes (Zhang et al., 2011). Nevertheless, CpG density may still contribute towards nucleosome positioning *in vivo*, as C- or G- containing dinucleotide pairs were enriched at nucleosome-forming sequences *in vivo* (Gupta et al., 2008). This may be consistent with a CpG- and GC-skew seen at promoters at the +1 nucleosome position (Segal et al., 2006, Hartono et al., 2015), but remains highly speculative. Although not directly addressing nuclease sensitivity in the same manner as this thesis, these varied observations suggest an altered affinity for nucleosome-binding of CpG-rich DNA that might reflect an instability of chromatin at CGIs which may be captured as elevated nuclease sensitivity. However, the best evidence to support a role for CpG density was recently generated by a highthroughput approach integrating 12,000 short (175 bp), exogenous DNA sequences into the ESC genome and

performing DNase-seq (Hashimoto et al., 2016). This experimental approach, similar to my attempts at RMCE-mediated integration of bacterial DNA sequences, revealed an over-representation of the CpG-dinucleotide in *k*-mer enrichment analysis. Intriguingly, this enrichment, although significant, was less than that for several transcription factor motifs, suggesting that CpG density might facilitate or support accessibility determined by transcription factor binding. A more careful exploration of this data, ideally combined with the DNA methylation state for the same DNA sequences, is required though to fully understand the breadth of this observation, but it is in fitting with a hypothesis that CpG density influences chromatin structure.

4.3.2 DNA methylation-dependent pathways to establish or constrain accessibility

If the CpG density of DNA sequences does contribute to chromatin accessibility, it appears to almost certainly depend upon the DNA methylation state of those sequences. This is supported by the loss of chromatin accessibility at regulatory elements upon the gain of DNA methylation (Palacios et al., 2010, Han et al., 2011, Thurman et al., 2012, Yang et al., 2012, Taberlay et al., 2014), enhanced nucleosome compaction of methylated DNA sequences *in vitro* (Lee and Lee, 2012, Collings et al., 2013), and the reactivation of distal regulatory elements and some repetitive classes upon loss of DNA methylation (Figure 4.2.5 & Figure 4.2.7; (Domcke et al., 2015, Maurano et al., 2015)). Although the majority of CGIs in any given cell type are non-methylated, the mechanisms by which they remain refractory to DNA methylation machinery are poorly understood (Bird, 2002, Schubeler, 2015). Given the breadth and coincidence of chromatin accessibility and hypomethylated DNA at CGIs across different cell types, it seems possible that the mechanisms by which CGIs avoid DNA methylation may be similar to the pathways by which they achieve an elevated chromatin accessibility. For example, there is some evidence to suggest that ZF-CxxC domain-containing proteins help to maintain the hypomethylated state at CGIs. For

example, CFP1 and MLL1/2 mediate H3K4me3 deposition at hypomethylated CGIs (Clouaire et al., 2012, Denissov et al., 2014), and the presence of this histone modification has been proposed to be antagonistic to the DNA methyltransferase DNMT3L (Ooi et al., 2007, Rose and Klose, 2014). Indeed, a H3K4me3-insensitive mutant of DNMT3A resulted in increased DNA methylation at normally hypomethylated CGI promoters (Noh et al., 2015). Similarly, deletion of KDM2B, a ZF-CxxC domain-containing protein which binds CGIs genome-wide, resulted in hypermethylation of a subset of CGIs (Boulard et al., 2015), suggesting the possibility that this family of proteins may protect against the DNA methylation machinery. This may not be a generic feature of ZF-CxxC domain proteins, as deletion of CFP1 does not result in this hypermethylation (Clouaire et al., 2012), and this model remains speculative. It seems possible that ZF-CxxC domain-containing proteins could function to directly or indirectly regulate chromatin accessibility. For example, knockdown of MLL1 results in reduced chromatin accessibility at a CpG-rich enhancer (Jeong et al., 2014), although this was probably due to altered transcription factor binding. This hypothesis will be addressed more specifically in Chapter 5 through the deletion of specific ZF-CxxC domain-containing proteins. Alternatively, there is evidence that methylation-dependent repressive pathways might constrain chromatin accessibility in the presence of DNA methylation. For example, reduced levels of methyl-CpG binding protein MeCP result in increased nuclease sensitivity at individual loci (Antequera et al., 1989) and the recruitment of HDACs or other chromatin remodelling complexes by methyl-CpG binding proteins could be similarly important (Ng et al., 1999, Yoon et al., 2003, Du et al., 2015, Ludwig et al., 2016). These observations offer some idea about how methylated or non-methylated CpG dinucleotides might constrain or facilitate an elevated chromatin accessibility of CGI promoters. However, significant additional work must be done to address these points with quantitative genome-wide chromatin accessibility assays.

Examination of chromatin accessibility in the absence of DNA methylation revealed the activation of many CpG-poor distal regulatory elements, but only a surprisingly small subset of CpG-rich repetitive elements. If non-methylated CpG-rich DNA was truly sufficient to establish chromatin accessibility, either through the presence of chromatin opening, or the absence of chromatin compacting, regulatory mechanisms, this would not have been expected. However, repetitive elements which did not gain accessibility were evolutionarily older, suggesting that other repressive mechanisms at these sites might limit accessibility. For example, many retroelements or retroviruses are specifically targeted by H3K9me3 methyltransferases with sequence-specific zinc-finger proteins (Wolf and Goff, 2009, Rowe et al., 2010, Rowe et al., 2013, Jacobs et al., 2014). In many cases this H3K9me3-mediated gene repression is largely maintained in the absence of DNA methylation (Matsui et al., 2010, Walter et al., 2016). Recently acquired repetitive elements may not yet have evolved this additional layer of repressive activity, and therefore may be more susceptible to reactivation and increased chromatin accessibility upon loss of DNA methylation. Indeed, the study of H1 turnover dynamics revealed that DNA methylation had little effect, while loss of H3K9me3 resulted in significant increase in histone turnover dynamics (Melcer et al., 2012), suggesting that this modification and associated processes might mediate chromatin compaction and structure away from CGI promoters (Verschure et al., 2005, Grewal and Jia, 2007, Hahn et al., 2013). Alternatively, repetitive elements can be recognised and bound by many different transcription factors and are subject to tissue-specific re-activation (Wang et al., 2007, Bourque et al., 2008, Schmidt et al., 2012, Wang et al., 2012, Jacques et al., 2013), and so the re-activation of only a subset of elements may also partly be determined by ESC-specific transcription factor binding events. Understanding these alternative pathways associated with DNA methylation-sensitive repetitive elements could therefore provide an insight into how DNA methylation may or may not contribute towards chromatin structure and accessibility at other regulatory elements.

4.3.3 Transcription factors to shape hypomethylated and hypersensitive chromatin

In addition to generic methylation-dependent or ZF-CxxC-mediated chromatin modifying activities, sequence-specific transcription factors have been implicated in maintaining both hypomethylated DNA and chromatin accessibility. Certain transcription factors have been demonstrated to bind methylated DNA sequences and subsequently block maintenance of methylation at those loci (Höller et al., 1988, Brandeis et al., 1994, Macleod et al., 1994, Stadler et al., 2011, Feldmann et al., 2013, Krebs et al., 2014, Long et al., 2016), presumably through a steric hindrance to DNMT function. It has also been widely established that transcription factor binding can either directly or indirectly establish chromatin accessibility at regulatory elements (Biddie et al., 2011a, Thurman et al., 2012, Pajoro et al., 2014, Schulz et al., 2015, Iwafuchi-Doi et al., 2016). This raises the possibility that CpG-rich chromatin might achieve its hypomethylated and nuclease hypersensitive state through the binding of such transcription factors. For example, the ubiquitously expressed SP1 and CTCF or the widely expressed REST, are transcription factors that have been demonstrated to be sufficient for both inhibiting DNA methylation and establishing chromatin accessibility (Höller et al., 1988, Macleod et al., 1994, Wood et al., 2003, Ooi et al., 2006, Wang et al., 2010, Stadler et al., 2011, Feldmann et al., 2013). Furthermore, CGI promoters are occupied by a large number and variety of transcription factors, partly because many transcription factors binding motifs are GC-rich (Deaton and Bird, 2011) or contain a CpG (Blattler and Farnham, 2013), but also due to the elevated DNA accessibility at these regulatory elements which facilitates access of transcription factors to the underlying DNA sequence motifs (Thurman et al., 2012).

In spite of the elevated transcription factor binding at CGIs, CpG-rich regulatory elements actually have significantly fewer sequence motifs than CpG-poor regulatory elements (unpublished observations). This apparent contradiction might make sense when other,

more generically permissive features of CGIs are considered. For example, at CpG-rich loci fewer transcription factor binding events might be required to establish a hypomethylated and accessible state, as transcription factor-driven chromatin accessibility could be reinforced by generic chromatin-modifying pathways found at CGIs. In contrast, due to their lower CpG density and subsequently lower levels of such supporting activities, non-CGI promoters or regulatory elements might require a higher number of cooperative transcription factor binding events to overcome DNA methylation and other repressive activities. The concerted action of ubiquitously or widely expressed transcription factors and generic chromatin remodelling enzymes may therefore allow the reinforcement of a robustly accessible chromatin state at CGI promoters across many different cell types, while activity and structure at CpG-poor promoters would be more dependent upon fine-tuning by tissue-specific transcription factors. Although this model has yet to be explored satisfactorily, some supporting evidence comes from the observation that the H3K4me3 methyltransferase MLL1 is required for FOXA1 and ER-α DNA binding at a CpG-rich enhancer (Jeong et al., 2014). Importantly, this model is not incompatible with DNA methylation blocking transcription factor binding events for specific factors (Iguchi-Ariga and Schaffner, 1989, Mancini et al., 1999, Bell and Felsenfeld, 2000, Palacios et al., 2010, Medvedeva et al., 2014, Domcke et al., 2015, Maurano et al., 2015), as initial binding of methylation-insensitive transcription factors could be sufficient to reduce DNA methylation sufficiently to allow methylation-sensitive DNA binding events. This requirement for cooperativity is supported by work on the methylation-sensitive NRF1, which requires the binding of neighbouring methylation-insensitive transcription factors (eg. CTCF or REST) in order to maintain normal binding (Domcke et al., 2015). Ultimately, substantial work remains to understand the underlying sequence-based principles of CGI and non-CGI recognition and binding by transcription factors, particularly in the context of DNA methylation and chromatin accessibility. Nevertheless, the binding and action of transcription factors at CGI promoters provides a compelling explanation

that could reconcile the highly coincident nature of DNA methylation and chromatin accessibility across a wide range of different tissues.

4.3.4 Conclusions and future directions

Although I was able to describe a clear relationship between DNA methylation, CpG density and chromatin accessibility, unfortunately I was unable to determine the precise contribution of non-methylated CpG-rich sequences towards CGI chromatin accessibility. Future work will therefore focus on development of a different high throughput assay to interrogate how CpG density and DNA methylation influence chromatin accessibility of exogenous sequences in the mammalian genome. A careful examination of transcription factor motifs and the DNA methylation status of such sequences will be essential. Alternatively, a candidate-based approach will be taken to examine chromatin-modifying enzymes or pathways that might regulate chromatin structure (eg. ZF-CxxC domain-containing proteins or methyl-binding proteins) (see Chapter 5). Finally, the interplay between transcription factor binding and chromatin-modifying pathways should be examined more closely, as a possible means by which chromatin structure or features could influence transcription factor binding and function at gene promoters. The elucidation of the molecular determinants of chromatin accessibility at CGI promoters will ultimately be essential to fully appreciate their contribution towards gene regulation and transcription.

CHAPTER 5 - REGULATION OF CGI ACCESSIBILITY BY CHROMATIN-MODIFYING PATHWAYS

5.1 Introduction

CpG islands (CGIs) are intimately associated with vertebrate gene promoters (Saxonov et al., 2006, Long et al., 2013b), and, in addition to having elevated chromatin accessibility, they display distinct chromatin features that are thought to be important for normal gene regulation (Blackledge and Klose, 2011, Deaton and Bird, 2011). This unique chromatin environment at CGIs is driven partly by ZF-CxxC domain-containing proteins, which are able to recognise and bind non-methylated CpG dinucleotides through their ZF-CxxC domain (Long et al., 2013a). This diverse group of proteins also exhibit histone-modifying activities, which contribute towards shaping the unique histone modification profiles observed at CGIs (Rose and Klose, 2014). For example, the histone demethylases KDM2A and KDM2B (which I will refer to collectively as KDM2A/B) are responsible for removing H3K36me1/me2 at CGI promoters (Tsukada et al., 2006, Blackledge et al., 2010), while MLL1, MLL2 and CFP1 contribute to the deposition of H3K4me3 (Mikkelsen et al., 2007, Illingworth et al., 2010, Thomson et al., 2010, Denissov et al., 2014). The concerted action of these ZF-CxxC domain-containing proteins creates a H3K36me2-depleted/H3K4me3-enriched chromatin state which may be important for normal transcription from CGI-associated gene promoters (Blackledge and Klose, 2011). However, CGIs also act as nucleation sites for polycomb repressive complexes 1 (PRC1) and 2 (PRC2), partially through the ZF-CxxC domain-containing KDM2B, which forms a complex with PRC1 and recruits it to CGI chromatin (Farcas et al., 2012, He et al., 2013, Wu et al., 2013). Although KDM2B occupies CGIs throughout the genome, enrichment of polycomb group (PcG) complexes and their associated histone modifications (H2AK119ub1 and H3K27me3 for PRC1 and PRC2 respectively) is only achieved at a

subset of CGIs, suggesting additional molecular cues must exist to transition these sites into a predominantly PcG-occupied state. Together these observations have established CGI chromatin as a nucleation site for chromatin-modifying complexes and has led to the proposal that antagonism between activating and repressive chromatin-modifying complexes, and the resolution of individual CGIs into specific states, may underpin how CGIs influence gene transcription (Blackledge and Klose, 2011, Klose et al., 2013, Voigt et al., 2013)

The presence of these activities at CGIs may also be relevant towards understanding chromatin accessibility at CGI promoters. For example, H3K4me3 is found at all CGI promoters regardless of gene activity (Guenther et al., 2007, Mikkelsen et al., 2007, Zentner and Henikoff, 2013), reminiscent of the largely invariant accessible chromatin state at CGIs (Chapter 4). Fittingly, H3K4me3 is often discussed in the context of marking open chromatin (Gaspar-Maia et al., 2009, Krishnakumar and Kraus, 2010, Bell et al., 2011, Wan et al., 2013), although a direct link between H3K4me3 and chromatin accessibility has yet to be convincingly demonstrated. Intriguingly, one study suggested that MLL1 may help to maintain nucleosome depletion at a non-methylated distal regulatory element (Jeong et al., 2014), although the ability to promote nuclease hypersensitivity by this and other ZF-CxxC domain-containing proteins at CGIs has not been examined. In fitting with an antagonism between active and repressive activities at CGI promoters, PcG-occupied CGIs have been described to exhibit reduced chromatin accessibility, consistent with PcG complexes playing a role in chromatin compaction (Dellino et al., 2004, Francis et al., 2004, Margueron et al., 2008, Grau et al., 2011). Despite the negative correlation between PcG occupancy and nuclease hypersensitivity (Bell et al., 2010), whether PcG complexes directly regulate chromatin accessibility at their CGI target sites *in vivo* remains unclear. Although ZF-CxxC domain-containing complexes and their associated histone-modifying activities have been heavily studied in the context of gene regulation, little effort has been placed on considering the possibility that they also

might serve to promote and/or constrain chromatin accessibility at CGI-associated gene promoters.

5.2 Results

5.2.1 ZF-CxxC domain-containing proteins and chromatin accessibility at CGIs

Previous studies, and work in this thesis (Chapter 4), have suggested that non-methylated CpG dinucleotides might contribute to chromatin accessibility at CGI promoters, and it seems possible that the action of ZF-CxxC domain-containing proteins could be required to achieve this. In order to address this possibility, I examined chromatin accessibility using assay for transposable-accessible chromatin (ATAC)-seq in mouse embryonic stem cells (ESCs) where I could conditionally ablate the expression of a panel of ZF-CxxC domain-containing proteins. In particular, I examined the contribution of FBXL19, of unknown function, the H3K4me3 methyltransferase complex subunit CFP1, and the histone demethylases KDM2A/B. As described previously, I used an experimentally determined set of non-methylated islands (NMIs), rather than the computationally predicted CGIs, in order to more accurately identify non-methylated, CpG-rich DNA at gene promoters in ESCs.

5.2.1.1 FBXL19 does not regulate chromatin accessibility in ESCs

One of the least studied ZF-CxxC domain-containing proteins is FBXL19 and how it contributes to CGI function remains unknown. Recent unpublished work in the Klose lab has confirmed that FBXL19 binds NMIs throughout the genome in ESCs and that this relies on its ZF-CxxC domain (Supplementary Figure 4; courtesy of E. Dimitrova). It was therefore interesting to examine whether FBXL19, and its ability to occupy CGIs, might contribute to chromatin structure and accessibility at NMI promoters. To test this hypothesis, I exploited a conditional mouse ESC line in which a single exon encoding the

ZF-CxxC domain of the *Fbxl19* gene is flanked by loxP recombination sequences, and which also expresses Cre-ERT, a tamoxifen-inducible form of Cre recombinase (*Fbxl19*$^{CxxC\text{-}flox}$; Figure 5.2.1 A). Addition of tamoxifen results in the translocation of Cre recombinase to the nucleus and excision of the ZF-CxxC domain-encoding exon. This effectively deletes the portion of the mRNA that encodes the ZF-CxxC domain but leaves the open reading frame in register and the remainder of the protein intact. Although, we can detect removal of this exon in genomic DNA and at mRNA level (Figure 5.2.1 B; courtesy of E. Dimitrova) there is currently no antibody of sufficient quality to confirm this deletion at the protein level. However, removal or mutation of the ZF-CxxC domain in FBXL19 transgenes abolishes its ability to bind NMIs and by extension this should render the protein produced in tamoxifen-treated *Fbxl19*$^{CxxC\text{-}flox}$ ESCs unable to bind chromatin (Supplementary Figure 4; courtesy of E. Dimitrova).

I therefore performed ATAC-seq upon the *Fbxl19*$^{CxxC\text{-}flox}$ ESCs with and without tamoxifen treatment (Figure 5.2.1 C). Visual examination of several NMI promoters did not reveal any apparent differences in the ATAC-seq signal between the tamoxifen-treated and wild type cells (Figure 5.2.1 C), and this was consistent with genome-wide analysis at NMI promoters (Figure 5.2.1 D). Although there were no significant quantitative changes in the accessibility of NMI promoters, I wondered whether there might be other qualitative differences, such as a change in nucleosome positioning. However, analysis aimed at addressing this possibility revealed no apparent change in nucleosome phasing or positioning upon loss of FBXL19 targeting (Figure 5.2.1 E). These observations suggest that FBXL19 contributes little to NMI chromatin structure or accessibility in ESCs. Interestingly however, recent work in the Klose lab has revealed that FBXL19 plays an important role in facilitating gene expression during ESC differentiation, but has few effects on gene expression in normally maintained ESC cultures (E. Dimitrova; unpublished observations). In future work it may be interesting to address whether FBXL19 contributes to chromatin features of NMI promoters during differentiation.

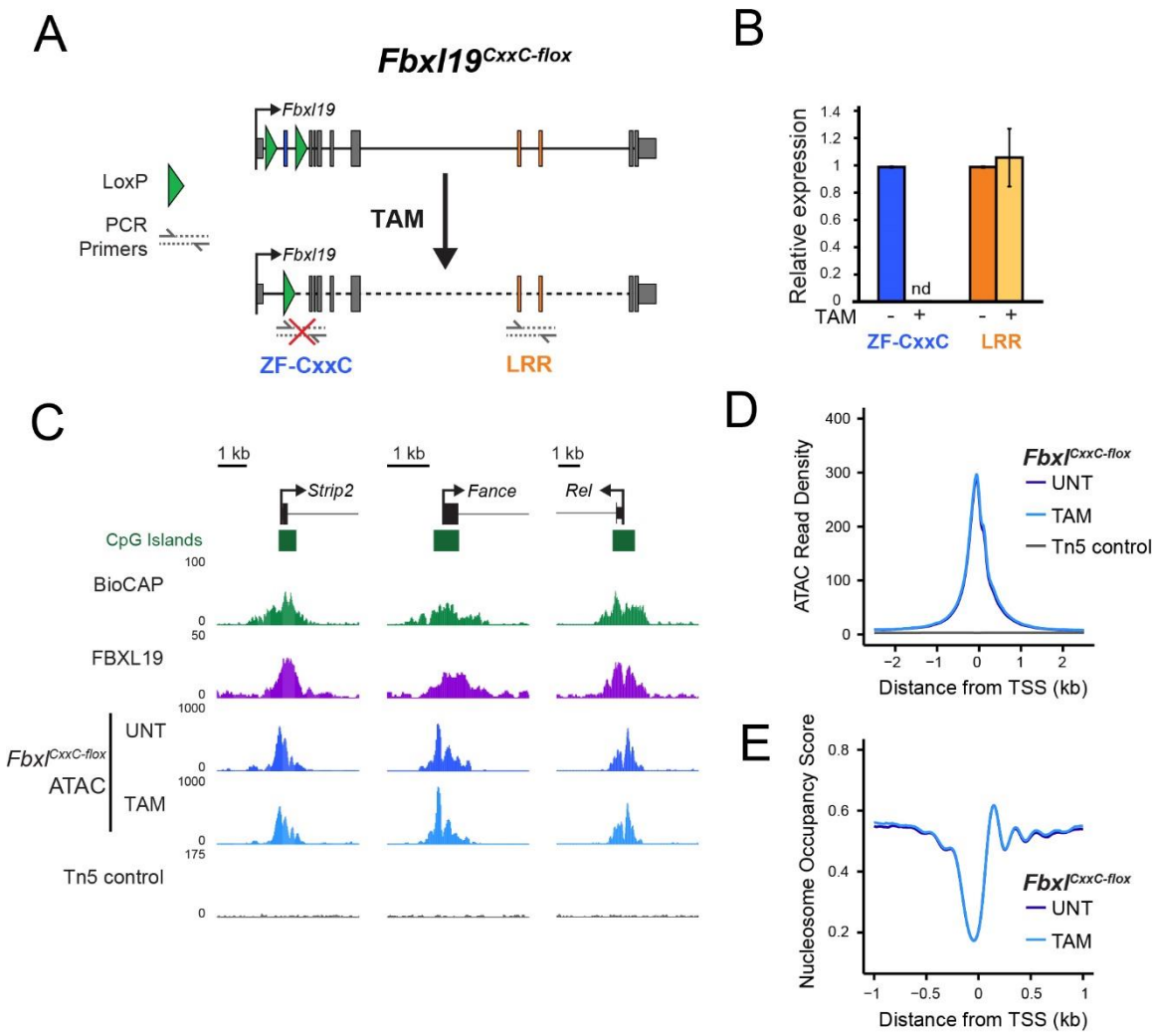


Figure 5.2.1 Loss of ZF-CxxC domain in FBXL19 does not affect chromatin accessibility.

A) Schematic depicting the *Fbxl19*$^{CxxC\text{-}flox}$ conditional ESCs. The ZF-CxxC domain-encoding exon is flanked by loxP recombination sites. In the presence of tamoxifen (TAM), Cre recombinase-ERT translocates to the nucleus and recombines the two loxP sites, excising the ZF-CxxC domain-encoding exon.

B) Quantitative PCR of *Fbxl19* mRNA levels using primers specific for exon(s) encoding either the ZF-CxxC or the leucine rich-repeat (LRR) domain, as depicted in (A). Treatment with tamoxifen (TAM) for 96 h, results in ablation of ZF-CxxC domain-encoding exons, while the remainder of mRNA is intact. Error bars represent ± standard deviation; *n* = 3. nd denotes “not detected”.

C) Genome snapshots of three NMI TSS, depicting enrichment of FBXL19, and ATAC-seq in the *Fbxl19*$^{CxxC\text{-}flox}$ ESCs after treatment with tamoxifen for 96 h.

D) Metaplot profile of *Fbxl19*$^{CxxC\text{-}flox}$ ATAC-seq signal at NMI TSS after treatment with tamoxifen for 96h.

E) Nucleosome positioning at NMI TSS in *Fbxl19*$^{CxxC\text{-}flox}$ ESCs after treatment with tamoxifen for 96h.

5.2.1.2 Chromatin accessibility at CFP1-dependent NMIs

The enrichment of H3K4me3 is a central feature of NMI promoters, and it was therefore important to consider the potential for this modification to contribute towards chromatin accessibility. The elevated H3K4me3 at NMIs is driven through the combined activity of MLL1, MLL2 and CFP1. As an initial step towards understanding the relationship between H3K4me3 and chromatin accessibility, I chose to perturb H3K4me3 levels through conditional deletion of CFP1, which disrupts SETD1 occupancy and activity in depositing H3K4me3. To achieve this, I took advantage of a mouse ESC line in which exons 4 and 5 of the *Cfp1* gene are flanked by loxP recombination sites, in addition to expressing Cre-ERT (*Cfp1*$^{fl/fl}$). Addition of tamoxifen to the *Cfp1*$^{fl/fl}$ ESCs introduces a frameshift in the open reading frame and CFP1 protein is lost from the cells within 96 hours (Figure 5.2.2 A). This results in a large (~50 %) reduction in total H3K4me3 (Figure 5.2.2 B; courtesy of Vincenzo di Cerbo). In collaboration with David Brown and Vincenzo di Cerbo, I performed ATAC-seq in the *Cfp1*$^{fl/fl}$ ESCs in the presence or absence of tamoxifen, while they profiled CFP1 chromatin occupancy and H3K4me3 by ChIP-seq (Figure 5.2.2 C). As expected from the western blot analysis of H3K4me3, CFP1-occupied NMI promoters demonstrated substantial reductions in H3K4me3 after loss of CFP1 (Figure 5.2.2 C). A visual comparison of the ATAC-seq signal between untreated and tamoxifen-treated cells revealed no apparent difference in accessibility at the majority of sites examined (Figure 5.2.2 C). Intriguingly, however, there was a small minority of sites, such as the promoter of the *Abhd16a* gene, which appeared to decrease slightly in their chromatin accessibility. When I examined the ATAC-seq signal across all NMI promoters no apparent change in the average level of accessibility was observed (Figure 5.2.2 D). This suggested that CFP1, and its regulation of SETD1-mediated H3K4me3 deposition, did not have major effects on chromatin accessibility at NMIs in ESCs. This was surprising, given the widespread association of H3K4me3 with chromatin accessibility and NMIs, and the reports linking H3K4me3 with open chromatin. Therefore, I set out to carry out a more careful analysis, taking into account the changes in H3K4me3 following deletion of CFP1.

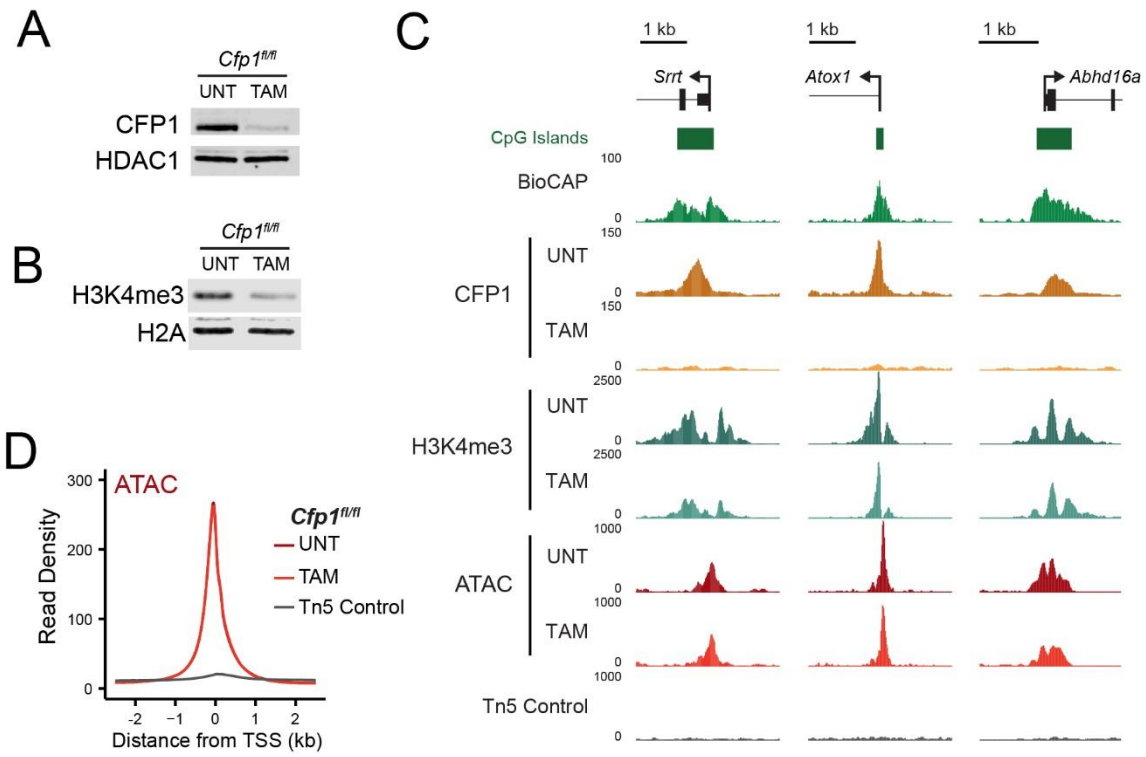


Figure 5.2.2 Conditional deletion of CFP1 results in reduced H3K4me3 with no effect on NMI accessibility.

A) Western blot analysis of *Cfp1*$^{fl/fl}$ ESC nuclear extracts after treatment with tamoxifen (96h).

B) Histone extracts analysed by western blot reveal global reduction in H3K4me3 levels after deletion of CFP1.

C) Genomic screenshots of several NMI promoters, depicting CFP1 and H3K4me3 ChIP-seq, and ATAC-seq, in the *Cfp1*$^{fl/fl}$ ESCs after treatment with tamoxifen (96h).

D) Metaplot analysis of *Cfp1*$^{fl/fl}$ ATAC-seq at ESC-active NMI promoters.

There appears to be a high degree of redundancy between CFP1/SETD1 and the other H3K4me3 methyltransferases MLL1/2 (Denissov et al., 2014). Although there are global reductions in H3K4me3 upon deletion of CFP1 (Figure 5.2.2 B), it was apparent that promoters were variably dependent upon CFP1 for H3K4me3 (Figure 5.2.2 C). Therefore, to better understand the contribution of CFP1 to H3K4me3, and to examine in finer detail the potential relationship between H3K4me3 and chromatin accessibility, I ranked gene promoters by their loss of H3K4me3 following deletion of CFP1 (Figure 5.2.3 A). This confirmed that although the majority of sites experienced some reductions in H3K4me3,

there were a subset of sites that were highly dependent upon CFP1 to maintain normal levels of H3K4me3 (Figure 5.2.3 B upper panel), and conversely, some sites that were independent of CFP1 (Figure 5.2.3 B lower panel). I reasoned that by examining these different groups in more detail I might uncover a relationship between H3K4me3 and chromatin accessibility at gene promoters that was not apparent from the meta-analysis in Figure 5.2.2. However, even the gene promoters with the largest reductions in H3K4me3 (Figure 5.2.3 B,C upper panels), showed only a very slight reduction in average accessibility (Figure 5.2.3 C). Clearly the relevance of such a small effect on accessibility is hard to interpret. Therefore, to examine this more extensively, differential analysis was performed to identify if any promoters were significantly affected in their chromatin accessibility upon deletion of CFP1. As expected, the vast majority of promoters showed no significant changes in their chromatin accessibility in the absence of CFP1, but there was a small number of promoters (n = 181) which showed significant reductions in their ATAC-seq signal (Figure 5.2.3 D). Intriguingly, gene promoters with reduced chromatin accessibility appeared to coincide with more dramatic H3K4me3 reductions than sites that were unchanged (Figure 5.2.3 E), and this difference was statistically significant (Figure 5.2.3 F). Promoters with CFP1-dependent ATAC-seq signal tended to have lower CpG densities (Figure 5.2.3 A,G), which might reflect a lower level of redundancy with MLL1/2 at these promoters (Supplementary Figure 5 (Denissov et al., 2014)), due to the co-transcriptional recruitment of CFP1/SETD1 by RNA polymerase II (Ng et al., 2003, Lee and Skalnik, 2008). However, the CFP1-dependent effects on chromatin accessibility were very modest, and the high degree of redundancy in H3K4me3 methyltransferase activity, as evidenced by considerable CFP1-independent H3K4me3, makes it difficult to confidently assign any clear relationship between H3K4me3 and chromatin accessibility. In order to understand the contribution of H3K4me3 to NMI chromatin structure and function it will therefore be essential to develop tools to conditionally ablate all H3K4me3 methyltransferase activity from cells.

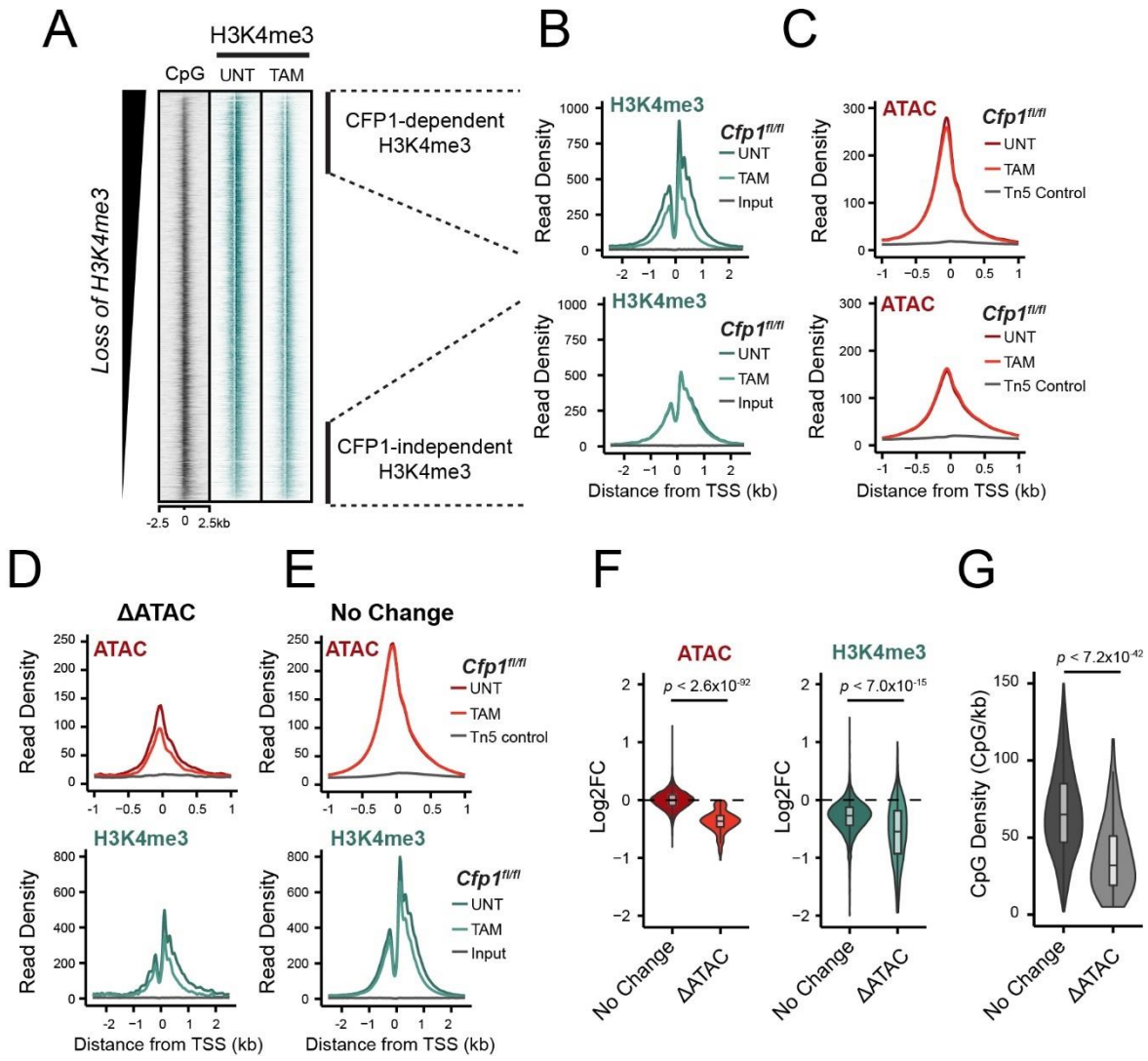


Figure 5.2.3 Chromatin accessibility at CFP1-dependent or CFP–independent H3K4me3 promoters.

A) Active TSS were ranked by their log2 fold change in H3K4me3 after tamoxifen treatment of *Cfp1*$^{fl/fl}$ ESCs. Sites at the top are most dependent upon CFP1 to maintain normal H3K4me3 levels. CpG density (CpG) is also profiled.

B) The top and bottom 20% of TSS, based upon their change in H3K4me3, reveal CFP1-independent H3K4me3 (bottom panel) and CFP1-dependent H3K4me3 (upper panel).

C) Same sites as in (B), but profiling of ATAC-seq reveals very small effect specific to CFP1-dependent H3K4me3 TSS.

D) Active TSS with a significant reduction in ATAC-seq signal (ΔATAC; *n* = 181) were profiled for ATAC-seq and H3K4me3 levels after treatment of *Cfp1*$^{fl/fl}$ ESCs with tamoxifen.

E) Same as in (D), but for active TSS with no significant change in ATAC-seq signal (No Change; *n* = 10726).

F) Comparison of log2 fold change in ATAC-seq and H3K4me3 ChIP-seq for CFP1-dependent (ΔATAC) or CFP1–independent (No Change) ATAC-seq signal at TSS ± 500 bp.

G) Comparison of CpG density for CFP1-dependent (ΔATAC) or CFP1–independent (No Change) TSS ± 500 bp.

5.2.1.3 The KDM2 histone demethylases regulate NMI chromatin accessibility

The ZF-CxxC domain-containing proteins KDM2A and its paralogue KDM2B (referred to collectively as KDM2A/B) play an important role in removing H3K36me2 from NMIs (Blackledge et al., 2010), but the outcomes of this and any other KDM2A/B-dependent activities in NMI promoter structure and function are unknown. I therefore set out to address if KDM2A/B might contribute towards the elevated chromatin accessibility of NMIs. However, in addition to the high degree of similarity between KDM2A and KDM2B, the study of KDM2A/B in mammalian cells is complicated by the presence of different isoforms, often referred to as long-form or short-forms (KDM2A/B-LF and KDM2A/B-SF respectively) (Long et al., 2013a). All isoforms are targeted to NMIs through their ZF-CxxC domains, but only the LF isoforms contain the catalytic JmjC histone demethylase domain. Otherwise, the biological role of the SF isoforms is presently unknown. To explore the relative contributions of the KDM2A/B LF and SF isoforms to NMI chromatin accessibility, I exploited a conditional mouse ESC line developed by Anne Turberfield in which the *Kdm2a* and *Kdm2b* genes have LoxP recombination sites flanking coding exons (Figure 5.2.4 A), and which also express the Cre-ERT recombinase. Addition of tamoxifen results in translocation of Cre-ERT to the nucleus, and recombination of the loxP sites to induce a frameshift in the LF isoforms of KDM2A/B. In this conditional line (*Kdm2.LF*$^{fl/fl}$) the TSS of the SF isoforms were then constitutively deleted by CRISPR/Cas9 (Anne Turberfield), generating a *Kdm2.LF*$^{fl/fl}$*;SF*$^{-/-}$ ESC line (Figure 5.2.4 A). Addition of tamoxifen to these *Kdm2.LF*$^{fl/fl}$*;SF*$^{-/-}$ cells therefore ablates KDM2A/B-LF isoforms, and result in ESCs which lack all KDM2A/B isoforms (Figure 5.2.4 A,B).

Therefore, in collaboration with Anne Turberfield, we performed ATAC-seq on the four different isoform combinations (Figure 5.2.4). Importantly, this ATAC-seq experiment was spiked with *Drosophila melanogaster* cells, as described previously for ChIP-seq approaches (Orlando et al., 2014, Hu et al., 2015), in order to have an internal calibration.

This calibration would therefore allow us to capture any subtle, genome-wide effects that might occur upon deletion of KDM2A/B. We chose to first examine the two extremes of this cell culture system, the *Kdm2.LF*$^{fl/fl}$ cells (with all KDM2A/B isoforms intact) and the *Kdm2.LF*$^{fl/fl/}$*;SF*$^{-/-}$ treated with tamoxifen, which lack both LF and SF isoforms, reasoning that if KDM2A/B had any contribution towards chromatin accessibility, it would be most apparent in this context. Intriguingly, visual examination of ATAC-seq signal at several NMI promoters revealed modest (*Ep400, Tango6*) or dramatic (*Cdh1*) reductions in their chromatin accessibility upon loss of KDM2A/B protein from ESCs (Figure 5.2.5 A). To explore the breadth of this observation, I compared ATAC-seq signal between *Kdm2.LF*$^{fl/fl}$ and *Kdm2.LF*$^{-/-}$*;SF*$^{-/-}$ cells. Surprisingly, this revealed that nearly all NMI promoters experienced reduced chromatin accessibility after deletion of KDM2A/B (Figure 5.2.5 B), and this reduction was significant for 29.2 % of NMI promoters (4205/14416; Figure 5.2.5 C). This seemed to indicate a potentially important contribution of these ZF-CxxC domain-containing proteins to chromatin accessibility at a large number of NMI promoters.

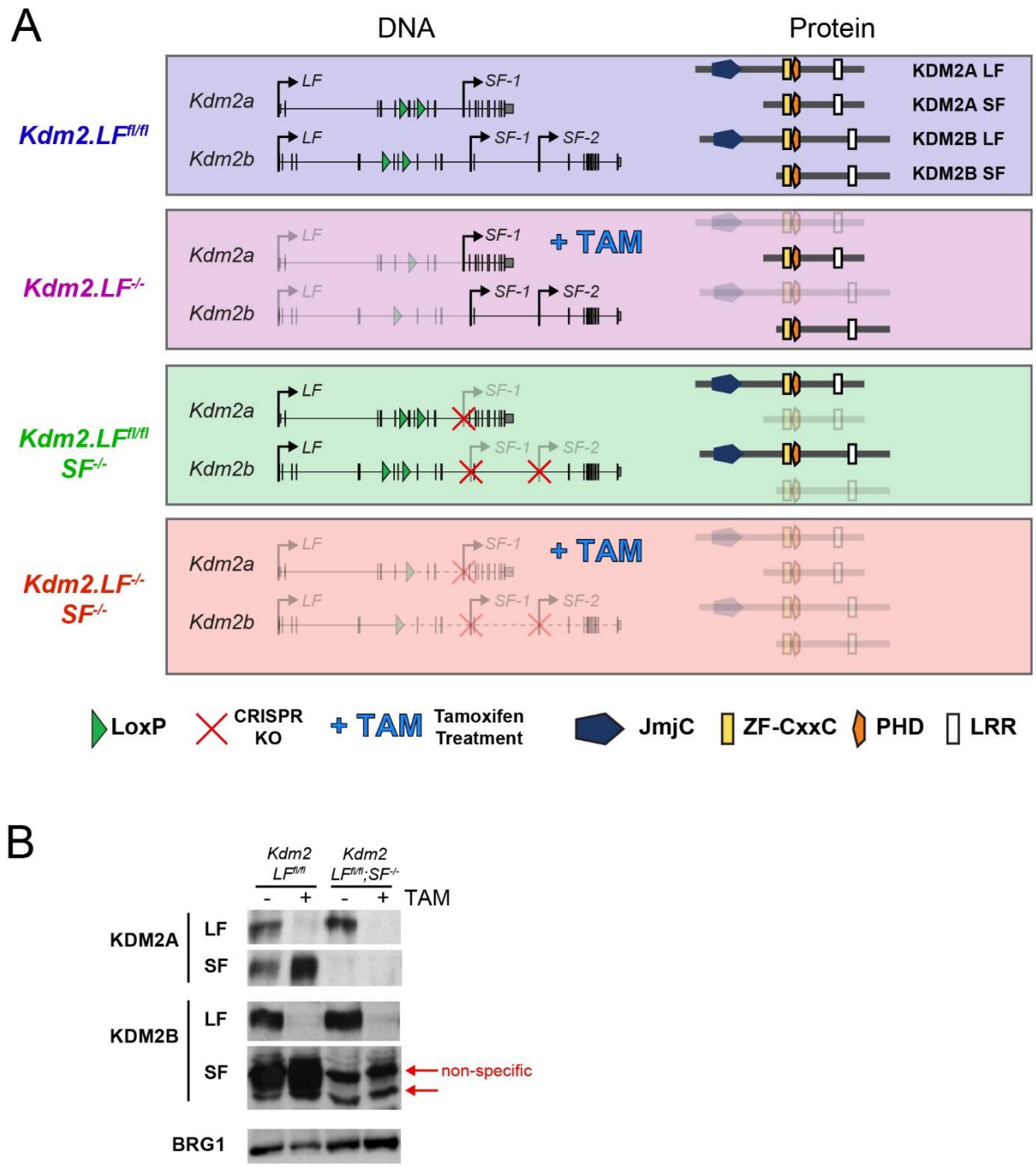


Figure 5.2.4 Mouse ESC cell lines to study KDM2A/B isoform function.

A) Schematic depicting conditional deletion of *Kdm2a* and *Kdm2b* long-form (LF) in the *Kdm2.LF*$^{fl/fl}$ ESCs, and constitutive null of the *Kdm2a* and *Kdm2b* short-forms (SF) with CRISPR/Cas9 in the *Kdm2.LF*$^{fl/fl}$*;SF*$^{-/-}$ ESCs. Treatment of either *Kdm2.LF*$^{fl/fl}$ or *Kdm2.LF*$^{fl/fl}$*;SF*$^{-/-}$ ESCs with tamoxifen (TAM) results in Cre-mediated recombination of loxP recombination sites specific to LF isoforms, resulting in a frameshift and loss of KDM2A/B-LF protein.

B) Western blot analysis of nuclear extracts from ESCs containing all wild type KDM2A/B isoforms (*Kdm2.LF*$^{fl/fl}$), or lacking all KDM2A/B (*Kdm2.LF*$^{-/-}$*;SF*$^{-/-}$) in untreated or tamoxifen (TAM)-treated cells (96 h).

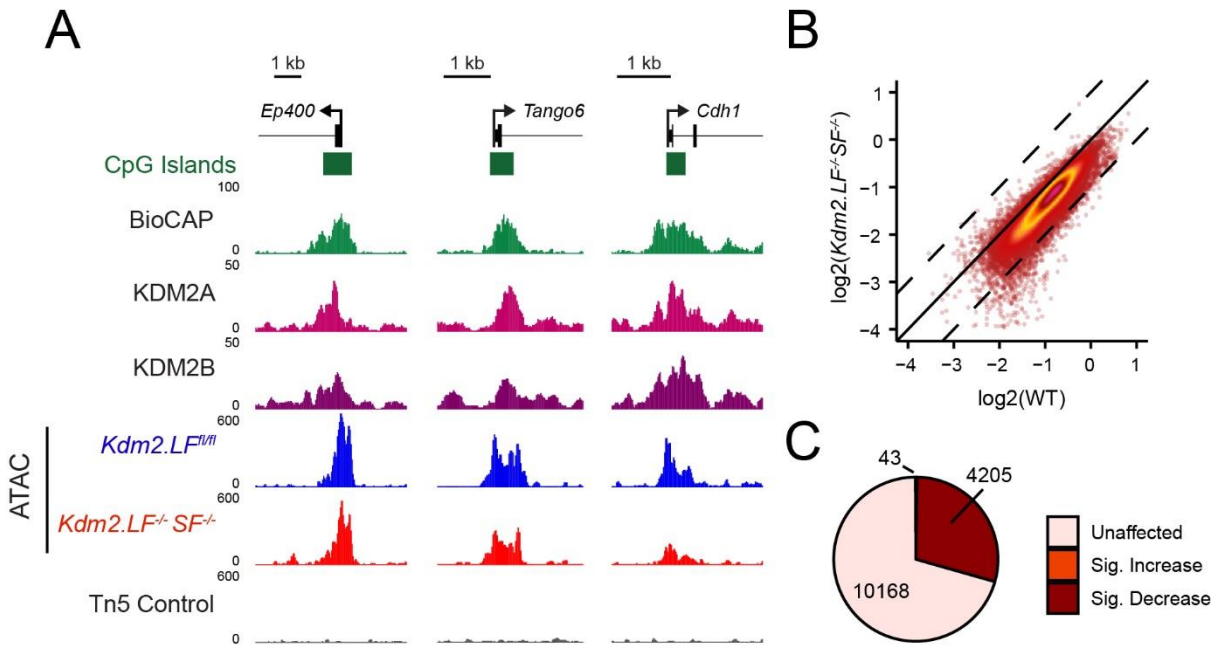


Figure 5.2.5 Chromatin accessibility in the absence of KDM2A and KDM2B.

A) Genomic snapshot of several NMI promoters, visualising wild type KDM2A and KDM2B ChIP-seq, and *Kdm2.LF$^{fl/fl}$* and *Kdm2.LF$^{-/-}$;SF$^{-/-}$* ATAC-seq.

B) Scatterplot comparing log2 ATAC-seq reads per kilobase (RPK) for wild type *Kdm2.LF$^{fl/fl}$* or *Kdm2.LF$^{-/-}$;SF$^{-/-}$* ESCs at NMI active promoters (ATAC hypersensitive peaks). Black line corresponds to 1:1 ratio between treated:untreated, while dotted lines represent 2-fold reduction or increase in signal.

C) Significant changes in calibrated ATAC-seq at NMI promoters were identified using DESeq2.

5.2.1.4 KDM2-dependent accessibility is driven by KDM2A/B-SF isoforms

Although we observed that KDM2A/B promoted chromatin accessibility at NMIs, it was unclear whether this was related to their role as histone demethylases. To address this question, I considered the ATAC-seq data generated from ESCs which lacked either the KDM2A/B LF or SF isoforms (*Kdm2.LF$^{-/-}$* or *Kdm2.LF$^{fl/fl}$;SF$^{-/-}$* respectively) (Figure 5.2.4). In this way, I could examine whether the LF isoforms, which contain the JmjC demethylase domain, or the SF isoforms were responsible for the KDM2-dependent chromatin accessibility. Quantitation of ATAC-seq revealed that loss of either the KDM2A/B LF or SF isoforms resulted in widespread reductions in promoter accessibility (Figure 5.2.6 A), although there was an appreciably greater effect in the *Kdm2.LF$^{fl/fl}$;SF$^{-/-}$* which lacked

KDM2A/B-SF expression. Indeed, there were many more significant reductions upon deletion of only the SF isoforms (22.4%; 3228/14416), although loss of the LF isoforms also resulted in a large number of significant changes (9.4%; 1349/14416) (Figure 5.2.6 B). Sites that relied on the LF or SF isoforms for their accessibility were not completely concordant (Figure 5.2.6 C). In general, quantitative comparison of the relative changes in ATAC-seq signal suggested that there did not appear to be a trend towards LF- or SF-specific effects (Figure 5.2.6 D). Importantly, the reductions in ATAC-seq signal tended to be larger upon loss of the SF isoforms, suggesting that the contribution of KDM2A/B towards promoter chromatin accessibility was most likely not related to their histone demethylase activity, but is rather driven by the C-terminal domains shared between LF and SF isoforms. The observation that deletion of KDM2A/B-SF resulted in the greatest reductions in chromatin accessibility is consistent with the fact that KDM2A/B-SF expression is approximately 5-fold higher than KDM2A/B-LF (Anne Turberfield personal communications). However, the mechanism by which the KDM2A/B SFs are able to promote chromatin accessibility at NMI promoters remains completely unknown. These observations nevertheless represent an exciting insight into how ZF-CxxC domain-containing proteins contribute to chromatin accessibility at NMIs.

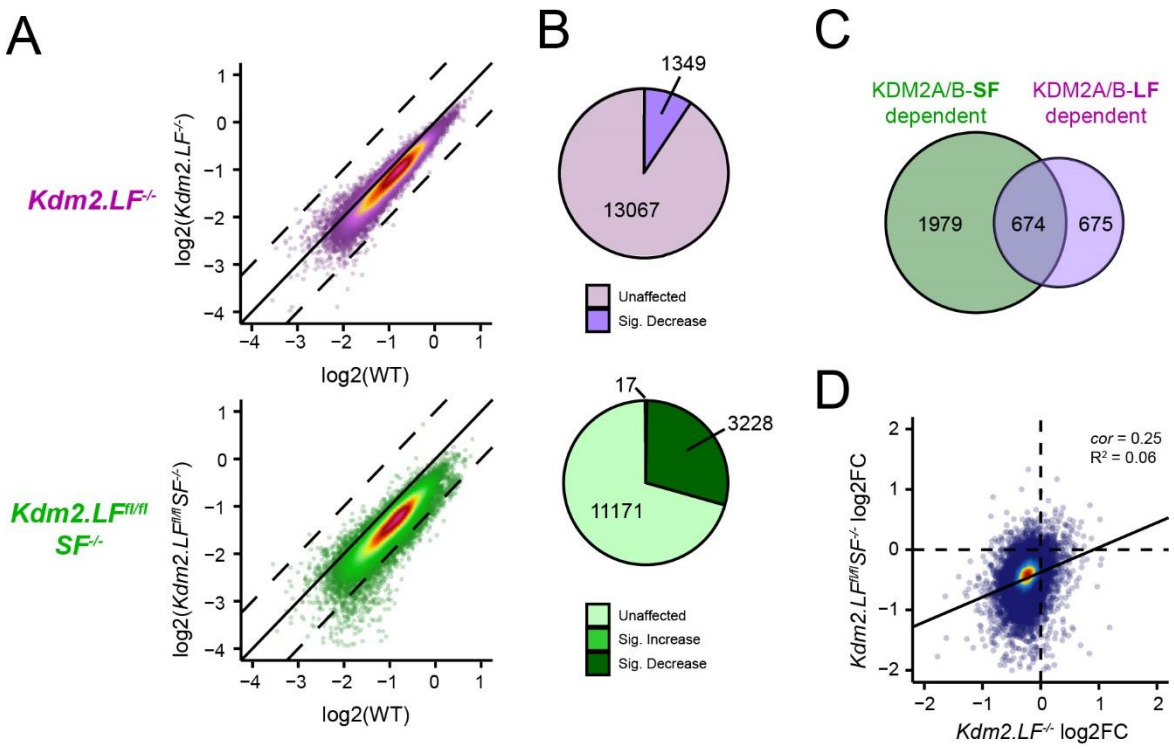


Figure 5.2.6 KDM2-SF isoforms are responsible for majority of KDM2-dependent chromatin accessibility.

A) Comparison of ATAC-seq RPK of ESCs with wild type (WT) KDM2A/B isoforms compared to ESCs expressing short form (SF)-only (*Kdm2.LF*$^{-/-}$) or long form (LF)-only (*Kdm2.LF*$^{fl/fl}$*;SF*$^{-/-}$). Intervals used for quantitation represent promoter ATAC hypersensitive sites.

B) Annotation of ESC-active promoter intervals with significant reductions in ATAC-seq signal for different *Kdm2* conditional or constitutive null ESCs.

C) Comparison of promoters significantly reduced in the *Kdm2.LF*$^{fl/fl}$*;SF*$^{-/-}$ or *Kdm2.LF*$^{-/-}$ ESCs.

D) Comparison of log2 fold changes in ATAC-seq signal for *Kdm2.LF*$^{-/-}$ with *Kdm2.LF*$^{fl/fl}$*;SF*$^{-/-}$ ESCs.

5.2.2 The regulation of chromatin accessibility by polycomb repressive complexes

NMIs are targeted by activities and complexes associated with both gene activation and repression, and this duality represents an important feature of these regulatory elements (reviewed in Klose et al., 2013, Voigt et al., 2013). Although the roles of several putatively "activating" ZF-CxxC domain-containing complexes are addressed above, it was important to examine chromatin accessibility at NMIs that are repressed in ESCs. In particular, a subset of NMIs are occupied by polycomb repressive complexes (PRC1 and

PRC2), and exhibit high levels of the PcG-associated histone modifications (H2AK119ub1 and H3K27me3 respectively). PcG proteins, particularly members of PRC1, have been implicated in chromatin compaction *in vitro* (Francis et al., 2004, Margueron et al., 2008, Bell et al., 2010, Grau et al., 2011), but a detailed dissection of chromatin accessibility at PcG targets *in vivo* has yet to be achieved. I therefore set out to characterise chromatin accessibility at PcG-occupied NMI promoters in ESCs and to examine the contribution of PRC1 and PRC2 complexes towards this PcG-associated state.

5.2.2.1 Polycomb-occupied NMIs have reduced chromatin accessibility

To begin to understand the relationship between PcG occupancy and chromatin accessibility, I examined ATAC-seq and DNase-seq profiles in wild type ESCs at PcG-occupied and non-PcG NMIs (Figure 5.2.7). This indicated that NMIs occupied by PcG components RING1B (PRC1), SUZ12 (PRC2) and marked by H3K27me3, tended to have reduced chromatin accessibility compared to NMIs of comparable size, CpG density and DNA methylation levels (Figure 5.2.7 A). Indeed, when NMI promoters were sorted by H3K27me3 levels, it was clear that NMIs enriched for H3K27me3 and PcG complexes exhibited reduced chromatin accessibility, as measured by DNase-seq or ATAC-seq (Figure 5.2.7 B,C), in fitting with previous reports (Bell et al., 2010, Calabrese et al., 2012, Beck et al., 2014, Deaton et al., 2016). Furthermore, PcG-occupied NMIs were enriched for ATAC-seq fragment sizes consistent with nucleosome occupancy, in contrast to non-PcG NMIs which tended to exhibit fragment sizes consistent with nucleosome-free DNA (Figure 5.2.7 D). This supports the contention that PcG-occupied NMIs exist in a state enriched for nucleosomes with reduced DNA accessibility.

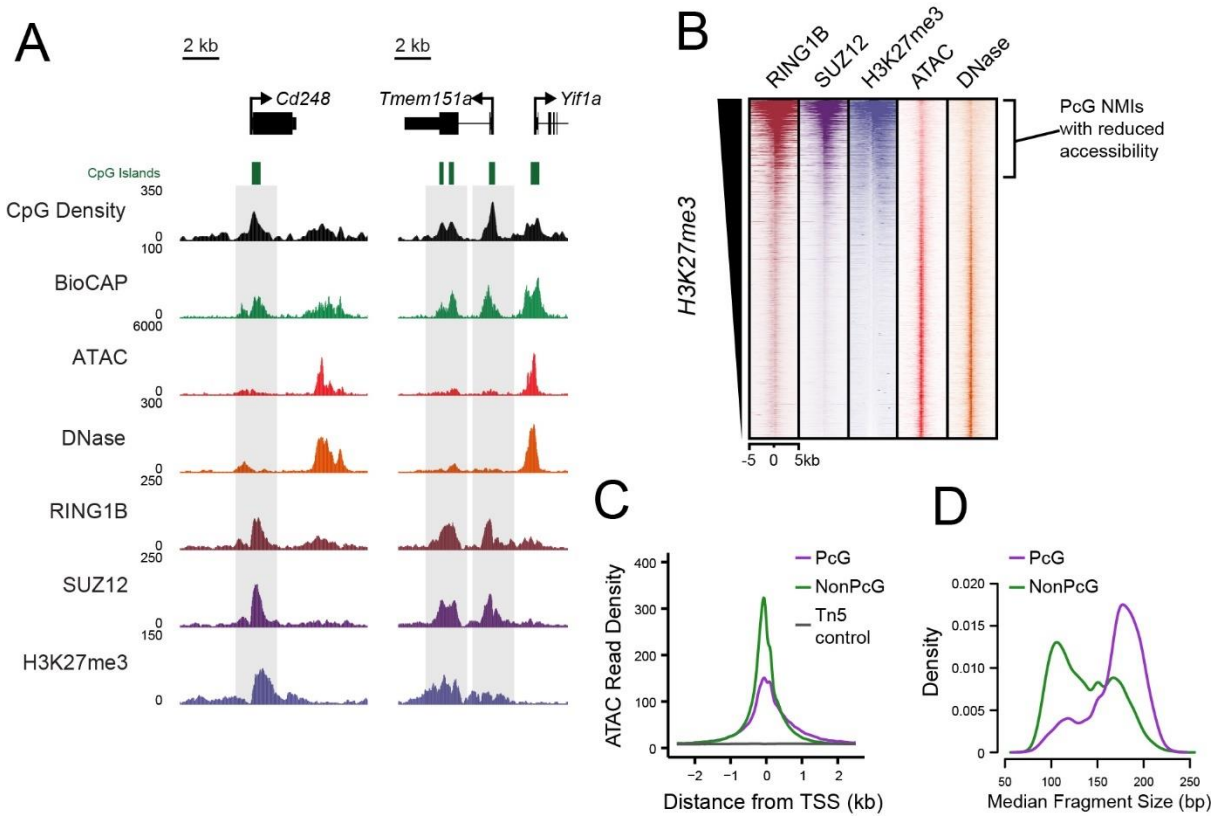


Figure 5.2.7 Reduced chromatin accessibility at PcG-occupied NMIs.

A) Genomic snapshot of several PcG-occupied NMI promoters (grey boxes). Two non-PcG NMIs are included for comparison.

B) Heatmap analysis of NMI promoters, ranked by enrichment of H3K27me3. RING1B, SUZ12 and H3K27me3 are included to depict PcG chromatin occupancy and activity, while ATAC-seq and DNase-seq provide two measures of chromatin accessibility.

C) Metaplot of ATAC-seq signal at PcG-occupied NMI or non-PcG NMI promoters. PcG targets were identified by presence of both RING1B (PRC1) and SUZ12 (PRC2) enrichment.

D) Median ATAC-seq fragment sizes for PcG or non-PcG NMI promoters. PcG-occupied NMIs have larger median fragment sizes consistent with nucleosomal fragment sizes.

5.2.2.2 Measuring chromatin accessibility in PRC1- or PRC2-null ESCs

Given the reduced chromatin accessibility at NMI promoters occupied by PcG complexes (Figure 5.2.7), and the evidence for *in vitro* compaction of nucleosomal arrays by PcG components, particularly PRC1 (Francis et al., 2004, Grau et al., 2011), it seemed likely that PRC1 and/or PRC2 might be important for limiting the chromatin accessibility of PcG target NMIs in ESCs. To test this hypothesis, I exploited two established ESC lines to ablate PRC1 or PRC2 function. To disrupt PRC1, I used *Ring1a*$^{-/-}$*;Ring1b*$^{fl/fl}$ ESCs, in which

the PRC1 catalytic core components (RING1A and RING1B) were either constitutively null (*Ring1a*$^{-/-}$) or could conditionally deleted upon addition of tamoxifen (*Ring1b*$^{fl/fl}$). Treatment of this cell line with tamoxifen for 48 hours is sufficient to ablate all detectable RING1B protein and PRC1-deposited H2AK119ub1 (Figure 5.2.8 A), while also reducing PRC2 occupancy and activity on chromatin (Blackledge et al., 2014). Alternatively, to examine the contribution of PRC2 towards chromatin accessibility, Eed conditional knockout cells (*Eed cKO*) were used, which express a doxycycline-sensitive *Eed* transgene (*Eed*^{TG}) in an *Eed*$^{-/-}$ background (Ura et al., 2008). In the presence of doxycycline, *Eed*^{TG} is not expressed, leading to loss of EED protein (Figure 5.2.8 B). Although EED is not the catalytic component of PRC2, its loss destabilises the PRC2 complex (Ura et al., 2008, Tavares et al., 2012), thereby resulting in loss of H3K27me3 (Figure 5.2.8 B). Having verified the capacity to ablate PRC activity using these two independent approaches in ESCs, I performed ATAC-seq in the *Ring1a*$^{-/-}$*;Ring1b*$^{fl/fl}$ and *Eed cKO* ESCs in order to understand whether PcG systems repress chromatin accessibility at NMIs. Surprisingly, when I compared the ATAC-seq profiles in the PRC1-null or PRC2-null cells to their matched wild type controls, there was no apparent change in chromatin accessibility at PcG-occupied NMIs (Figure 5.2.8 C), and this appeared to be true at PcG-occupied NMIs genome-wide (Figure 5.2.8 D). This result was unexpected given the extensive body of work describing a role PcG systems in chromatin compaction and genome-wide correlations between PcG occupancy and reduced chromatin accessibility.

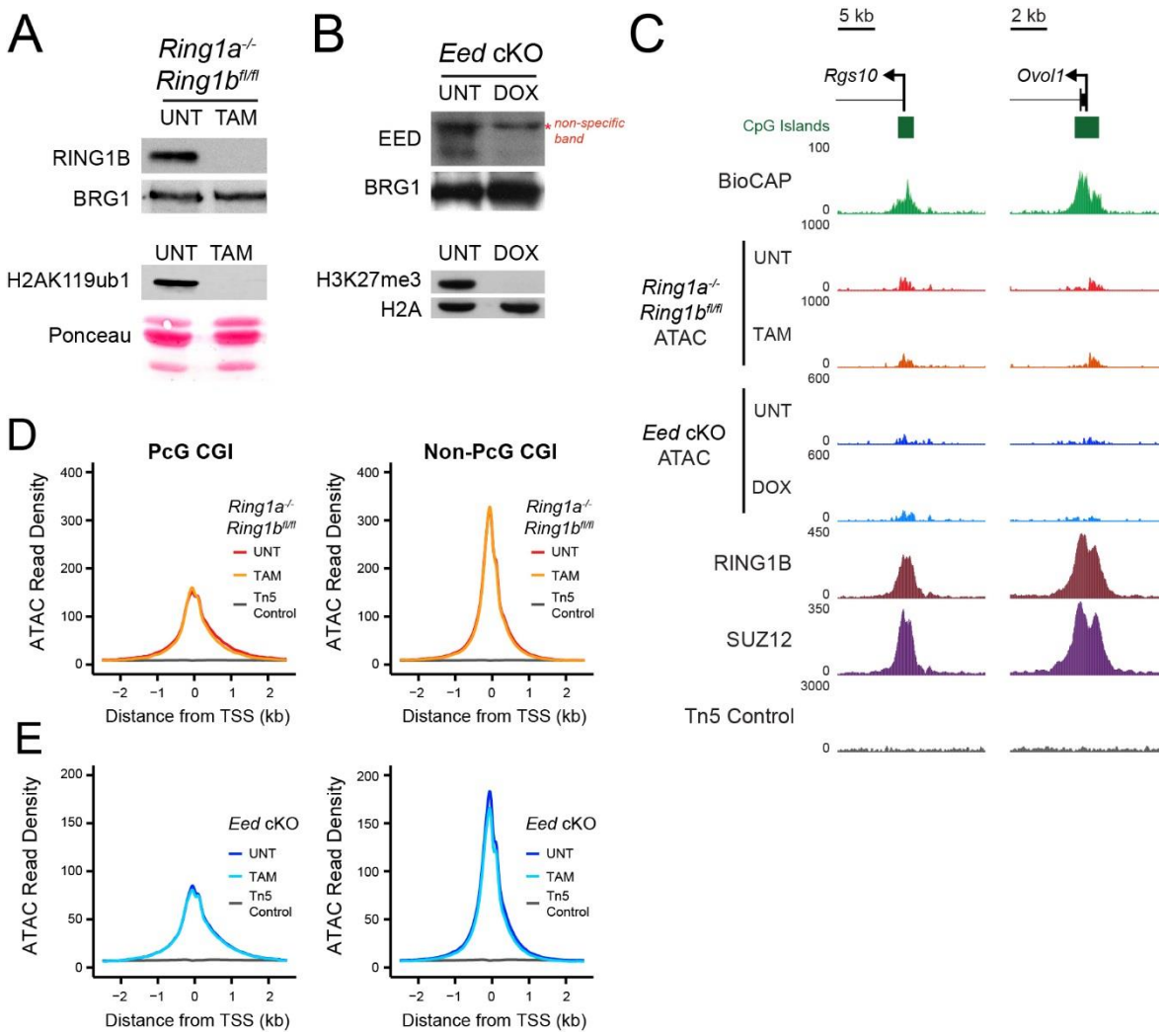


Figure 5.2.8 Loss of PRC1 or PRC2 does not increase chromatin accessibility of PcG-occupied NMIs.

A) Western blotting for *Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ ESCs after 48 h tamoxifen (TAM) treatment. Upper panel consists of nuclear extracts probed with RING1B and BRG1 (loading control), and lower panel consists of histone acid extracts probed for H2AK119ub1, with Ponceau S stain as a loading control.

B) Western blotting for *Eed* cKO ESCs after 14 days doxycycline (DOX) treatment, using either nuclear extracts (upper panel) or histone acid extracts (lower panel).

C) Genomic snapshots for two PcG-occupied NMI promoters depicting ATAC-seq in *Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ ESCs after 48 h TAM treatment, and *Eed* cKO ESCs after 14 days DOX treatment.

D) Metaplot analysis at PcG-occupied NMIs and non-PcG-occupied NMIs for *Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ ATAC-seq.

E) Same as in (D), but for *Eed* cKO ATAC-seq.

Although there were no apparent increases in chromatin accessibility of PcG-occupied NMIs after deletion of PRC1 or PRC2, several large PcG domains actually demonstrated decreases in their ATAC-seq signal (Figure 5.2.9 A), directly in contradiction with the hypothesis that PcG systems constrain chromatin accessibility at NMI promoters. This was observed in the absence of either PRC1 or PRC2, suggesting that this is a direct outcome of loss of PRC complexes. While this decrease in chromatin accessibility was only significant for a small subset of sites (101/3931 and 38/3931 PcG-occupied NMIs in the *Ring1a*$^{-/-}$*;Ring1b*$^{fl/fl}$ and *Eed cKO* ESCs respectively), it appeared to be specific to the largest PcG domains (Figure 5.2.9 B,C). These observations indicate that deletion of PRC1 or PRC2 can result in a chromatin state that is surprisingly less accessible to nuclease digestion. This unexpected result most likely represents a form of nuclease hypersensitivity influenced by the high frequency of chromatin binding and modification by PRC1 and PRC2. This fits with the low, but above background, levels of chromatin accessibility at PcG-occupied NMIs (Figure 5.2.7), which may be driven by interactions of PRC1 and PRC2 with chromatin. This may also have broader implications for understanding and interpreting nuclease hypersensitivity assays, particularly in the context of regulatory elements with high occupancy of *trans*-acting factors.

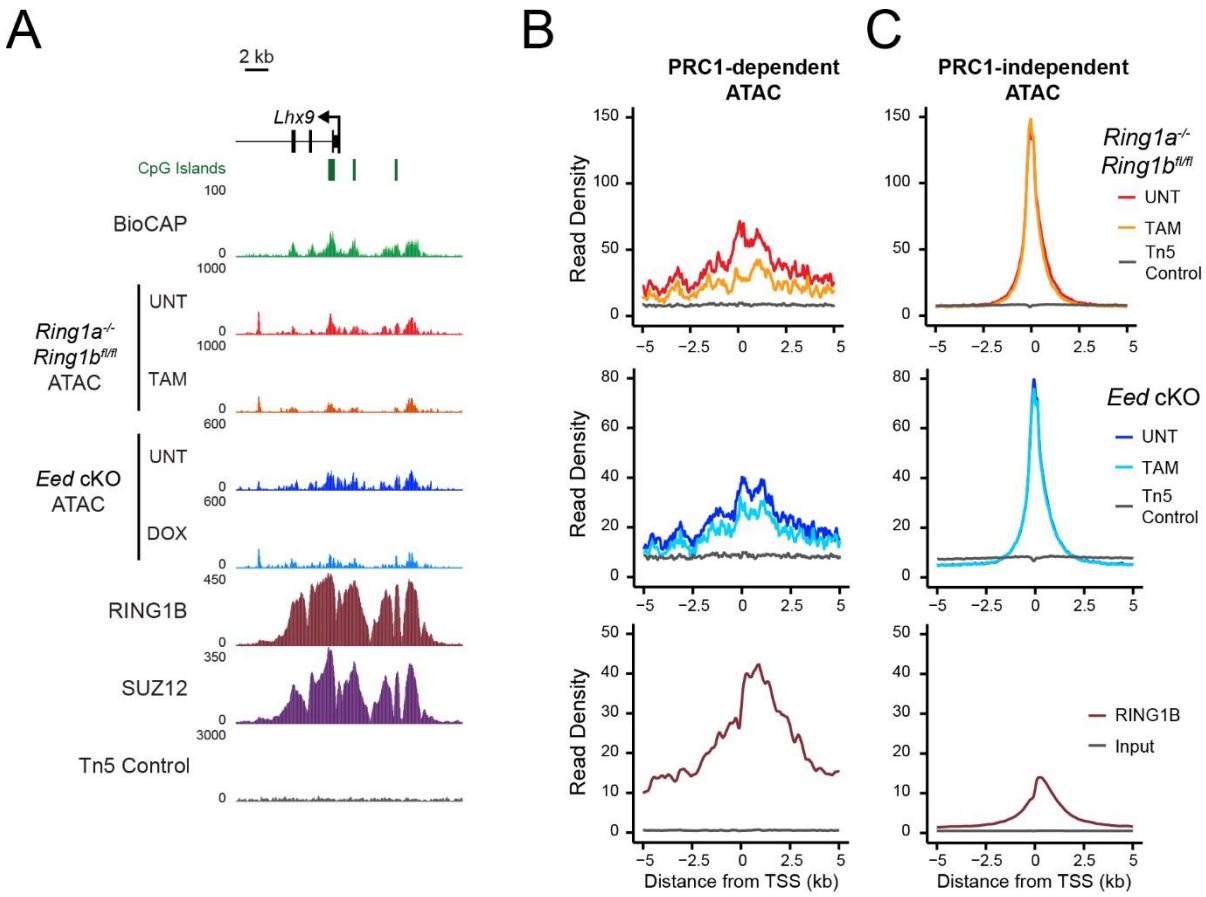


Figure 5.2.9 PcG complexes contribute to nuclease sensitivity of PcG-occupied NMIs.

A) Genomic snapshot of the *Lhx9* PcG domain, which exhibits decreased ATAC-seq signal after deletion of either PRC1 (*Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$) or PRC2 (*Eed* cKO) from ESCs.

B) Metaplot analysis at PcG-occupied NMIs which experience a significant decrease in ATAC-seq signal after tamoxifen treatment of *Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ ESCs (*n* = 101/3931). Wild type RING1B enrichment is also visualised.

C) Metaplot analysis at PcG-occupied NMIs which exhibit no significant change in ATAC-seq signal after tamoxifen treatment of *Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ ESCs (*n* = 101/3830).

By examining the contribution of PRC1 and PRC2 independently, I had hoped to identify any specific contribution of either complex to the constraint of chromatin accessibility at PcG target sites. However, there is also a certain degree of redundancy between PRC1/2 activity in gene repression (Leeb et al., 2010), and deletion of PRC1 results in wide-spread reductions, but not complete loss, of PRC2 recruitment at PcG-occupied NMIs in ESCs (Blackledge et al., 2014), and *vice versa* in PRC2-null cells (Tavares et al., 2012). This redundancy could potentially mask PcG protein function in regulating chromatin

accessibility. To address this possibility, I constitutively deleted *Eed* from the *Ring1a*$^{-/-}$*;Ring1b*$^{fl/fl}$ ESCs (Figure 5.2.10 A-D). Addition of tamoxifen to this cell line should therefore fully ablate both PRC1 and PRC2 complex formation and activities (Figure 5.2.10 E). Unfortunately, due to time constraints I was unable to perform ATAC-seq on these cells to test for redundancy between PRC1 and PRC2 in potentially limiting chromatin accessibility. An alternative explanation to such potential redundancy is that a PcG-associated factor, which is able to function independently of the PRC1 or PRC2 catalytic activities or core complex, is responsible for constraining chromatin accessibility at PcG targets. Therefore, in addition to performing ATAC-seq upon the *Eed*$^{-/-}$*;Ring1a*$^{-/-}$*;Ring1b*$^{fl/fl}$ ESCs, future work should focus upon examining different PcG auxiliary factors for any changes in chromatin accessibility at PcG-occupied NMIs.

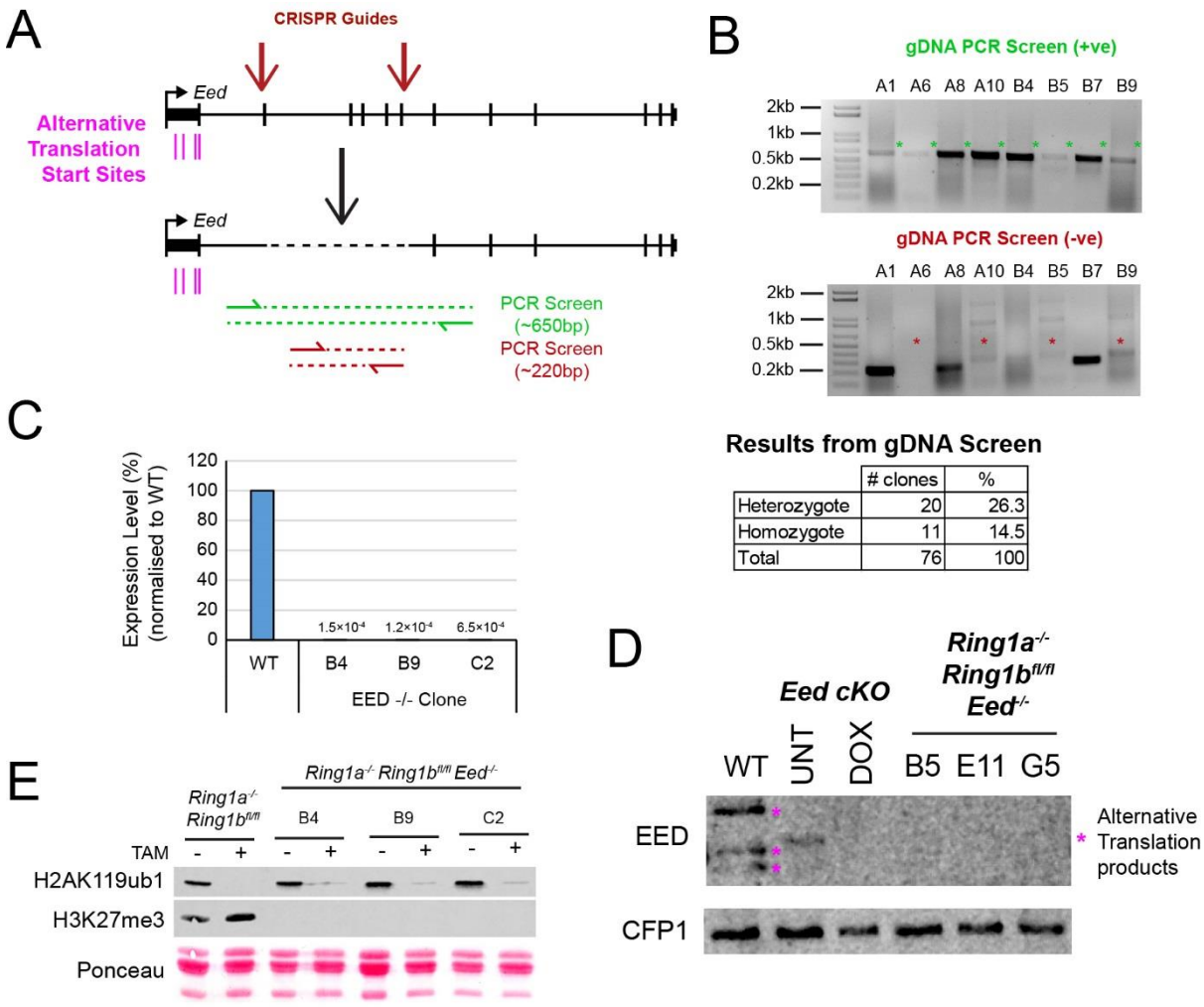


Figure 5.2.10 CRISPR/Cas9 deletion of *Eed* in *Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ ESCs.

A) Schematic detailing CRISPR/Cas9 strategy for deletion and screening of endogenous *Eed* in mouse *Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ ESCs. Guide RNAs were designed to flank five exons, and successful CRISPR/Cas9-mediated cleavage would result in a frameshift in translation products. Genomic DNA PCR screening strategy is detailed for positive PCR screen (green; gain of ~650bp product) or negative PCR (red; loss of ~220bp product).

B) Genomic DNA PCR screen for a selection of putative *Eed* KO clones, using either positive PCR product (upper panel) or negative PCR strategy (lower panel). NB Negative control is not included here, but confirmed specificity of PCR products, depicted with green * for positive band, or red * for lack of wild type product.

C) Reverse-transcriptase quantitative PCR for three clones identified by genomic DNA screen to be *Eed*$^{-/-}$, compared to wild type parental ESCs.

D) Western blotting of nuclear extracts from three *Eed*$^{-/-}$;*Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ clones. Parental nuclear extract (WT), and *Eed* cKO ESCs before and after doxycycline treatment, are included for reference.

E) Histone extracts used for western blotting confirm loss of H3K27me3 deposition in three *Eed*$^{-/-}$;*Ring1a*$^{-/-}$;*Ring1b*$^{fl/fl}$ clones. Treatment with tamoxifen (TAM) for 48 h also ablates H2AK119ub1 levels after deletion of RING1B.

5.3 Discussion

Non-methylated CGI promoters exhibit a unique signature that is broadly associated with an open and accessible chromatin structure (Chapter Four), but the causes and consequences of this accessibility have not been widely examined. Here, I have attempted to address the hypothesis that CGI chromatin accessibility is regulated by ZF-CxxC domain-containing proteins, and that the action of these diverse proteins might explain the close relationship between non-methylated, CpG-rich DNA and chromatin accessibility. Although FBXL19 and CFP1 did not appear to contribute towards the accessibility of CGI promoters (Figure 5.2.1, Figure 5.2.2 & Figure 5.2.3), deletion of KDM2A/B resulted in widespread reductions in CGI chromatin accessibility (Figure 5.2.5 & Figure 5.2.6), supporting a role for these CGI-bound proteins in promoting the elevated chromatin accessibility of CGIs. In contrast, PcG complex occupancy correlated with reduced chromatin accessibility at CGIs (Figure 5.2.7). Surprisingly, however, ablation of either PRC1 or PRC2 followed by ATAC-seq suggested that the core PcG complexes and their catalytic activity appeared not to be required for this repression of chromatin accessibility (Figure 5.2.8 & Figure 5.2.9). Together, these observations are consistent with an antagonism of chromatin accessibility by different features of CGI chromatin, although the precise mechanisms, particularly with relation to PcG-dependent chromatin compaction, remain elusive.

5.3.1 Potential mechanisms for the regulation of chromatin structure at CGIs

ZF-CxxC domain-containing proteins are responsible for shaping a unique chromatin environment at non-methylated CGIs of many gene promoters, and one the main ways they do this is through histone-modifying activities. One of the best described of these activities is the deposition of H3K4me3 by the CFP1/SETD1 complex, and the histone methyltransferases MLL1 and MLL2. In fitting with a potential role for H3K4me3 and these complexes in promoting chromatin accessibility at CGIs, recent experiments have

identified that deletion of SETD1A (which interacts with CFP1) in erythroid lineages results in reduced chromatin accessibility, as measured by FAIRE at individual loci (Li et al., 2016). This was proposed to be mediated through the reduced ability to recruit the chromatin remodelling complex NURF, but the precise contribution of H3K4me3 towards this remains to be characterised. An additional study suggested that ASH2L (a component of SETD1/MLL-containing complexes) plays a key role in limiting heterochromatin formation, and this was interpreted to be important for maintaining normal open chromatin in cells (Wan et al., 2013). MLL1 also appears to be important for facilitating nucleosome depletion at a single distal regulatory element (Jeong et al., 2014), although this may have been driven by altered transcription factor binding. However, it was unclear from these studies if reduced H3K4me3 deposition was responsible for these alterations in chromatin structure or accessibility, or if other non-enzymatic functions were involved. Certainly I was only able to observe minor effects on chromatin accessibility upon deletion of CFP1, although there were major reductions in H3K4me3 (Figure 5.2.2 & Figure 5.2.3). This may have been due to redundancy between CFP1/SET1 and MLL1/2 for H3K4me3 deposition (Denissov et al., 2014) or the partial retention of SETD1 in the absence of CFP1 (V. di Cerbo unpublished observations). There is reasonable evidence that H3K4me3 may act as a signal to recruit chromatin remodelling complexes containing H3K4me3-specific chromodomains (CHD) (Santos-Rosa et al., 2003, Flanagan et al., 2005, Pray-Grant et al., 2005, Sims et al., 2005) or plant homeodomains (PHD) (Wysocka et al., 2005, Li et al., 2006, Wysocka et al., 2006). Indeed, one such chromatin remodeller, CHD1, has been demonstrated to be required to maintain nuclease sensitivity at gene promoters (Siggens et al., 2015) or repress heterochromatic domains (Gaspar-Maia et al., 2009), although the precise contribution of H3K4me3 in recruiting this remodeller has recently been called into question (Morettini et al., 2011, Skene et al., 2014, Siggens et al., 2015). Intriguingly, another remodelling complex which specifically recognises H3K4me3, nucleosome remodelling factor (NURF), was recently linked with SETD1-dependent chromatin accessibility (Li et al., 2006, Wysocka et al., 2006, Li et al., 2016). However, the difficulty

in ablating all H3K4me3 in living cells makes it difficult to truly understand the biological relevance of H3K4me3 towards the recruitment of such chromatin remodellers, and chromatin accessibility in general.

Although ZF-CxxC-dependent histone modifications could potentially act to recruit specific chromatin remodellers, ZF-CxxC domain-containing proteins might also regulate chromatin accessibility through the deposition or removal of histone modifications that antagonise repressive pathways responsible for chromatin compaction. For example, H3K4me3 is antagonistic to DNMT and PcG activities (Ooi et al., 2007, Schmitges et al., 2011, Noh et al., 2015), potentially limiting DNA methylation- or PcG-dependent repressive activities. CGIs are depleted of H3K36me2, which given a putative role in gene repression through recruitment of HDACs (Li et al., 2009a), had seemed a likely candidate for the regulation of chromatin accessibility. However, even though KDM2A/B promote CGI accessibility, this appears to be independent of their histone demethylase domain, suggesting that the removal of H3K36me2 may not play an important role in mediating chromatin structure in mammalian cells.

If the demethylase activity of KDM2A/B is not essential for CGI chromatin accessibility, how does KDM2A/B regulate this important feature of CGI chromatin? Intriguingly, the demethylase domain of KDM2A/B-LF does not appear to be required for normal gene expression, while deletion of KDM2A/B-SF results in widespread reductions in transcription (Anne Turberfield, unpublished observations). KDM2A/B both interact with SKP1, an important factor involved with ubiquitylation and proteasomal degradation of proteins through SCF complexes (Farcas et al., 2012; Anca Farcas and Emilia Dimitrova unpublished observations, Wong et al., 2016). Intriguingly, it seems possible that KDM2A/B that may be capable of recognising substrates through their LRR domains and regulating protein turnover, as was recently proposed for KMD2B, potentially by acting directly as an E3 ligase (Janzer et al., 2012, Han et al., 2016). If KDM2A/B substrates

were associated with chromatin remodelling activities, this might provide a mechanism by which KDM2A/B could regulate CGI chromatin accessibility. Alternatively, the simple occupancy of KDM2A/B at target sites could help to maintain the hypomethylated state, as was recently proposed for KDM2B (Boulard et al., 2015). This could prevent recruitment of MBD proteins which might act to remodel chromatin and constrain chromatin accessibility (Antequera et al., 1989, Ng et al., 1999, Yoon et al., 2003, Du et al., 2015, Ludwig et al., 2016).

One intriguing possibility is that high concentrations of DNA binding events by ZF-CxxC domain-containing proteins could distort the thermal properties and/or interactions of DNA with the nucleosomal surface, rendering it sensitive to nuclease digestion. This model has previously been proposed to describe altered chromatin accessibility in the presence of transcription factor binding events on nucleosomal DNA (Polach and Widom, 1996, Anderson and Widom, 2000, Sekinger et al., 2005, Albert et al., 2007, Jiang and Pugh, 2009). This may explain the observation that the largest PcG domains experience decreased chromatin accessibility in PRC1- or PRC2-null cells (Figure 5.2.9), but might also be the case for KDM2-dependent chromatin accessibility, given their high expression levels and enrichment at CGI promoters. This could be consistent with the elevated state of CGI promoters in general, which are highly occupied by many transcription factors and chromatin-modifying enzymes (Thurman et al., 2012, Sheffield et al., 2013, Chen et al., 2014c). However, at this stage it is difficult to address this model without a better examination of the active remodelling processes that might be involved.

5.3.2 PcG-dependent or –independent chromatin compaction at CGIs

While the establishment of open and accessible chromatin at CGIs may possibly be important for CGI promoter activity, conversely, a subset of CGIs exist in a less accessible,

more compact state that is coincident with high levels of PcG repressive complexes (Figure 5.2.7) (Bell et al., 2010, Calabrese et al., 2012, Beck et al., 2014, Deaton et al., 2016). Analysis of nucleosomal DNA by electron microscopy has revealed that both *Drosophila* (Francis et al., 2004, Beh et al., 2012) and mammalian (Margueron et al., 2008, Grau et al., 2011) PcG components are capable of compacting nucleosomes *in vitro*, and combined with *in vivo* measurements of higher order chromatin interactions mediated by PRC1 (Eskeland et al., 2010, Schoenfelder et al., 2015), chromatin compaction is believed to be an important feature of PcG-dependent gene repression (King et al., 2005, Simon and Kingston, 2013). However, deletion of either PRC1 or PRC2 does not result in increased nuclease sensitivity at PcG-occupied CGIs in mouse ESCs (Figure 5.2.8), as measured by ATAC-seq It is possible that the measurement of chromatin accessibility by the Tn5 transposase may only partially capture the compact state at PcG-occupied CGIs, given an apparent size-dependent occlusion of polymerases at PcG targets (Fitzgerald and Bender, 2001). For example, the use of a nuclease with a greater molecular weight may emphasize the reduced chromatin accessibility at PcG targets and allow a greater dynamic range to detect changes in chromatin accessibility in the absence of PRC1/2. Nevertheless, using ATAC-seq it was possible to observe a reduced chromatin accessibility at PcG-occupied CGIs (Figure 5.2.7), but the accessibility of these sites did not increase upon deletion of either PRC1 or PRC2. It is therefore important to try to reconcile this surprising result with the current understanding of PcG complexes and their function at CGIs.

The PcG-dependent compaction of nucleosomes appears to be independent of the histone-modifying activities of both PRC1 and PRC2, as evidenced by compaction of tail-less histone-containing nucleosomes (Francis et al., 2004, Margueron et al., 2008) and by non-catalytic PcG components (Grau et al., 2011). How this might relate to the histone methylation-dependent chromatin compaction observed for the PcG component L3MBTL1 (Trojer et al., 2007) is unclear. In the context of my experiments, I have not only

disrupted the activity, but also the chromatin occupancy, of PRC1 and PRC2 core components, with significant effects on the alternative complex (Tavares et al., 2012, Blackledge et al., 2014). For example, in the case of PRC1-depleted cells, there is nearly a complete loss of SUZ12, EZH2 and EED (PRC2) chromatin occupancy (Endoh et al., 2008, Blackledge et al., 2014). However, the potential for the many different auxiliary PcG components to occupy chromatin independent of the core PcG complexes and their histone-modifying activities is poorly understood. Such residual chromatin binding by auxiliary PcG factors such as Polyhomeotic, BCOR and L3MBTL1/2, and even PCGF core components, could explain the maintenance of the PcG-associated chromatin state in the *Ring1a*$^{-/-}$*;Ring1b*$^{fl/fl}$ and *Eed* cKO ESCs. This will be a key area for future work, not only to understanding PcG-dependent chromatin structure at gene promoters, but for a more general appreciation of the potentially varied roles of different PcG components in the regulation of gene expression.

Studies examining PRC1-dependent nucleosome compaction in mammalian systems have identified that CBX2 (M33) and PCGF4 (BMI1) are sufficient to compact nucleosome arrays *in vitro* (Grau et al., 2011). This appears to be driven by the presence of a highly basic and disordered protein sequence that is conserved across different PcG components in different species (Grau et al., 2011, Beh et al., 2012). CBX2 and PCGF4 are both members of canonical PRC1 complexes, which are generally recruited to PcG domains through the recognition of H3K27me3 by CBX proteins (Fischle et al., 2003b, Wang et al., 2004, Bernstein et al., 2006). However, both CBX2 and PCGF4 are expressed at low levels and they form only a small minority of PRC1 complexes in ESCs (Morey et al., 2012, Kloet et al., 2016), so it is unclear what their contribution towards PcG-dependent chromatin compaction could be in this cell type. Although chromatin occupancy of CBX2 and PCGF4 in the absence of PRC1 or PRC2 has not been directly examined, chromatin occupancy of the similar PCGF2 and CBX7 are lost in the absence of PRC2 (Rebecca Lea, Neil Blackledge unpublished observations; Tavares et al., 2012),

suggesting that the function of these canonical PRC1 complexes is likely to be disrupted in both of my genetic ablation experiments with PRC1 and PRC2. However, one recent study identified that CBX2 occupancy at mitotic chromosomes was independent of both PRC1 and PRC2 (Zhen et al., 2014), suggesting that it may exist and be able to function by itself without the catalytic core of PRC1. In general, the activities or functions of PcG subunits independent of catalytic core complexes remain unclear. A better understanding of this will be essential to understand the regulation of chromatin structure and accessibility at PcG target sites.

An alternative explanation for the reduced chromatin accessibility at PcG target sites *in vivo* is the presence of additional repressive chromatin-modifying complexes that co-occupy PcG target sites in ESCs. For example, the nucleosome remodelling and deacetylase (NuRD) complex co-occupies many PcG target sites (Yildirim et al., 2011) and is believed to facilitate PRC2 recruitment and activity by removing acetylation of H3K27 (Morey et al., 2008, Reynolds et al., 2012). The chromatin remodelling activity of NuRD is determined by CHD4 (Mi2β) (Tong et al., 1998, Wade et al., 1998, Xue et al., 1998), and this activity is required for NuRD-mediated repression of transcription (Linder et al., 2007, Musselman et al., 2012, Ramírez et al., 2012). Several studies have identified a role for CHD4 in limiting chromatin accessibility at regulatory elements (Gao et al., 2009, Ramírez et al., 2012, Morris et al., 2014, Hainer and Fazzio, 2015), particularly at bivalent gene promoters (de Dieuleveult et al., 2016), suggesting that the enrichment of NuRD complexes at PcG target sites could be responsible for reduced chromatin accessibility, potentially independent of PcG complexes themselves. Intriguingly, L3MBTL2 appears to be able to interact with both PRC1 and NuRD complexes (Trojer et al., 2011, Qin et al., 2012) and is able to compact nucleosomes *in vitro* (Trojer et al., 2011), similar to L3MBTL1 (Trojer et al., 2007). This may represent a functional link between PcG- and NuRD-dependent processes that will warrant further investigation relating to the potential modulation of chromatin accessibility at CGIs.

Although the presence of repressive activities may play an important role, the absence of activating signals could also be critical to understanding chromatin structure at PcG-associated CGIs. For example, the removal of histone acetylation by deacetylases, NuRD-dependent or –independent, could also be important for limiting chromatin accessibility at PcG-occupied CGIs. Acetylation has been demonstrated to limit nucleosome compaction (Lee et al., 1993, Shogren-Knaak et al., 2006) and chromatin condensation (Görisch et al., 2005, Llères et al., 2009), and to promote chromatin accessibility (Rincon-Arano et al., 2012, Lennartsson et al., 2015, Frank et al., 2016, Pattenden et al., 2016). Indeed, it seems possible that reduced histone acetyltransferase and increased deacetylase activity at repressed or inactive NMI promoters could result in reduced chromatin accessibility, and that this could be co-incident with PcG recruitment in response to reduced transcriptional activity (Riising et al., 2014). Reduced chromatin accessibility at PcG-occupied CGIs might also reflect the absence of activating transcription factors at these loci (Guertin and Lis, 2013). For example, PcG-occupied CGIs in ESCs tend to be enriched for motifs of transcription factors not expressed in embryonic tissues (Ku et al., 2008, Mendenhall et al., 2010), and so their reduced accessibility may result from an absence of transcription factor-dependent chromatin remodelling or pioneering activities (Biddie et al., 2011a, Thurman et al., 2012, Pajoro et al., 2014, Schulz et al., 2015, Iwafuchi-Doi et al., 2016). The enrichment, or lack of, expressed transcription factor motifs at PcG targets may also reflect why only a subset of PcG-occupied promoters are re-activated upon deletion of PcG complexes (Leeb et al., 2010, Blackledge et al., 2014, Schoenfelder et al., 2015), although this remains to be explored satisfactorily. Finally, there is some evidence that chromatin compaction precedes PcG recruitment and activity, as during a time course of PcG-dependent repression decreases in chromatin accessibility were observed slightly prior to PcG complex occupancy and activity on chromatin (Yuan et al., 2012). Furthermore, PRC2 is stimulated *in vitro* nucleosome arrays with higher degrees of compaction (Yuan et al.,

2012), suggesting that PcG complexes may recognise pre-existing compact chromatin states rather than playing a key role in establishing them. It would be interesting to modulate the chromatin accessibility of different CGI promoters *in vivo* to test whether this is indeed true. However, due to the difficulty in modulating chromatin accessibility directly, and other confounding features at CGI promoters, this experiment remains technically very challenging.

5.3.3 Conclusions and Future Directions

Here I have described varied contributions of ZF-CxxC domain-containing proteins towards chromatin accessibility at CGIs, with KDM2A/B appearing to play a widespread role in promoting chromatin accessibility. Future work will certainly require examining whether KDM2A or KDM2B contribute equally to CGI chromatin accessibility, and whether specific domains within the SF isoforms are required to mediate this effect, through either recruitment or turnover of chromatin remodelling enzymes, disruption of DNA methylation-dependent processes, or the steric alteration of chromatin structure. Further work should also focus on the potential for H3K4me3 to regulate chromatin structure, particularly in the context of redundancy between MLL1, MLL2 and CFP1. Conversely, the causes and consequences of reduced chromatin accessibility at PcG-occupied CGIs remain unclear. In addition to addressing potential redundancy between PRC1 and PRC2, a better characterisation of auxiliary PcG components will be necessary, particularly CBX2, PCGF4 and L3MBTL1/2, as well as assessing the interplay between PcG and NuRD-dependent remodelling. Ultimately, the functional relevance of chromatin accessibility towards the function of CGIs at gene promoters still requires further work. However, it will be difficult to examine this question until we achieve a more defined understanding of the molecular determinants shaping chromatin structure at CGIs.

CHAPTER 6 - REGULATION OF CHROMATIN ACCESSIBILITY BY THE PIONEER FACTOR OCT4

6.1 Introduction

In addition to promoters and CGIs, many genomic regulatory elements exist in an open, chromatin state, such as enhancers, insulators and silencers. These elements are often distal to gene promoters, bound by sequence-specific transcription factors, and serve to regulate gene expression at promoters through long-range interactions (Spitz and Furlong, 2012, Todeschini et al., 2014). Although the mechanisms by which this is achieved remain to be elucidated, transcription factor binding is believed to be essential for the normal structure and function of these distal regulatory elements (Radman-Livaja and Rando, 2010, Cockerill, 2011, Guertin and Lis, 2013). In terms of accessing their DNA motif sequences, many transcription factors rely upon pre-existing chromatin accessibility or structure at their target sites to bind and exert their gene regulatory function (Guccione et al., 2006, Lupien et al., 2008, Guertin and Lis, 2010, Wu et al., 2011b), and this is consistent with work describing nucleosomes as a substantial barrier for DNA binding factors (Anderson and Widom, 2000, Sekinger et al., 2005, Struhl and Segal, 2013). For example, DNA binding of the glucocorticoid receptor in response to hormone treatment is limited to pre-accessible regions of the genome (Biddie et al., 2011b, John et al., 2011). In some cases, such DNA binding events require the action of ATP-dependent chromatin remodelling enzymes to reposition or evict nucleosomes from transcription factor binding sites to allow factors to recognise and bind their motif sequences (Wang et al., 1996, Olave et al., 2002, Lickert et al., 2004, Wu et al., 2007, Ho et al., 2009), although the requirement on such processes by different transcription factors remains to be fully understood.

While many transcription factors appear to bind pre-accessible regulatory elements, during cellular and developmental transitions, transcription factors must be able to access and bind previously inaccessible and nucleosome-occluded DNA sequences in order to establish new regulatory elements and transcriptional programs. To achieve this, cells rely upon a class of specialized transcription factors which can access their DNA sequence motifs even when associated with a nucleosome or found in inaccessible chromatin (Magnani et al., 2011, Zaret and Carroll, 2011). These so-called 'pioneer' transcription factors are thought to function by binding motifs positioned directly on the nucleosomes, or alternatively, by repositioning or displacing nucleosomes to provide access to the DNA sequence (Zaret and Carroll, 2011, Voss and Hager, 2014). This defining characteristic enables them to engage with their sequence motifs in nucleosome-rich or inaccessible regions of the genome and allow them to function as master transcriptional regulators to establish gene regulatory networks during developmental transitions or cellular reprogramming (Iwafuchi-Doi and Zaret, 2014). One of the best characterised pioneer factors is FOXA1, which was first demonstrated to be able to generate an open chromatin state in the absence of active chromatin remodelling activities (Cirillo et al., 2002), and is also capable of binding directly to nucleosomal DNA *in vivo* (Li et al., 2011b, Iwafuchi-Doi et al., 2016). Through the application of genome-wide chromatin occupancy or accessibility assays, the list of pioneer transcription factors has expanded in recent years (Iwafuchi-Doi and Zaret, 2014), but the mechanisms by which such pioneer factors interact with chromatinised DNA to establish chromatin accessibility remain unclear (Voss and Hager, 2014).

An important paradigm that has been used to understand pioneer transcription factor function is that of the master pluripotency transcription factor OCT4 (POU5F1), which is essential for normal embryonic development and the maintenance of stem cell pluripotency (Jerabek et al., 2014, Wu and Schöler, 2014). Its role as a pioneer transcription factor was established during experiments reprogramming somatic cells into

induced pluripotent stem cells (iPSC) by the Yamanaka transcription factors (OCT4, KLF4, SOX2 and c-MYC; OKSM) (Takahashi and Yamanaka, 2006, Buganim et al., 2013), of which OCT4 is both irreplaceable and sufficient for establishing pluripotency (Nakagawa et al., 2008, Kim et al., 2009, Li et al., 2011a, Jerabek et al., 2014). With the exception of c-MYC, these transcription factors are capable of binding a subset of their targets sites even when occupied by nucleosomes prior to the induction of cellular reprogramming (Soufi et al., 2012, Soufi et al., 2015). In the case of OCT4, it is thought that this pioneering binding activity may pave the way for further transcription factor binding and nucleosome displacement in order to establish functional distal regulatory elements (You et al., 2011, Shakya et al., 2015, Simandi et al., 2016). Although, OCT4's capacity to engage with previously inaccessible regions of chromatin reveals its activity as a pioneering factor during induced reprograming, these experiments have relied upon overexpression of OCT4, and the biological relevance of many of the pioneered target sites in this system remains unclear (Soufi et al., 2012, Soufi et al., 2015). Furthermore, the precise contribution of OCT4 to establishing distal regulatory elements through creating chromatin accessibility and facilitating transcription factor co-binding has yet to be explored in detail.

6.2 Results

6.2.1 OCT4 regulates chromatin structure and transcription factor binding

While the majority of OCT4 binding sites in ESCs are distal to promoters and are not located at CGIs, it seemed both important and interesting to consider how sequence-specific transcription factors might act to shape chromatin accessibility profiles in a given cell type. Whether this would offer insights into the potential mechanisms for CGI chromatin accessibility was unclear. Nevertheless, OCT4s essential role in determining the cell fate and transcriptional program of ESCs, and its reported role as a pioneer factor, made it an excellent candidate to study how sequence-specific transcription factors can regulate chromatin accessibility.

6.2.1.1 OCT4 is required for chromatin accessibility at its binding sites

The atomic structure of OCT4 and its DNA binding activity *in vitro* have been characterized in detail (Klemm et al., 1994, Reményi et al., 2001, Esch et al., 2013), yet there is still a poor mechanistic understanding of how it engages its motif sequences to regulate transcriptional activity in living cells. To understand how this interesting transcription factor engages with the genome to regulate chromatin structure, I exploited a conditional ESC line where addition of a small molecule (doxycycline) leads to loss of OCT4 expression (ZHBTC4 ESCs; Niwa et al., 2000). Prolonged removal of OCT4 from ESCs results in loss of pluripotency and cellular differentiation (Niwa et al., 2000, Adachi et al., 2013). I therefore identified an acute treatment condition where following 24 hours of doxycycline treatment cells lacked OCT4 protein (Figure 6.2.1 A) but retained normal ESC morphology, were alkaline phosphatase positive (Figure 6.2.1 B), and expressed wild type levels of the pluripotency transcription factors SOX2 and NANOG (Figure 6.2.1 A). The

maintenance of stem cell features under these treatment conditions provided an opportunity to examine where OCT4 normally engages the ESC genome and to ask how this relates to chromatin accessibility and transcription factor co-occupancy. To address these important questions, I first set out to identify and characterise a *bone fide* set of OCT4 binding sites in ESC chromatin. To achieve this, I performed chromatin immunoprecipitation coupled with massively parallel sequencing (ChIP-seq) for OCT4 in the untreated and doxycycline-treated OCT4 conditional cells. Under wild type conditions it was possible to identify 15,920 high-confidence OCT4 binding sites that were lost in doxycycline-treated cells (Figure 6.2.1 C). The majority of these binding events (75%) had histone modifications usually associated with distal regulatory elements, like enhancers (Figure 6.2.1 D), consistent with previous reports that OCT4 binds extensively to distal, non-promoter, regions of the genome (Chen et al., 2008, Göke et al., 2011).

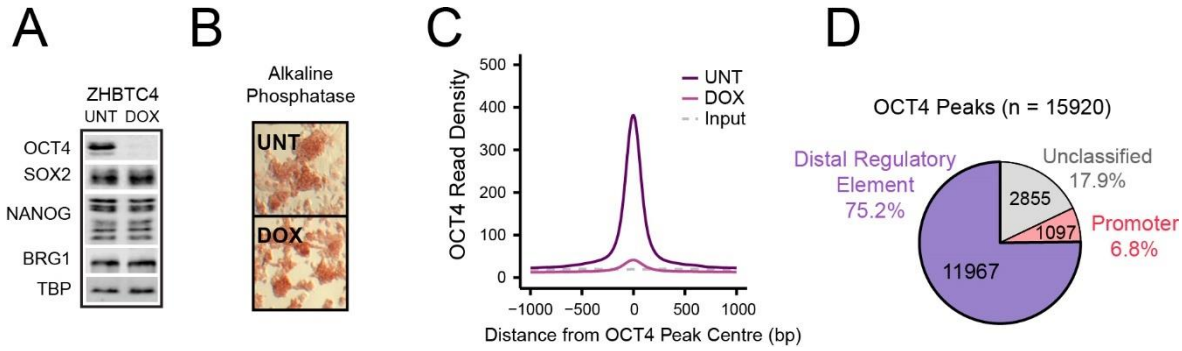


Figure 6.2.1 Deletion of OCT4 using ZHBTC4 ESCs.

A) Western blot of nuclear extracts from ZHBTC4 ESCs grown with or without doxycycline (DOX) for 24 h.

B) ZHBTC4 ESCs show similar levels of alkaline phosphatase activity after 24 h DOX treatment, suggesting they retain a pluripotent, ESC-like state.

C) Average OCT4 ChIP-seq profile over OCT4 peaks (n=15920) reveals near complete loss of OCT4 after 24 h DOX treatment.

D) Annotation of OCT4 peaks as promoters or distal regulatory elements (enhancers) based on H3K4me3 and H3K4me1 histone modifications.

OCT4 is essential for the correct regulation of the pluripotency transcriptional network in mouse ESCs and its function at distal regulatory elements appears to be important for this process (Göke et al., 2011, Jerabek et al., 2014, Chen et al., 2016, Simandi et al., 2016). One indication of its potential function is that during reprogramming, OCT4 acts as a pioneer transcription factor capable of binding inaccessible chromatin (Soufi et al., 2012, Jerabek et al., 2014, Soufi et al., 2015). Therefore, I hypothesized, based on the function of other pioneer factors (Cirillo et al., 2002, Biddie et al., 2011a, Pajoro et al., 2014), that OCT4 might establish local chromatin accessibility at its target sites. To examine this hypothesis, I performed assay for transposase accessible chromatin followed by massively parallel sequencing (ATAC-seq) in the ZHBTC4 cells with and without doxycycline treatment. Intriguingly, examination of several distal OCT4 target sites neighbouring the pluripotency-associated genes *Tbx3* and *Klf2* revealed that loss of OCT4 resulted in substantial reductions in their chromatin accessibility (Figure 6.2.2 A). Comparison of the average ATAC-seq signal at OCT4 peaks in untreated or doxycycline-treated cells confirmed that there were major reductions in the chromatin accessibility of OCT4 target sites genome-wide (Figure 6.2.2 B). Indeed, the majority of OCT4 targets experienced reductions in chromatin accessibility (Figure 6.2.2 C), which was significant for 72 % of OCT4 sites, consistent with OCT4 acting as a pioneer transcription factor to establish chromatin accessibility at these loci. Putative distal regulatory elements bound by OCT4 were particularly affected, and were more likely to exhibit greater reductions in chromatin accessibility than OCT4-bound promoters (Figure 6.2.2 D,E), consistent with a higher number of regulatory inputs at promoters. These observations highlighted a requirement for OCT4 to maintain chromatin accessibility at a large number of distal regulatory elements in mouse ESCs, and were supported by several previous studies (You et al., 2011, Chen et al., 2014b, Shakya et al., 2015, Lu et al., 2016). Importantly, the fact that OCT4 target sites are largely inaccessible in the absence of OCT4 provides convincing evidence that OCT4 does indeed act as a pioneer factor in normal ESCs (as opposed to in over-expression experiments during reprogramming), and that OCT4 is

required to establish and/or maintain normal chromatin structure at the majority of its target sites. It was then important to explore whether these changes in chromatin accessibility related to changes in transcription factor binding and gene expression upon loss of OCT4.

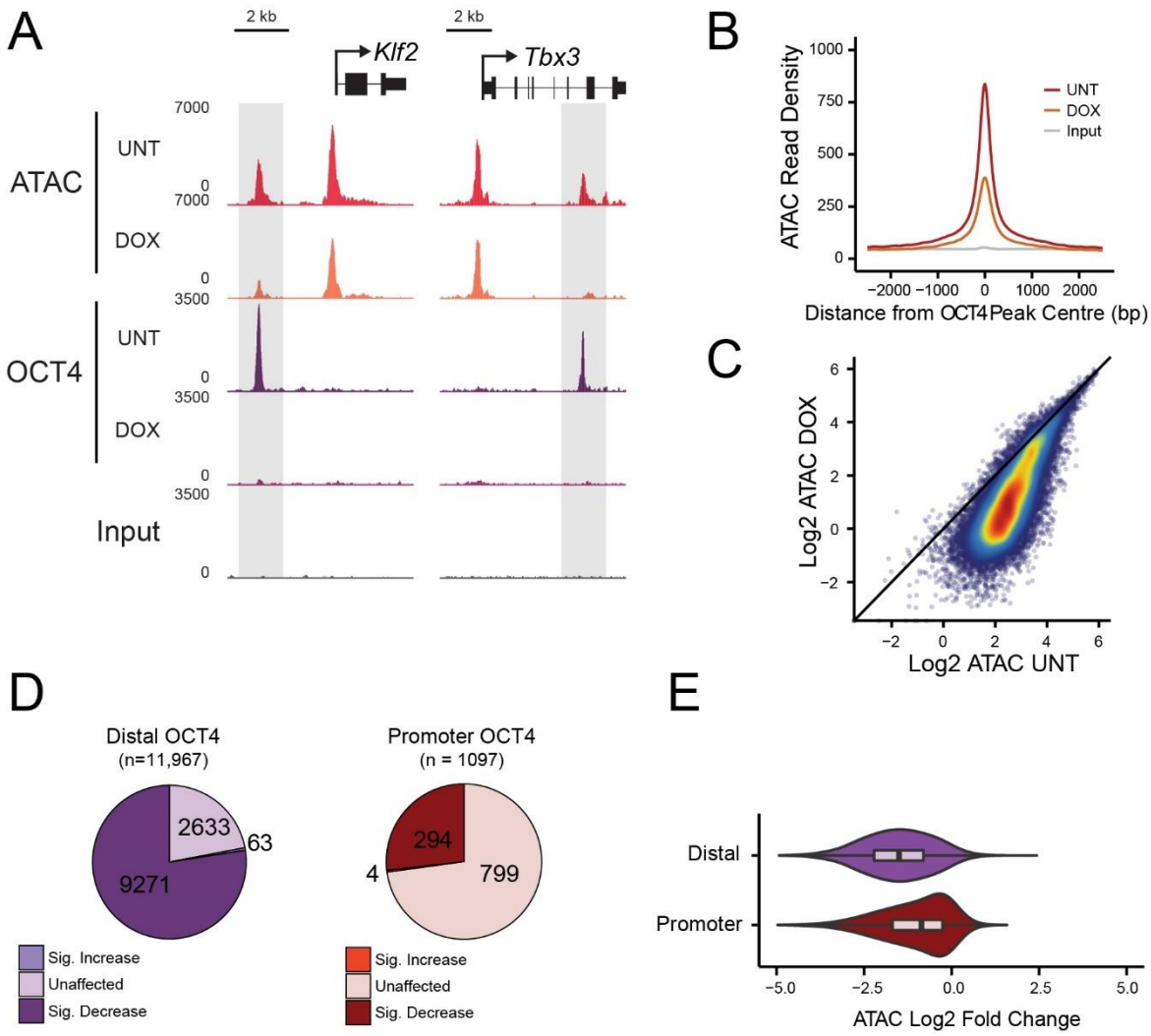


Figure 6.2.2 OCT4 regulates chromatin accessibility at distal regulatory sites.

A) Genomic snapshot of ATAC-seq and OCT4 ChIP-seq in ZHBTC4 ESCs after 24 h doxycycline (DOX) treatment. OCT4-bound regulatory elements are highlighted by grey boxes.

B) Metaplot at OCT4 peaks (n=15,920) reveals reduced ATAC-seq signal genome-wide.

C) Comparison of untreated (UNT) and doxycycline (DOX) ZHBTC4 ATAC-seq RPKM values reveals large number of affected OCT4 target sites.

D) The majority of OCT4-bound distal regulatory elements lose chromatin accessibility while fewer OCT4-bound promoters are affected.

E) Comparison of ATAC log2 fold change after deletion of OCT4 at either OCT4-bound promoters or enhancers.

6.2.1.2 OCT4 supports pluripotency factor binding and gene regulation

Many OCT4 binding sites are co-occupied by SOX2 and NANOG (Chen et al., 2008, Göke et al., 2011), and this combinatorial binding is believed to be important for regulation of the pluripotency transcriptional network in ESCs (Buganim et al., 2013, Smith et al., 2016). In fact, OCT4 and SOX2 act cooperatively to enhance the affinity of each other to their DNA template, and can physically interact to bind a dual OCT4/SOX2 motif sequence (Ng et al., 2012, Aksoy et al., 2013, Tapia et al., 2015). It still however remains poorly understood the extent to which individual pluripotency-associated transcription factors rely on each other for target site binding *in vivo* and how this relates to gene regulation. Given that removal of OCT4 had widespread effects on ATAC-seq signal at its target sites, I wondered whether this might be reflected in changes in the binding of other pluripotency-associated transcription factors. To address this possibility, I examined the binding profiles of SOX2 and NANOG in the presence or absence of OCT4 using ChIP-seq. By ranking OCT4 binding sites by their relative change in chromatin accessibility, it was possible to observe that some OCT4 targets were more significantly affected than others were (Figure 6.2.3 A). These sites appeared to be almost completely inaccessible in the absence of OCT4, consistent with its function as a pioneer factor (Soufi et al., 2012, Soufi et al., 2015). Importantly, the sites with the largest reductions in chromatin accessibility revealed major reductions in SOX2 and NANOG binding (Figure 6.2.3 A). These alterations in SOX2 and NANOG binding correlated extremely well with the loss of ATAC-seq signal (Figure 6.2.3 B,C) suggesting that OCT4 plays a central role in supporting combinatorial binding at these sites, potentially through its regulation of chromatin structure.

To understand if this loss of transcription factor binding was relevant to regulation of the pluripotency transcriptional network, I performed nuclear RNA-seq in wild type and OCT4-depleted cells. Loss of OCT4 resulted in a similar number of genes significantly up-regulated (1616) and down-regulated (1436), with only the down-regulated genes strongly

associated with pluripotency-related pathways (Supplementary Figure 6 A,B), in fitting with OCT4s role as a transcriptional activator (Boyer et al., 2005). When I compared gene expression changes at the genes closest to distal OCT4 targets, there was a very clear trend towards down-regulation of gene expression at the sites most dependent upon OCT4 for chromatin accessibility and transcription factor binding (Figure 6.2.3 D). For example, the distal OCT4 binding site neighbouring the pluripotency-associated *Utf1* gene loses chromatin accessibility, SOX2 and NANOG binding upon deletion of OCT4, which is coincident with a large decrease in *Utf1* RNA levels (Figure 6.2.3 E). In fact, genes neighbouring affected OCT4 binding sites were significantly more likely to be down-regulated after loss of OCT4 than genes neighbouring OCT4 targets that retained chromatin accessibility and transcription factor binding (Figure 6.2.3 F). Furthermore, reductions in the expression of genes near OCT4-bound distal sites was highly coincident (82%) with reductions in chromatin accessibility (Figure 6.2.3 G). Interestingly, however, reductions in chromatin accessibility at an OCT4-bound site did not always lead to alterations in the expression of neighbouring gene, suggesting that some bound sites may not function in gene regulation or that they may have regulatory capacity that extends beyond the nearest gene. Intriguingly, these gene expression defects may reflect an inability to maintain promoter-enhancer interactions in the absence of OCT4 (Levasseur et al., 2008). Together, these observations establish that OCT4 plays an essential role in shaping both chromatin accessibility and cooperative transcription factor binding at its target sites throughout the ESC genome. This pioneer activity appears to be essential for OCT4 to achieve its gene regulatory function of activating the pluripotency transcriptional network.

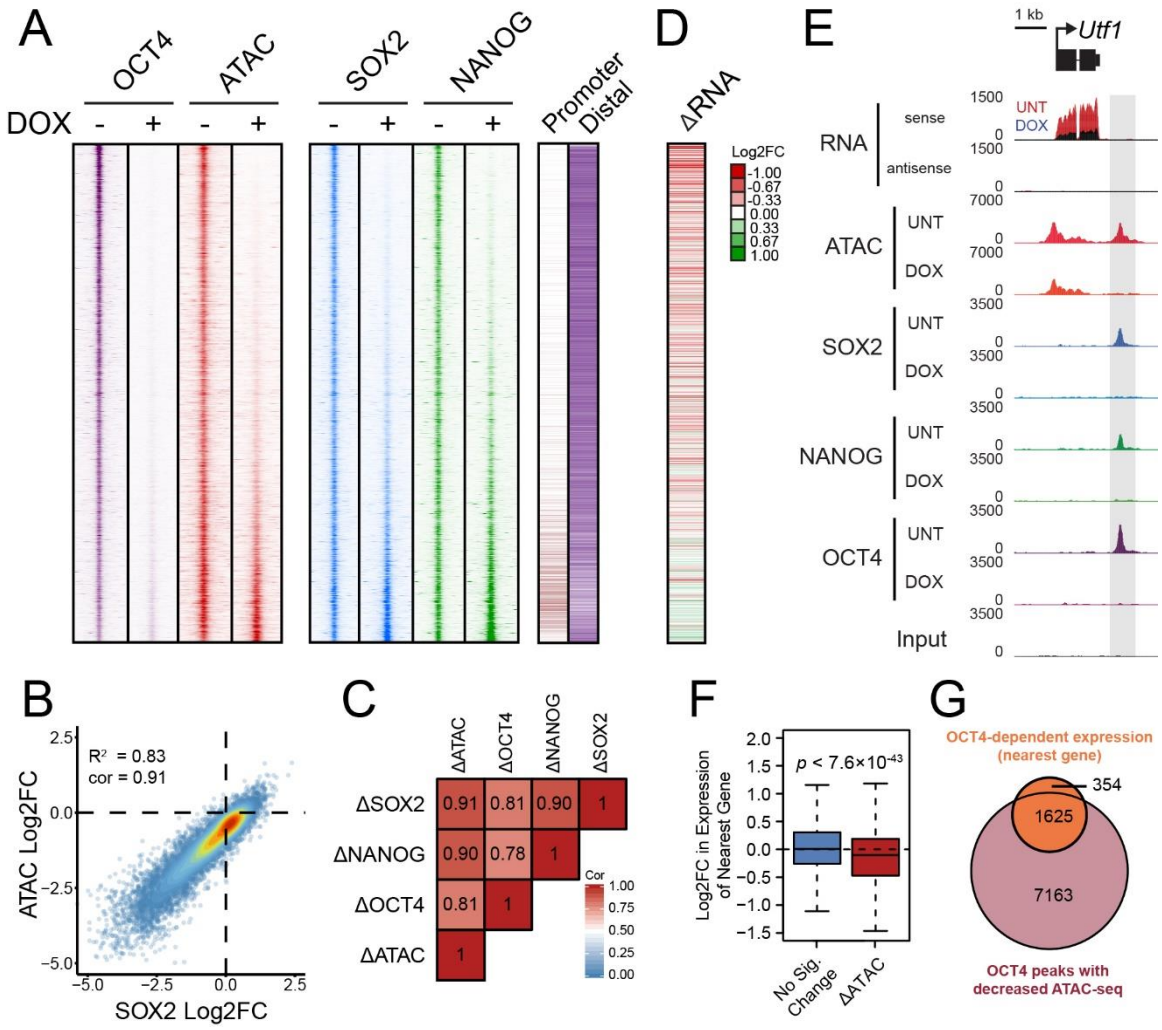


Figure 6.2.3 Loss of OCT4 results in altered pluripotency factor occupancy and down-regulation of neighbouring genes.

A) Heatmap analysis for OCT4 targets ranked by their loss of chromatin accessibility (ATAC) after 24 h doxycycline (DOX) treatment, including their annotation as promoters or distal regulatory elements (right).

B) Scatterplot comparing loss of SOX2 ChIP-seq signal with loss of ATAC-seq signal in ZHBTC4 ESCs at OCT4-SOX2-NANOG co-bound targets (n=5675).

C) Correlation matrix depicting pearson correlation coefficients (Cor) between the log2 fold changes of SOX2, NANOG or OCT4 ChIP-seq with log2 fold changes in ATAC-seq signal at OCT4-SOX2-NANOG co-bound targets (n=5675).

D) Distal OCT4 targets sorted as in (A), depicting log2 fold change in RNA-seq for nearest gene.

E) Genomic snapshot for pluripotency-associated Utf1, which is decreased in expression following loss of OCT4. Distal OCT4 target site is highlighted in grey.

D) Comparison of gene expression changes (nucRNA log2FC) for genes closest to distal OCT4 target sites, with (ΔATAC) or without (No Sig. Change) a significant reduction in chromatin accessibility after deletion of OCT4.

E) A Venn diagram comparing the OCT4 binding sites for which the nearest gene has significantly reduced expression after OCT4 ablation (orange; n=1979) and the OCT4 targets with significant reductions in chromatin accessibility (red; n=8788). Only OCT4 target sites for which the nearest gene has sufficient RNA-seq coverage are included.

6.2.2 OCT4 targets a chromatin remodeller to regulate chromatin accessibility

The ability of OCT4 to regulate chromatin accessibility and transcription factor binding at its target sites, particularly distal regulatory elements, appears to be critical for its ability to regulate transcriptional activity. However, the mechanism by which it might achieve this pioneering activity and subsequent remodelling of chromatin structure is not known. There are several potential modalities by which this could be achieved. For example, some pioneer factors appear to be able to recognise and bind their DNA motifs directly on top of nucleosomes (Espinosa and Emerson, 2001, Cirillo et al., 2002, Soufi et al., 2015, Iwafuchi-Doi et al., 2016), which in some cases can lead to the eviction or repositioning of nucleosomes through steric mechanisms (Cirillo et al., 2002, Iwafuchi-Doi and Zaret, 2014, Voss and Hager, 2014). Alternatively, there are several precedents for the activity of chromatin remodelling enzymes playing a role in this process (Kadam and Emerson, 2003, Marathe et al., 2013, Ceballos-Chávez et al., 2015, Laurette et al., 2015). To date, few transcription factors have been demonstrated to be completely sufficient for chromatin remodelling or nucleosome repositioning (Cirillo et al., 2002), so it seemed most sensible to examine whether the activity of a chromatin-remodelling complex might be responsible for OCT4-dependent chromatin accessibility.

6.2.2.1 OCT4-dependent chromatin accessibility correlates the chromatin remodeller BRG1

Recent studies have identified an increasing number of transcription factors to interact with and target chromatin-remodelling complexes to DNA (Kadam and Emerson, 2003, Ooi et al., 2006, Biddie et al., 2011a, Hu et al., 2011, Marathe et al., 2013, Ceballos-Chávez et al., 2015, Laurette et al., 2015). I therefore sought to examine how different chromatin remodellers might relate to OCT4-dependent chromatin accessibility. Initially, I reasoned that if a chromatin-remodelling complex was responsible or contributed towards

OCT4-dependent accessibility, that remodeller might be enriched at the sites most affected in their ATAC-seq signal after deletion of OCT4. To examine this possibility, I compared the chromatin occupancy of nine different chromatin-remodelling complexes (BRG1, CHD1, CHD2, CHD4, CHD6, CHD8, CHD9, EP400, INO80) in wild type mouse ESCs with the changes in chromatin accessibility observed upon deletion of OCT4 (Figure 6.2.4 A) (Wang et al., 2014, de Dieuleveult et al., 2016). This revealed a negative correlation between levels of BRG1, the catalytic ATPase subunit of the BAF remodelling complex (SWI/SNF), and the changes in ATAC signal after OCT4 loss, indicating that sites with the largest losses of chromatin accessibility were more enriched for BRG1 than sites with OCT4-independent accessibility. Indeed, comparison of OCT4-dependent and OCT4-independent ATAC peaks confirmed a greater level of BRG1 enrichment at OCT4-dependent sites (Figure 6.2.4 B). Furthermore, wild type levels of OCT4 and BRG1 ChIP-seq profiles are highly correlated genome-wide (Figure 6.2.4 C) (Ho et al., 2009, Kidder et al., 2009, de Dieuleveult et al., 2016) and at OCT4 peaks (Figure 6.2.4 D and Figure 6.2.6 B), suggesting that they co-localise throughout the genome. These observations support a previously suggested functional relationship between OCT4 and BRG1 (Buganim et al., 2013, Jerabek et al., 2014), and are strengthened by the identification of BRG1 as an OCT4 interactor by mass spectrometry (Pardo et al., 2010, van den Berg et al., 2010, Ding et al., 2012). Furthermore, BRG1 has been demonstrated to be an important factor for maintaining pluripotency in embryonic stem cells (Ho et al., 2009, Kidder et al., 2009, Singhal et al., 2014, Zhang et al., 2014), as well as improving the efficiency of reprogramming experiments using OCT4 (Singhal et al., 2010). Given these multiple lines of evidence connecting OCT4 and BRG1, it was important to examine how BRG1 might contribute towards the regulation of chromatin structure and transcription factor binding at OCT4 target sites in ESCs.

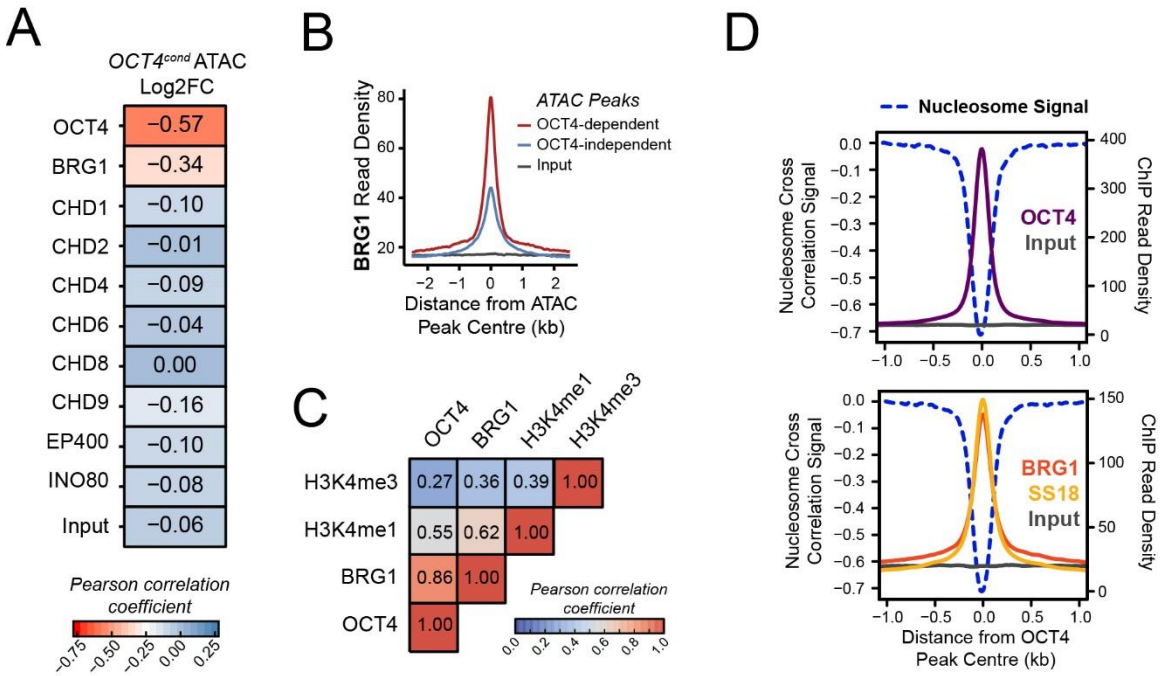


Figure 6.2.4 OCT4-dependent chromatin accessibility is correlated with BRG1.

A) Correlation matrix comparing log2 fold change in ATAC-seq signal (ΔATAC) after deletion of OCT4 in ZHBTC4 ESCs with wild type ChIP-seq for OCT4 and nine chromatin remodellers, at wild type ZHBTC4 ATAC peaks (n=68438).

B) Metaplot of BRG1 ChIP-seq enrichment at OCT4-dependent or OCT4-independent ATAC peaks confirms enrichment for BRG1 at OCT4 targets with significant decreases in chromatin accessibility after deletion of OCT4 (OCT4-dependent).

C) Genome-wide correlation for OCT4, BRG1, H3K4me1 and H3K4me3 in 2 kb windows reveals high degree of co-localization between OCT4 and BRG1.

D) A high resolution metaplot illustrating nucleosome occupancy, OCT4, BRG1 and BRG1-associated factor SS18 at OCT4 peaks (n=15920; 10bp windows) demonstrates that OCT4 and BRG1/BAF signal co-localises to the nucleosome-depleted region at OCT4 peaks.

6.2.3 BRG1 regulates chromatin accessibility at OCT4 target sites

BRG1, like OCT4, is essential for maintaining pluripotency of ESCs (Ho et al., 2009, Kidder et al., 2009, Singhal et al., 2014, Zhang et al., 2014), supports early embryonic development (Bultman et al., 2006), and improves the efficiency of iPSC reprogramming (Singhal et al., 2010). Based on these observations and the fact that BRG1 was specifically enriched at sites that rely on OCT4 for chromatin accessibility (Figure 6.2.4), it seemed likely that BRG1 might be required to shape chromatin accessibility at OCT4 target sites. To examine this important question, I exploited an established mouse

embryonic stem cell line where BRG1 expression can be conditionally ablated upon the addition of tamoxifen (*Brg1*$^{fl/fl}$) (Ho et al., 2009). Tamoxifen treatment of the *Brg1*$^{fl/fl}$ cells resulted in near complete loss of BRG1 protein levels within 72h (Figure 6.2.5 A). Importantly, there were no apparent reductions in the protein levels of OCT4 or SOX2, and *Brg1*$^{fl/fl}$ cells retained alkaline phosphatase staining after 72 h (Figure 6.2.5 B), establishing that the ESC state was largely unaffected by loss of BRG1 at this time point. I therefore performed ATAC-seq in the *Brg1*$^{fl/fl}$ cells with or without tamoxifen treatment (Figure 6.2.5 C). Examination of several distal OCT4 target sites revealed substantial reductions in ATAC-seq signal upon deletion of BRG1 (Figure 6.2.5 C), indicating that BRG1 does contribute to chromatin accessibility at OCT4 target sites. When this analysis was extended to OCT4 target sites genome-wide, it was clear that nearly all OCT4-bound sites experienced a reduction in their chromatin accessibility when BRG1 was deleted (Figure 6.2.5 D), although this reduction only reached my statistical significance threshold for 45% of OCT4 peaks (FDR < 0.05; 1.5-fold reduction). The changes in chromatin accessibility upon deletion of BRG1 were similar to those observed upon deletion of OCT4 (Figure 6.2.5 C,E,F). These results are similar to a requirement for BAF complexes in mediating chromatin accessibility and transcription factor binding in ESCs (Ho et al., 2011, You et al., 2011, Hainer and Fazzio, 2015), but this is the first characterisation of such a functional relationship with OCT4. In order to identify if BRG1s catalytic activity was required for these changes in chromatin accessibility, I attempted to establish a stable over-expression of the K798R catalytic mutant in the *Brg1*$^{fl/fl}$ ESCs. In spite of several attempts, I was unable to isolate such a line (data not shown), likely due to a dominant negative effect described for this mutation (Bultman et al., 2000, Morris et al., 2014). Nevertheless, the observations from the ATAC-seq experiment indicate that BRG1 plays a widespread role in shaping chromatin accessibility at OCT4 target sites in ESCs and suggests that the pioneering activity of OCT4 to remodel chromatin relies upon BRG1.

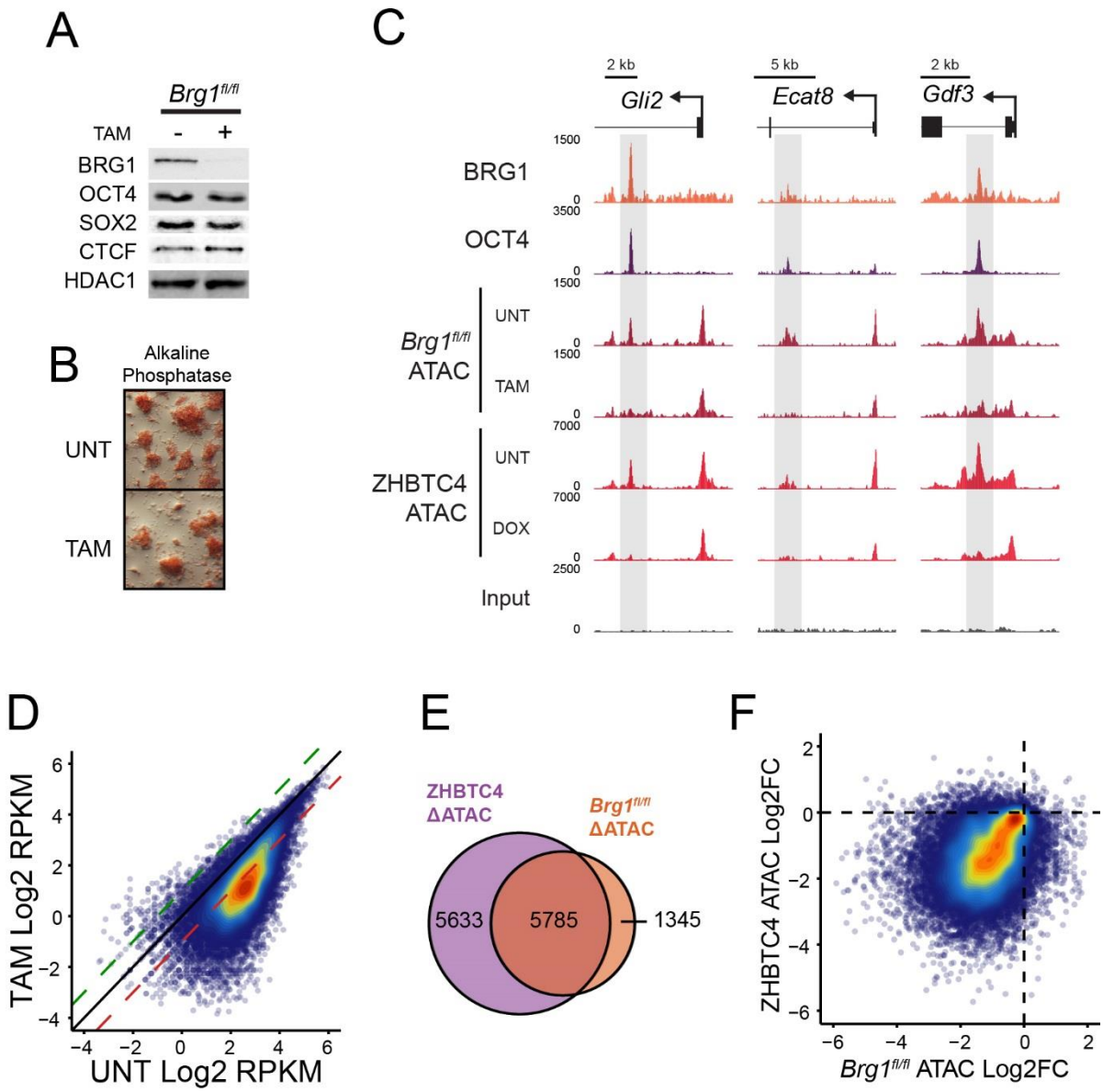


Figure 6.2.5 BRG1 regulates chromatin accessibility at OCT4 target sites.

A) Conditional deletion of BRG1 using *Brg1*$^{fl/fl}$ mouse ESCs. Treatment of cells with tamoxifen (TAM) for 72 h confirms deletion of BRG1 by western blotting.

B) Alkaline phosphatase staining of *Brg1*$^{fl/fl}$ ESCs after 72 h tamoxifen treatment shows comparable alkaline phosphatase activity, indicative of retained ESC state.

C) Genomic snapshot of distal OCT4 target sites near pluripotency-associated *Gli2*, *Ecat8*, and *Gdf3*. Deletion of BRG1 by tamoxifen treatment of *Brg1*$^{fl/fl}$ results in reduced chromatin accessibility, as measured by ATAC-seq. This is comparable with the reductions in chromatin accessibility that occur upon deletion of OCT4 (ZHBTC4) after treatment with doxycycline (DOX).

D) Comparison of untreated (UNT) and tamoxifen-treated (TAM) *Brg1*$^{fl/fl}$ ESCs ATAC-seq RPKM values reveals large number of affected OCT4 target sites.

E) Venn diagram overlap of OCT4 targets which significantly lose ATAC-seq signal (FDR < 0.05; 1.5-fold reduction) upon deletion of OCT4 (ZHBTC4) or BRG1 (*Brg1*$^{fl/fl}$).

F) Scatterplot analysis of changes in ZHBTC4 and *Brg1*$^{fl/fl}$ ATAC-seq at OCT4 target sites.

6.2.4 OCT4 recruits BRG1 to chromatin

Biochemical studies have identified that OCT4 and BRG1 physically interact (Pardo et al., 2010, van den Berg et al., 2010, Ding et al., 2012), and there is a high degree of co-localisation throughout the ESC genome between these two factors (Figure 6.2.4 and Figure 6.2.6B) (Ho et al., 2009, Kidder et al., 2009, de Dieuleveult et al., 2016). It therefore seemed possible that OCT4 might directly recruit BRG1 to OCT4 target sites throughout the genome. To address this, I treated ZHBTC4 mouse ESCs with doxycycline to ablate OCT4 expression, and performed BRG1 ChIP-seq after 24 hours. Although total nuclear levels of BRG1 were unaffected after doxycycline treatment (Figure 6.2.1 A), examination of several OCT4 binding sites revealed large reductions in BRG1 chromatin occupancy upon loss of OCT4 (Figure 6.2.6 A), with sites lacking OCT4 retaining BRG1 chromatin occupancy (Figure 6.2.6 A; Supplementary Figure 6). Visualisation of BRG1 ChIP-seq at OCT4 binding sites before and after doxycycline treatment revealed substantial reductions in BRG1 levels across all OCT4 peaks (Figure 6.2.6 B), while regulatory elements without OCT4 retained BRG1 (Figure 6.2.6 C). This indicates that OCT4 is required for BRG1 chromatin occupancy, and suggests that OCT4 recruits BRG1 to chromatin at these loci, presumably to remodel chromatin. Indeed, loss of BRG1 chromatin occupancy was highly correlated with loss of ATAC-seq signal in the absence of OCT4 (Figure 6.2.6 D), supporting the conclusion that OCT4 regulates chromatin accessibility at its target sites through recruitment of BRG1.

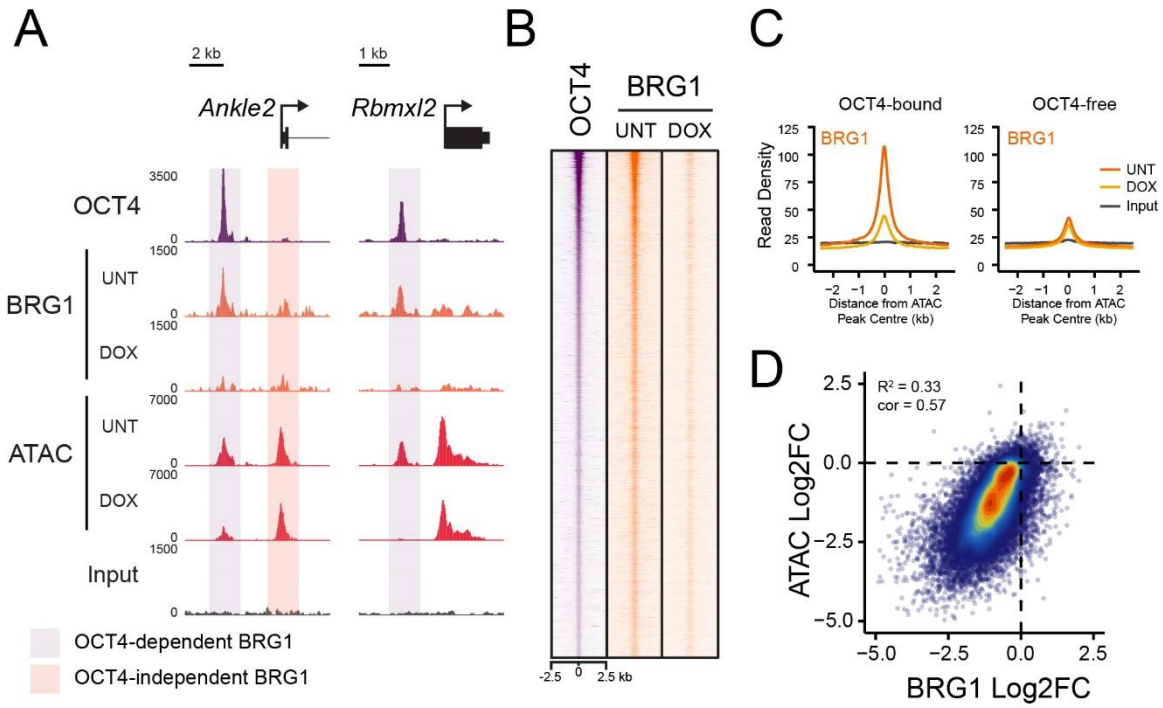


Figure 6.2.6 OCT4 recruits BRG1 to chromatin.

A) Genomic screenshot of BRG1 ChIP-seq in ZHBTC4 ESCs in the presence (UNT) or absence (DOX) of OCT4. Purple and orange rectangles highlight OCT4-dependent BRG1 and OCT4-independent BRG1 respectively.

B) Heatmap analysis of OCT4 peaks (n = 15920) confirms genome-wide loss of BRG1 enrichment upon deletion of OCT4.

C) Metaplot of ZHBTC4 BRG1 ChIP-seq at OCT4-bound ATAC peaks (n = 11666) or ATAC peaks lacking OCT4 (OCT4-free; n = 56494).

D) Scatterplot demonstrating high degree of correlation between reductions in BRG1 chromatin occupancy and chromatin accessibility (ATAC-seq) at OCT4 peaks.

6.2.5 BRG1 is required to mature transcription factor binding

The discovery that OCT4 recruits BRG1 to regulate chromatin accessibility at OCT4 target sites in mouse ESCs revealed an interesting mechanism by which a sequence-specific transcription factor can influence chromatin architecture at its binding sites. This has important implications for understanding gene regulation by transcription factors (Spitz and Furlong, 2012), and is particularly relevant to understanding the function of master developmental transcription factors such as OCT4 (Iwafuchi-Doi and Zaret, 2014). However, what remained unclear was to what extent OCT4 required BRG1 to occupy and

bind DNA. For example, BRG1 might be required to remodel chromatin at OCT4 motif sites to destabilise nucleosome:DNA contacts, as previously described (Côté et al., 1998, Narlikar et al., 2001), which could prevent OCT4 and its co-factors from binding their motif sequences. Alternatively, OCT4 might be able to recognise and bind its target sequences without requiring nucleosome re-positioning, as suggested recently (Soufi et al., 2015). This alternative model is more in fitting with the current view of pioneer transcription factors, which have been invoked to be sufficient to access their DNA motifs and to establish new regulatory elements in inaccessible chromatin in the absence of chromatin remodelling activities (Magnani et al., 2011, Iwafuchi-Doi and Zaret, 2014). Therefore, understanding whether OCT4's ability to bind DNA requires chromatin remodelling by BRG1 is essential to appreciate its function as a pioneering transcription factor.

6.2.5.1 BRG1 is required for OCT4/SOX2 DNA binding and gene regulation

The OCT4-dependent recruitment of BRG1 to OCT4 target sites suggested that BRG1 might be important for regulating binding and/or function of pluripotency transcription factors. To address whether this was the case, I performed ChIP-seq for OCT4 and SOX2 in the *Brg1*$^{fl/fl}$ ESCs, with or without tamoxifen treatment. Importantly, total nuclear levels of OCT4 and SOX2 remained unchanged after 72 hours of tamoxifen treatment of the *Brg1*$^{fl/fl}$ ESCs (Figure 6.2.5). I chose to focus my subsequent analysis on distal OCT4 targets, as these were the sites most dependent upon OCT4 for their chromatin accessibility and function (Figure 6.2.2), and I wanted to avoid any confounding effects of promoter-dependent regulatory features. Visual examination of distal OCT4 targets revealed two general modes of OCT4 and SOX2 binding in the absence of BRG1 (Figure 6.2.7). Firstly, the majority of OCT4 target sites showed dramatic reductions in OCT4 and SOX2 ChIP-seq signal after deletion of BRG1 (Figure 6.2.7 A), consistent with a requirement for BRG1 to mediate their binding at these loci. In contrast, a subset of sites

retained normal transcription factor binding in the absence of BRG1 (Figure 6.2.7 B). Genome-wide there was a general trend towards decreased binding (Figure 6.2.7 C), and approximately 60% of distal OCT4 targets experienced significant reductions in OCT4 occupancy after deletion of BRG1. SOX2 binding was also significantly reduced following BRG1 removal, although the total number of sites in which SOX2 binding was significantly affected was less dramatic (18.5%; 1276/6908 of OCT4-SOX2 co-bound targets). The majority of target sites did not completely lose OCT4 or SOX2 DNA binding, suggesting that a small amount of transcription factor binding was able to occur even in the absence of BRG1 (Figure 6.2.7 D). This indicates that although OCT4 and SOX2 are still able to recognise and partially bind their sequence motifs in the absence of BRG1, they require the OCT4-dependent recruitment of BRG1 to remodel chromatin and establish normal regulatory elements.

Having demonstrated that OCT4 and SOX2 binding was largely dependent upon BRG1, it was then important to examine how this related to OCT4-dependent regulation of the pluripotency transcriptional network. Somewhat surprisingly, previous work had not supported a direct relationship between the loss of BRG1 and the reduced activity of OCT4 target genes (Ho et al., 2009, Ho et al., 2011, Zhang et al., 2014, Hainer et al., 2015). However, it was apparent that following loss of BRG1, OCT4 and SOX2 binding was variably affected (Figure 6.2.7 A-C). Therefore, it seemed possible that transcriptional changes following loss of BRG1 might only be evident at sites that lose normal OCT4 and SOX2 binding. To examine this possibility I carried out nuclear RNA-seq in wild type and BRG1-depleted cells. As described earlier, loss of OCT4 expression resulted in the down-regulation of many genes in close proximity to OCT4 binding sites (Figure 6.2.7 E, lower panel), in fitting with OCT4 functioning to activate transcription at its target genes (Boyer et al., 2005). When I examined genes associated with reduced OCT4 and SOX2 binding in the absence of BRG1, it was clear that their expression was substantially reduced (Figure 6.2.7 E, upper panel). In contrast, sites that maintained OCT4 and SOX2 binding

in the absence of BRG1 showed no changes or increases in expression (Figure 6.2.7 E), potentially driven through increased SOX2 binding at these sites (Figure 6.2.7 D). This supports a model whereby BRG1 is required for pluripotency-associated transcription factor binding at distal regulatory elements to activate transcriptional activity of nearby genes, and likely reflects a general mechanism by which chromatin remodelling complexes influence transcription by modulating transcription factor binding. Ultimately, these observations are consistent with a pioneering activity of OCT4 and SOX2 to initially engage with inaccessible chromatin, but that OCT4-dependent BRG1 recruitment and remodelling is subsequently required to mature these inaccessible sites into functional distal regulatory elements to regulate the pluripotency transcriptional network.

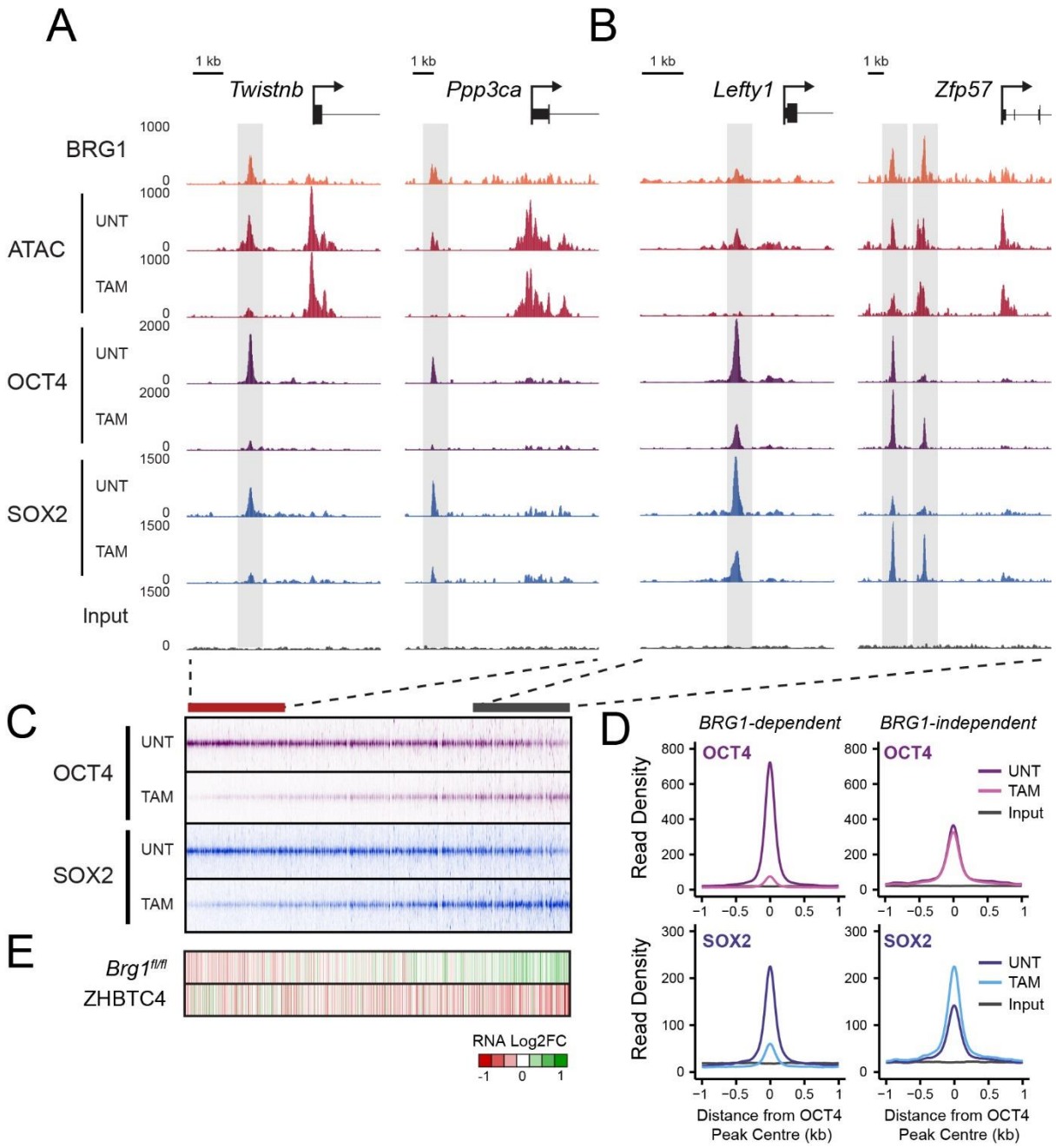


Figure 6.2.7 OCT4 and SOX2 require BRG1 to bind DNA and regulate gene expression.

A) Genomic snapshots depicting *Brg1*$^{fl/fl}$ ATAC-seq and OCT4 and SOX2 ChIP-seq after tamoxifen (TAM) treatment. Two distal OCT4 targets which significantly lose OCT4 and SOX2 are illustrated.

B) Same as (A), but for two distal OCT4 targets which maintain OCT4 and SOX2 binding.

C) Heatmap analysis of distal OCT4 targets ranked by their relative change in OCT4 ChIP-seq after deletion of BRG1. OCT4 and SOX2 ChIP-seq are visualised, and broadly demonstrate a trend of BRG1-dependent and BRG1-independent OCT4 and SOX2 binding.

D) Metaplot profiles at BRG1-dependent or BRG1-independent distal OCT4 targets.

E) Heatmap of log2 fold change of *Brg1*$^{fl/fl}$ and ZHBTC4 RNA-seq for genes nearest to distal OCT4 targets, ranked by their relative change in OCT4 ChIP-seq following deletion of BRG1.

6.2.6 Distinct modes of OCT4 binding and function during reprogramming and development

Through studying OCT4 in iPSC reprogramming it has been proposed that there may be distinct phases of OCT4 binding which are influenced by pre-existing chromatin states in somatic cells (Koche et al., 2011, Soufi et al., 2012, Buganim et al., 2013, Chen et al., 2016). For example, OCT4 initially occupies a subset of sites modified by histone H3 lysine 4 dimethylation or histone H3 lysine 27 acetylation, which are thought to be associated with more accessible chromatin during the early stages of reprogramming. This is followed by OCT4 binding at sites lacking pre-existing chromatin modifications that are thought to become accessible during the later stages when pluripotency is being established. This suggests that different OCT4 binding modalities may be exploited to support regulatory element function during reprogramming and development. In light of the observation in ESCs that OCT4 binding events can be separated into those that are BRG1-dependent and BRG1-independent, I wanted to examine how these distinct types of binding were related to OCT4-associated gene expression and binding events during cellular reprogramming and early development. To achieve this I first analysed gene expression during the reprogramming of mouse fibroblasts into iPSCs via OCT4/SOX2/KLF4/MYC (Cieply et al., 2016) and asked when genes associated with BRG1-independent and BRG1-dependent OCT4 binding at distal regulatory elements were expressed. Intriguingly, genes associated with BRG1-independent OCT4 binding were activated early during reprogramming and BRG1-dependent sites later (Figure 6.2.8 A). This suggests that OCT4 expression in somatic cells leads to a more rapid transcriptional response from genes associated with BRG1-independent OCT4 binding sites. In contrast, genes that require BRG1 for OCT4 binding were expressed later during reprogramming, perhaps because OCT4 requires a more substantial remodelling of chromatin to establish transcription factor binding and chromatin structure at these sites. In fitting with this possibility, BRG1-dependent OCT4 binding sites were associated with

genes expressed at later developmental stages during early mouse development than genes associated with BRG1-independent OCT4 binding (Figure 6.2.8 C), suggesting that these two modalities might exist in the developing embryo and are not only observed in cell culture systems. To examine this possibility, I considered the activation of BRG1-independent and BRG1-dependent OCT4 target sites in the early mouse embryo using chromatin accessibility as a proxy for OCT4 binding and regulatory activity (Wu et al., 2016). Surprisingly, this confirmed that BRG1-dependent OCT4 targets become accessible at later stages of embryonic development (Figure 6.2.8 D). Together these observations suggest two distinct forms of OCT4 engagement with the genome during reprogramming and development. The first relies on binding to sites that are already accessible and display more rapid gene expression kinetics. The second mode of binding appears to rely on the pioneering activity of OCT4 to engage with previously inaccessible chromatin, whereby the remodelling activity of BRG1 is required to facilitate additional transcription factor binding and to functionally mature these regulatory elements to support pluripotency gene expression later in development or reprogramming. Importantly, these observations suggest that pre-existing chromatin environments dictate different mechanisms that transcription factors must employ to read DNA-encoded information to regulate gene expression.

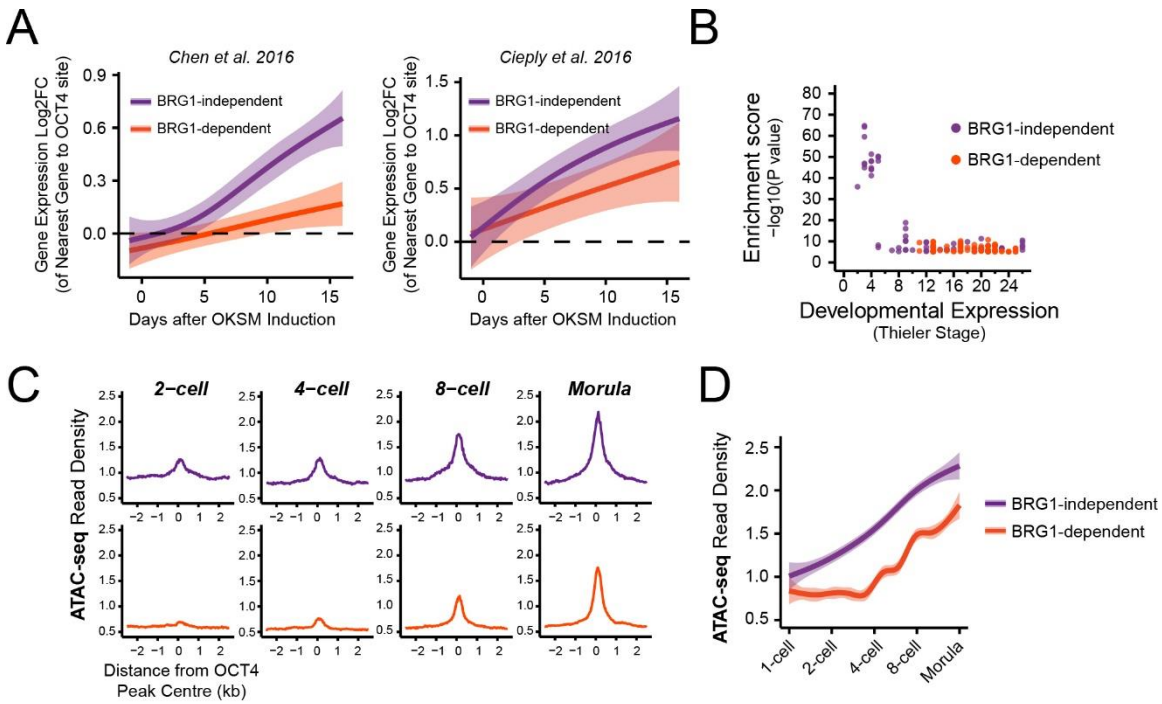


Figure 6.2.8 OCT4s requirement for BRG1 chromatin remodelling reveals distinct timing of regulatory events during reprogramming and development.

A) A time course of gene expression changes (log2 fold change) during OKSM-mediated reprogramming of mouse embryonic fibroblasts to iPSCs (Chen et al., 2016, Cieply et al., 2016). The expression changes of OCT4-dependent genes neighbouring the 20% of distal OCT4 targets that were most dependent upon BRG1 for normal OCT4 binding (BRG1-dependent) or the 20% of distal OCT4 targets that were least dependent (BRG1-independent) following BRG1 removal were quantified and visualised as a smoothed trendline ± 95% confidence interval.

B) Genomic Regions Enrichment of Annotations Tool (GREAT) for BRG1-dependent and BRG1-independent distal OCT4 targets. Plotted are the enrichment (-log10(P Value)) against the MGI Expression profile, or development stage, for genes associated with distal OCT4 targets.

C) Metaplot profiles of ATAC-seq signal during early mouse embryonic development (Wu et al., 2016) at BRG1-independent (upper) and BRG1-dependent (lower) distal OCT4 target sites. BRG1-independent sites gain accessibility earlier than BRG1-dependent sites.

D) A quantitation of ATAC-seq read density at distal OCT4 targets depicted in (C) during early embryonic development plotted as a smoothed trendline ± 95% confidence interval.

6.3 Discussion

Tissue-specific transcription factors are required to recognise and bind DNA motif sequences in order to regulate gene expression during important developmental transitions. Their ability to stably interact with DNA in the context of chromatin, and to displace or compete with nucleosomes to establish this DNA binding and residency is essential to this process, but remains very poorly understood. Here I have described evidence that the master pluripotency factor OCT4 binds previously inaccessible chromatin and then acts as a pioneer to alter chromatin structure at its target sites and facilitate binding of other pluripotency factors. It achieves this by recruiting the BAF chromatin remodelling complex, and deletion of the BAF catalytic core BRG1 phenocopies the loss of OCT4-dependent chromatin accessibility. Furthermore, there is a clear requirement for BRG1 in allowing OCT4, and the additional pluripotency factor SOX2, to establish normal binding at distal regulatory elements in ESCs, and this perturbation of transcription factor binding has variable, but destabilising effects, on the pluripotency transcriptional network. Finally, OCT4 targets that are dependent on BRG1 for their chromatin occupancy are established later in reprogramming and development, compared to BRG1-independent OCT4 targets which are able to access their DNA motifs independent of BRG1-facilitated remodelling events. This has broad implications for understanding how OCT4 regulates the pluripotency transcriptional network in ESCs, but also how pioneer transcription factors exploit chromatin remodelling activities to access DNA-encoded information in the genome.

6.3.1 Cooperation between OCT4 and BRG1/BAF complexes

For many years OCT4 has been known as one the master regulators of embryonic development and stem cell pluripotency (Jerabek et al., 2014, Wu and Schöler, 2014), and although its interfaces with DNA motif sequences have been elucidated (Klemm et al., 1994, Reményi et al., 2001, Esch et al., 2013), the mechanism by which it establishes

DNA binding throughout the genome, particularly in the context of chromatin, has not been well-described in living cells. Establishing that OCT4 is required to maintain chromatin accessibility in normally inaccessible regions of chromatin is an important observation towards confirming that OCT4 acts as a pioneer transcription factor in the normal pluripotent state. While regulation of chromatin structure or nucleosome positioning by OCT4 had been observed before at individual loci (You et al., 2011, Shakya et al., 2015) or genome-wide (Chen et al., 2014b), the effects were rather modest. In contrast, I observed an essential role for OCT4 in determining the chromatin accessibility at many thousands of its binding sites, and that this important activity results from the recruitment of the BAF remodelling complex by OCT4. BRG1 and the embryonic BAF complex have been identified in multiple proteomic analyses of OCT4 (Pardo et al., 2010, van den Berg et al., 2010, Ding et al., 2012, Esch et al., 2013), and several groups have proposed models whereby OCT4 might directly recruit BRG1 to chromatin (Buganim et al., 2013, Jerabek et al., 2014, Singhal et al., 2014, Smith et al., 2016). Indeed, BRG1, and its cofactor BAF155, enhance OKSM-mediated reprogramming and facilitate greater chromatin occupancy of exogenous OCT4 (Singhal et al., 2010). Another study showed only a modest contribution of BAF155 towards nucleosome positioning, as well as reduced BAF155 occupancy at two OCT4 targets after OCT4 deletion (You et al., 2011), although this was likely due to reduced BAF155 protein levels. Importantly, by measuring BRG1s contribution towards chromatin remodelling at OCT4 target sites using ATAC-seq, combined with establishing the direct recruitment of BRG1 by OCT4 genome-wide, I have been able to link these disparate observations to form a more mechanistic understanding of OCT4 binding and function at its target sites throughout the ESC genome. Although the precise mechanisms by which BRG1 promotes OCT4 and SOX2 binding remains to be explored, previous work has proposed that BRG1/BAF complexes may function not through nucleosome displacement and eviction, but rather through dynamic transitioning between nucleosomal states with different accessibility features, potentially through weakening histone:DNA contacts (Côté et al., 1998, Narlikar et al., 2001, Clapier and

Cairns, 2009). This would be consistent with BRG1-dependent remodelling enhancing the binding of OCT4/SOX2 to motif sequences positioned on top of nucleosomes (Ishii et al., 2015, Soufi et al., 2015).Furthermore, this mechanistic detail may reflect an important observation regarding the interpretation of differential ATAC-seq experiments described here, suggesting that altered chromatin accessibility may not simply be a measure of nucleosome displacement or depletion.

6.3.2 Tissue-specific chromatin remodelling for sequence-specific regulation

The functional interaction between BRG1/BAF and OCT4 provides us with an important insight into how transcription factors exploit chromatin remodelling machines to bind their DNA sequences and effect gene expression, and this may be particularly relevant for tissue-specific gene regulation. For example, BAF remodelling complexes exist in various tissue-specific forms (Wang et al., 1996, Olave et al., 2002, Lickert et al., 2004, Wu et al., 2007, Ho et al., 2009), and this appears to be important for their ability to interact with a wide variety of tissue-specific transcription factors (Ooi et al., 2006, Hu et al., 2011, Weider et al., 2012, Ballaré et al., 2013, Marathe et al., 2013, Bao et al., 2015, Hainer and Fazzio, 2015, Laurette et al., 2015). The full extent of tissue specificity for BAF components remains to be seen (Wu, 2012, Kadoch and Crabtree, 2015), but the startling diversity of potential combinations reveals a flexibility for the recruitment of BAF complexes, and other chromatin remodellers (Morris et al., 2014, Wang et al., 2014, Ceballos-Chávez et al., 2015, Laurette et al., 2015), to shape transcription factor binding and gene expression across varied cellular states during development. BRG1/BAF appears to be essential for establishing OCT4 DNA binding with other pluripotency-associated transcription factors in ESCs. Although reminiscent of studies which suggest that a single pioneer factor establishes chromatin accessibility to facilitate binding of additional factors (Biddie et al., 2011a, Hu et al., 2011, Schulz et al., 2015, Iwafuchi-Doi et al., 2016), it appears as if OCT4-dependent remodelling by BRG1 might instead act to stabilise or reinforce nascent

OCT4-SOX2 binding events in inaccessible chromatin, consistent with data suggesting that SOX2 binding often precedes that of OCT4 (Chen et al., 2014a). This “maturation” of transcription factor binding appears to be a significant contribution of BRG1 towards normal expression of the pluripotency transcription network (Ooi et al., 2006, Ho et al., 2011, Hainer and Fazzio, 2015).

6.3.3 Models for maturation of transcription factor binding by BRG1

How then do transcription factors and chromatin remodellers cooperate together to establish pioneer transcription factor binding? One possibility is an “assisted search” model, whereby a transcription factor-chromatin remodeller complex engages in transient, non-specific “on-off” interactions with chromatin. During these transient interactions, the chromatin remodeller might reposition or evict nucleosomes non-specifically to facilitate easier access to the underlying DNA sequence for the transcription factor. However, live-cell imaging of remodellers in human cells suggested that the majority of interactions with the genome are not involved with active remodelling activity (Erdel et al., 2010, Erdel and Rippe, 2012). This suggests that the majority of chromatin remodelling activity might therefore be limited to transcription factor targets by the initial recognition and motif binding by the sequence-specific factor. The subsequent increase in residency time for the chromatin remodeller, combined with potential conformational changes of the transcription factor or recognition of epigenetic modifications, might then allow the remodeller to remodel the local chromatin environment, potentially unmasking co-factor sites, improving transcription factor binding affinity and establishing an active regulatory element. Such a stochastic process might explain the delay in activation of BRG1-dependent compared to BRG1–independent OCT4 targets (Figure 6.2.8). For example, at BRG1-independent sites OCT4 gains easier access to the motif sequences and so is able to establish binding sooner than BRG1-dependent sites, which require a more stochastic-driven process to resolve OCT4 DNA binding. Importantly, the specificity of

OCT4 target sites depends upon the presence of cooperative binding partners such as SOX2 or OTX, and changes in expression of these factors during development results in changes in OCT4 binding profiles (Jauch et al., 2011, Aksoy et al., 2013, Yang et al., 2014a). This establishes that co-factor binding is likely to play a role in promoting OCT4s pioneering activity in inaccessible chromatin, probably by increasing OCT4 DNA-binding affinity and residency time. While OCT4 appears to rely in many instances on chromatin remodelling to function as a pioneer factor, it remains unclear how many other putative pioneer factors rely upon similar interactions, either directly or indirectly, with chromatin remodellers.

6.3.4 Conclusions and future directions

While the precise sequence of events for pioneer-mediated chromatin remodelling by OCT4/BRG1 remains to be elucidated, these observations provide convincing evidence that transcription factors are capable of modulating chromatin structure and accessibility to facilitate their binding to DNA sequences. Although this work does not directly address the CGI chromatin accessibility discussed throughout this thesis, given OCT4s binding to predominantly distal regulatory elements, it does highlight a general principle by which cells can regulate chromatin structure. This may have important implications for understanding CGI chromatin accessibility. For example, ubiquitously expressed transcription factors may have similar interactions with chromatin remodellers and could make important contributions towards CGI chromatin accessibility across different cell or tissue types. It will be interesting to explore how common such cooperation between sequence-specific factors and chromatin-modifying machinery might be, in order to gain a better appreciation of how cells coordinate chromatin structure, transcription factor binding, and ultimately, the regulation of gene expression.

CHAPTER 7 - CONCLUSIONS AND FUTURE DIRECTIONS

Chromatin accessibility is a characteristic feature of regulatory elements in eukaryotic genomes. While both active promoters and distal regulatory elements may be accessible to *trans*-acting factors in a given cell type, CGI promoters tend to be accessible across a wide range of cell types independent of transcriptional activity (Chapter 4). It has previously been proposed that this may be an inherent feature of CGI sequences (Ramirez-Carrozzi et al., 2009), and I have attempted to address this question through the genetic ablation of DNA methyltransferases and recombination experiments using CGI-like sequences (Chapter 4). However, although preliminary results from these recombination experiments were suggestive, I was unable to sufficiently interrogate enough CGI-like sequences, or to compare methylated with non-methylated sequences, due to experimental difficulties. Nevertheless, by examining several ZF-CxxC domain-containing proteins, I was able to identify that KDM2A/B promote chromatin accessibility at CGIs genome-wide in ESCs (Chapter 5), potentially representing a generic mechanism by which non-methylated, CpG-rich DNA sequences could achieve an elevated chromatin accessibility. In contrast, the subset of CGI promoters occupied by polycomb repressive complexes have a more compact chromatin structure, which suggests that polycomb-associated features may exist to antagonise normal CGI accessibility (Chapter 5), although I was unable to identify the factors responsible for this. As an alternative approach to understanding chromatin accessibility in ESCs, I also considered the mechanism by which the pioneer factor OCT4 establishes open chromatin at its target sites (Chapter 6). Intriguingly it achieves this through direct cooperation with a chromatin remodelling complex, and this is essential for normal OCT4/SOX2 binding (Chapter 6). Ultimately, these experiments have revealed a multi-layered picture of how chromatin accessibility can be determined at gene regulatory elements.

The complexity of chromatin-associated pathways at gene regulatory elements, particularly CGIs, makes it challenging to understand the direct mechanisms by which chromatin accessibility is regulated *in vivo*. For example, chromatin remodellers, transcription factors, histone variants, histone acetylation, DNA structure, and chromatin-modifying enzymes have all been implicated in promoting chromatin accessibility (Anderson and Widom, 2001, Segal and Widom, 2009, Grøntved et al., 2013, Hu et al., 2013, Oldfield et al., 2014, Schulz et al., 2015, Siggens et al., 2015, Frank et al., 2016, Li et al., 2016). However, due to the interplay and feedback mechanisms between many of these different features in many cases it is challenging to infer the precise molecular mechanisms underlying changes in chromatin structure resulting from the perturbation of a single feature. For example, increases in accessibility following inhibition of HDAC activity could result from direct disruption of histone:DNA interactions by increased acetylation (Shogren-Knaak et al., 2006), or through the increased recruitment of chromatin remodelling complexes (Marmorstein and Berger, 2001, Hassan et al., 2002). Similarly, the absence of DNA methylation at CGIs could serve to promote accessibility by the recruitment of ZF-CxxC domain-containing proteins (Chapter 5), increased transcription factor binding (Palacios et al., 2010, Medvedeva et al., 2014), the absence of methylation-dependent repressive pathways (Ng et al., 1999, Yoon et al., 2003, Du et al., 2015, Ludwig et al., 2016) or altered nucleosomal DNA properties (Lee and Lee, 2012, Collings et al., 2013). Alteration of a single one of these features could directly lead to changes in another, and in many cases this might be difficult to predict. It is therefore important that future work should consider chromatin structure in a more holistic manner, and try to integrate our knowledge of interdependency between chromatin-associated factors and chromatin itself.

While the defined causes of CGI chromatin accessibility may be confounded by the complexity of these regulatory platforms, the consequences of this chromatin accessibility are even less clear. Chromatin accessibility has traditionally been viewed as an important

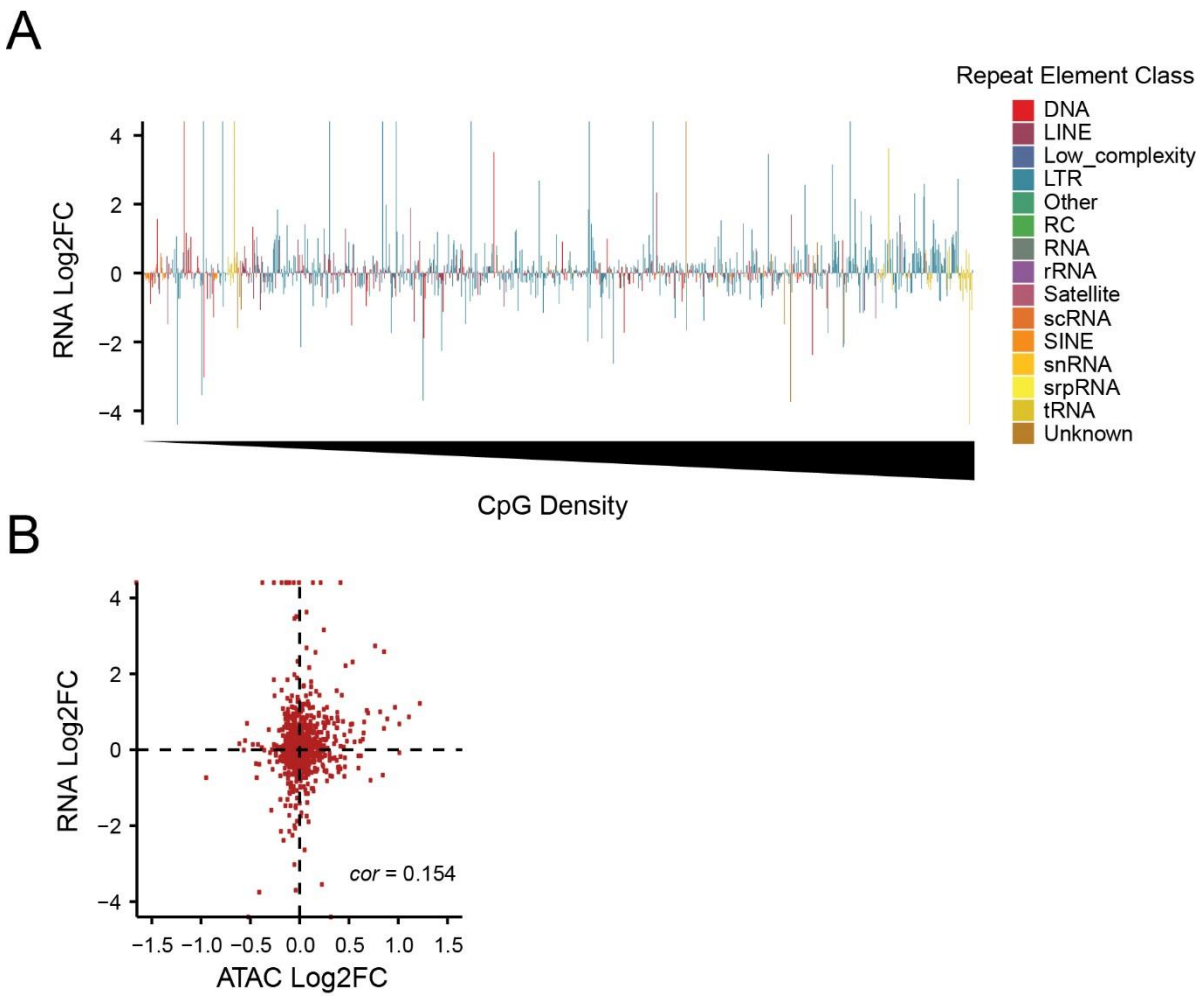
feature of regulatory elements to allow the binding of non-pioneer transcription factors. Typically, this has been ascribed to active processes driven by ATP-dependent remodellers which evict or reposition nucleosomes to uncover DNA motif sequences. While this can clearly contribute towards the overall chromatin accessibility of CGIs and other regulatory elements (Chapter 6, Morris et al., 2014, de Dieuleveult et al., 2016), I propose an alternative model through which chromatin accessibility can also be shaped by the dynamic interactions of transcription factors and chromatin-modifying complexes with nucleosomal DNA at regulatory elements. This is in fitting with the surprising observation that loss of PcG complexes can result in decreases in the already limited chromatin accessibility of CGIs occupied with the largest concentration of PcG complexes (Chapter 5). It seems possible that the dynamic nature and frequency of chromatin interactions by PcG complexes (as measured by live-cell imaging; Fonseca et al., 2012) might contribute to some level of nuclease hypersensitivity, potentially by altering nucleosomal:DNA interactions. The biochemical work characterising altered histone:DNA contacts in the presence of transcription factor binding events (Polach and Widom, 1995, Li and Widom, 2004) may also apply for more generic chromatin binding events, such as binding of non-methylated CpGs by ZF-CxxC domain-containing proteins, or even chromatin-modifying proteins which recognise and bind specific histone modifications. In general, CGI promoters are highly occupied by sequence-specific transcription factors and other generic chromatin-associated complexes (Wang et al., 2012, Chen et al., 2014c). Furthermore, live cell imaging of transcription factors (McNally et al., 2000, Voss et al., 2011, Swinstead et al., 2016) and chromatin-binding complexes (Cheutin et al., 2003, Deal and Henikoff, 2010, Fierz, 2016) in many cases reveals rapid binding and dissociation from chromatin. This dynamic association with chromatin could create transient fluctuations or alterations in chromatin structure that might be captured as chromatin accessibility by nuclease hypersensitivity assays. This hypothesis of a passive, non-specific effect on chromatin accessibility by chromatin-binding complexes could partly

explain the elevated chromatin accessibility of CGIs, given the high concentration of *trans*-acting factors at these loci.

So is chromatin accessibility a cause or consequence of regulatory activity at genomic elements? Most likely it represents both. Cells can specifically target active remodelling processes to CGIs and other regulatory elements in order to support new transcription factor binding events by evicting nucleosomes, and this may be supported by generic chromatin-interacting factors passively altering nucleosomal DNA. A combination of such features is likely to be important to maintain these chromatin states given the rapid binding kinetics of many *trans*-acting factors. Importantly, while many distal regulatory elements appear to be highly dependent upon a single factor or remodeller, CGIs are likely to have a greater redundancy for both active and passive effectors of chromatin accessibility, and this may be reflected in their relatively ubiquitous accessibility across different tissues. Alternatively, it remains possible that the hypomethylated state and underlying CpG density of CGIs may also be involved in facilitating such chromatin accessibility, but whether such features a sufficient for chromatin accessibility requires further work. The clear requirement of pre-existing chromatin accessibility for non-pioneer transcription factors to bind their target sites is the most apparent and direct consequence of chromatin accessibility, although others may exist. Ultimately, cells establish chromatin accessibility profiles that both reflect and instruct their regulation of gene expression through a combination of targeted, active remodelling and generic chromatin-modifying pathways.

APPENDICES

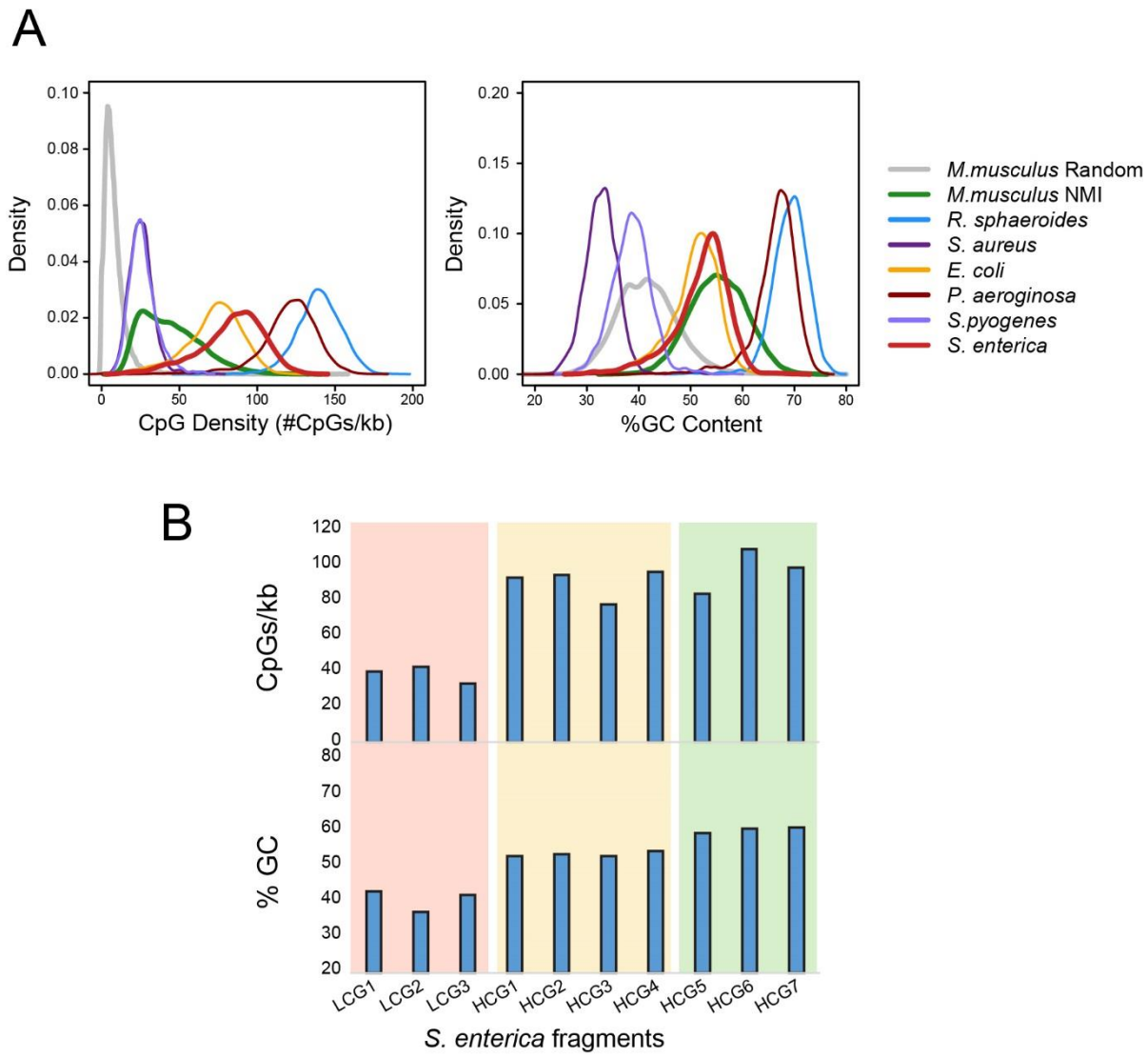
Appendix One – Supplementary Figures



Supplementary Figure 1 Reactivation of repetitive elements only partially explains changes in chromatin accessibility in DNMT TKO cells.

A) Visualisation of log2 fold change in RNA-seq between DNMT TKO and J1 wild type ESCs at repetitive elements, ranked according to CpG density.

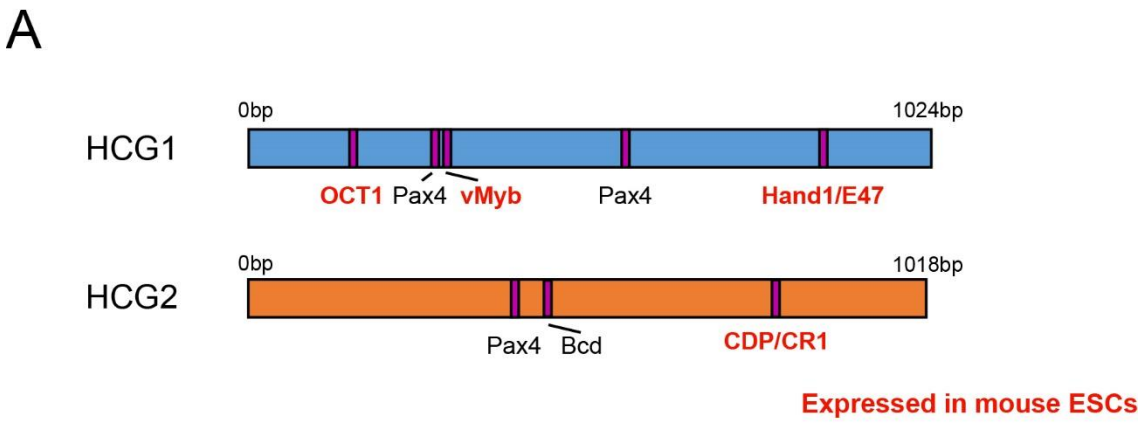
B) Scatterplot comparing log2 fold changes in RNA-seq and ATAC-seq between the DNMT TKO and J1 ESCs at repetitive elements. cor depicts Pearson correlation coefficient.



Supplementary Figure 2 Experimental design of bacterial DNA sequences.

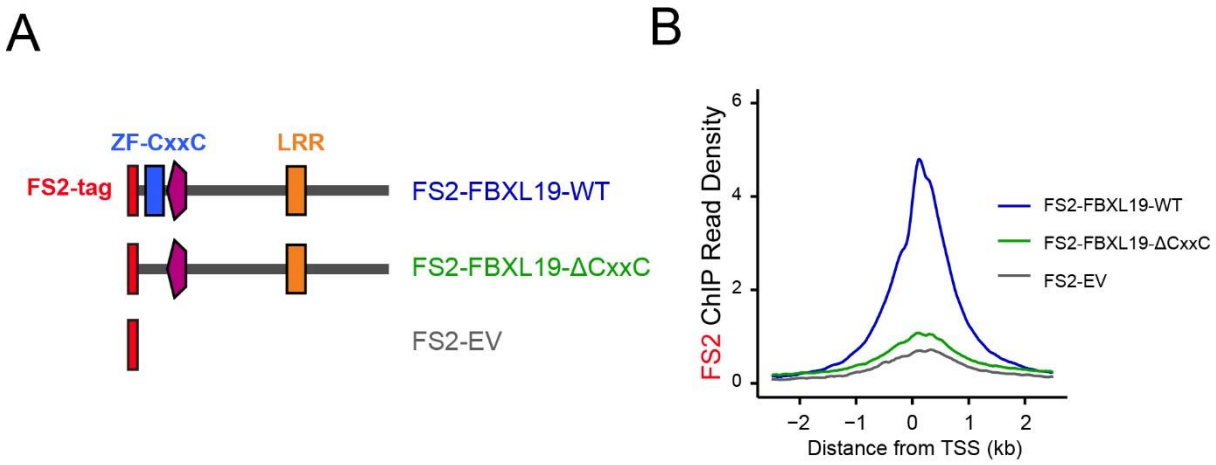
A) Bacterial genomes were bioinformatically fragmented into 1 kb sequences, and their CpG density and GC content were compared with that of mouse ESC NMI sequences or random genomic intervals.

B) Comparison of CpG density and GC content for the ten *S. enterica* fragments cloned for RMCE targeting into ESCs.



Supplementary Figure 3 Transcription factor motifs in RMCE bacterial sequences.

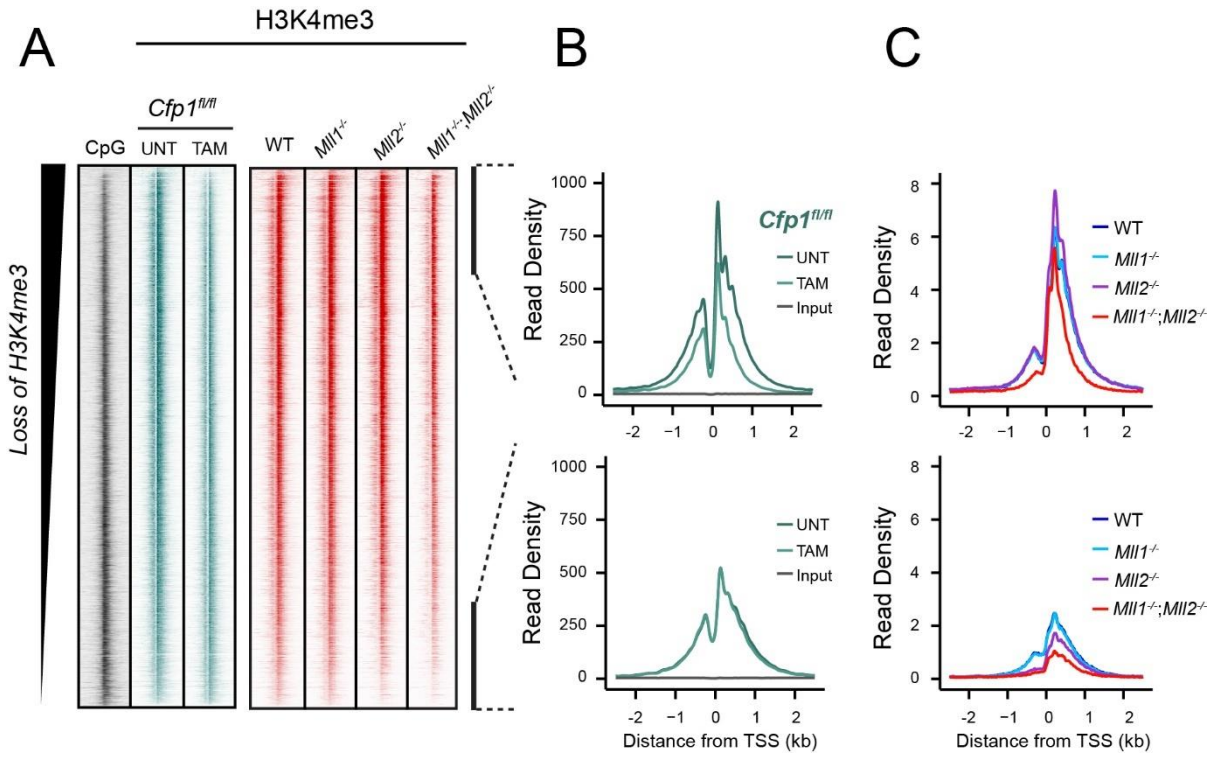
A) Transcription factor motifs were identified in the HCG1 and HCG2 recombinant sequences using MATCH^{TM}. Transcription factors which are expressed in ESCs are highlighted in red.



Supplementary Figure 4 The ZF-CxxC domain of FBXL19 is essential for occupancy at NMI promoters.

A) FS2-FBXL19 constructs used to generate ChIP-seq tracks for wild type FBXL19 (FS2-WT) and ZF-CxxC deletion (FS2-ΔCxxC).

B) Metaplot at NMI promoters in ESCs expressing FS2-FBXL19-WT, FS2-FBXL19-ΔCxxC or FS2-EV. (Courtesy of Dr Emilia Dimitrova)



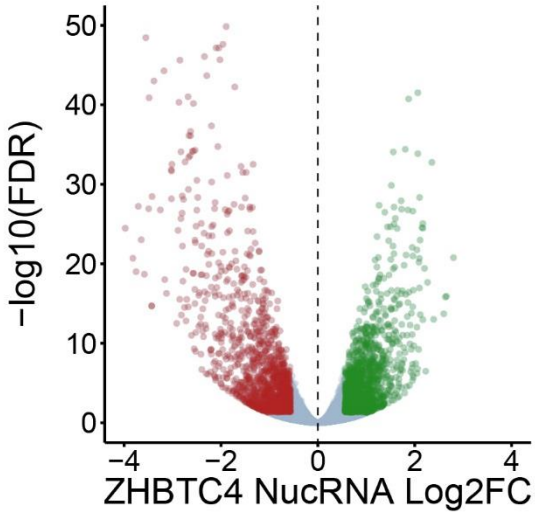
Supplementary Figure 5 Redundancy between CFP1, MLL1 and MLL2 for deposition of H3K4me3.

A) Heatmap of ESC-active TSS ranked by their relative loss of H3K4me3 after treatment of *Cfp1*$^{fl/fl}$ ESCs with tamoxifen for 96 h. H3K4me3 ChIP-seq from *Cfp1*$^{fl/fl}$ ESCs or ESCs from an independent study (Denissov et al., 2014) lacking either MLL1 (*Mll1*$^{-/-}$), MLL2 (*Mll2*$^{-/-}$) or both MLL1/2 (*Mll1*$^{-/-}$;*Mll2*$^{-/-}$).

B) Metaplot analysis of CFP1-dependent (upper panel) or CFP1-independent (lower panel) H3K4me3 at TSS in the *Cfp1*$^{fl/fl}$ ESCs.

C) As in (B), but for H3K4me3 ChIP-seq from *Mll1*$^{-/-}$, *Mll2*$^{-/-}$ and *Mll1*$^{-/-}$;*Mll2*$^{-/-}$ ESCs.

A



B

ZHBTC4 Down-regulated

TermID	Pathway	logP
WP1763	Pluripotency Network	-32.02
WP1498	Dopminergic Neurogenesis	-6.23
WP339	ESC Pluripotency Pathways	-5.16

TermID	Biological Process	logP
GO:0007275	multicellular organismal development	-30.47
GO:0032502	developmental process	-27.71
GO:0048731	system development	-27.54
GO:0030154	cell differentiation	-26.93
GO:0048513	organ development	-25.98
GO:0009888	tissue development	-25.13
GO:0048869	cellular developmental process	-24.75
GO:0010628	positive regulation of gene expression	-20.37

C

ZHBTC4 Up-regulated

TermID	Pathway	logP
WP1270	Endochondral Ossification	-5.14
WP1770	Carbon metabolism	-4.68
WP553	Calcium Regulation Cardiac Cell	-4.27

TermID	Biological Process	logP
GO:0007275	multicellular organismal development	-39.28
GO:0048731	system development	-38.85
GO:0032502	developmental process	-33.17
GO:0009653	anatomical structure morphogenesis	-29.67
GO:0007399	nervous system development	-27.75
GO:0048513	organ development	-26.77
GO:0072359	circulatory system development	-24.60
GO:0072358	cardiovascular system development	-24.60
GO:0030154	cell differentiation	-23.77
GO:0050793	regulation of developmental process	-23.62
GO:0048869	cellular developmental process	-22.20
GO:0022008	neurogenesis	-21.54

Supplementary Figure 6 OCT4-dependent regulation of pluripotency transcriptional network.

A) Gene expression changes after treatment of ZHBTC4 ESCs with doxycycline for 24 h. Log2 fold change is plotted against –log10 false discovery rate (FDR).

B) Gene ontology enrichment of significantly down-regulated genes after loss of OCT4 expression.

C) Same as in (B), but for ZHBTC4 up-regulated genes.

REFERENCE LIST

Adachi K, Nikaido I, Ohta H, Ohtsuka S, Ura H, Kadota M, Wakayama T, Ueda Hiroki R & Niwa H 2013. Context-Dependent Wiring of Sox2 Regulatory Networks for Self-Renewal of Embryonic and Trophoblast Stem Cells. *Molecular Cell,* 52**,** 380-392.

Adelman K & Lis JT 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet,* 13**,** 720-731.

Adey A, Morrison H, Asan, Xun X, Kitzman J, Turner E, Stackhouse B, MacKenzie A, Caruccio N, Zhang X*, et al.* 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology,* 11**,** R119.

Aksoy I, Jauch R, Chen J, Dyla M, Divakar U, Bogu GK, Teo R, Leng Ng CK, Herath W, Lili S*, et al.* 2013. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *The EMBO Journal,* 32**,** 938-953.

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC & Pugh BF 2007. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomycescerevisiae genome. *Nature,* 446**,** 572-576.

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U & Zoghbi HY 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet,* 23**,** 185-188.

Anderson JD, Lowary PT & Widom J 2001. Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites1. *Journal of Molecular Biology,* 307**,** 977-985.

Anderson JD & Widom J 2000. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *Journal of Molecular Biology,* 296**,** 979-987.

Anderson JD & Widom J 2001. Poly(dA-dT) Promoter Elements Increase the Equilibrium Accessibility of Nucleosomal DNA Target Sites. *Molecular and Cellular Biology,* 21**,** 3830-3839.

Antequera F, Boyes J & Bird A 1990. High levels of De Novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell,* 62**,** 503-514.

Antequera F, Macleod D & Bird AP 1989. Specific protection of methylated CpGs in mammalian nuclei. *Cell,* 58**,** 509-517.

Ardehali MB, Mei A, Zobeck KL, Caron M, Lis JT & Kusch T 2011. Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *The EMBO Journal,* 30**,** 2817-2828.

Armstrong JA, Papoulas O, Daubresse G, Sperling AS, Lis JT, Scott MP & Tamkun JW 2002. The Drosophila BRM complex facilitates global transcription by RNA polymerase II. *The EMBO Journal,* 21**,** 5245-5254.

Arndt PF, Burge CB & Hwa T 2003. DNA Sequence Evolution with Neighbor-Dependent Mutation. *Journal of Computational Biology,* 10**,** 313-322.

Arndt PF & Hwa T 2004. Regional and time-resolved mutation patterns of the human genome. *Bioinformatics,* 20**,** 1482-1485.

Auerbach RK, Euskirchen G, Rozowsky J, Lamarre-Vincent N, Moqtaderi Z, Lefrançois P, Struhl K, Gerstein M & Snyder M 2009. Mapping accessible chromatin regions using Sono-Seq. *Proceedings of the National Academy of Sciences,* 106**,** 14926-14931.

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH & Dhe-Paganon S 2008. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature,* 455**,** 822-825.

Bach C, Mueller D, Buhl S, Garcia-Cuellar MP & Slany RK 2009. Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. *Oncogene,* 28**,** 815-823.

Ball DJ, Gross DS & Garrard WT 1983. 5-methylcytosine is localized in nucleosomes that contain histone H1. *Proceedings of the National Academy of Sciences,* 80**,** 5490-5494.

Ballabio E & Milne TA 2012. Molecular and Epigenetic Mechanisms of MLL in Human Leukemogenesis. *Cancers,* 4**,** 904.

Ballaré C, Castellano G, Gaveglia L, Althammer S, González-Vallinas J, Eyras E, Le Dily F, Zaurin R, Soronellas D, Vicent Guillermo P*, et al.* 2013. Nucleosome-Driven Transcription Factor Binding and Gene Regulation. *Molecular Cell,* 49**,** 67-79.

Bao X, Rubin AJ, Qu K, Zhang J, Giresi PG, Chang HY & Khavari PA 2015. A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. *Genome Biology,* 16**,** 1-17.

Bao Y & Shen X 2007. INO80 subfamily of chromatin remodeling complexes. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 618**,** 18-29.
Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I & Zhao K 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell,* 129**,** 823-837.
Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M & Kouzarides T 2010. Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. *Cell,* 143**,** 470-484.
Baubec T, Ivánek R, Lienert F & Schübeler D 2013. Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. *Cell,* 153**,** 480-492.
Beck S, Lee B-K, Rhee C, Song J, Woo AJ & Kim J 2014. CpG island-mediated global gene regulatory modes in mouse embryonic stem cells. *Nature Communications,* 5**,** 5490.
Beh LY, Colwell LJ & Francis NJ 2012. A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. *Proceedings of the National Academy of Sciences,* 109**,** E1063-E1071.
Bell AC & Felsenfeld G 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature,* 405**,** 482-485.
Bell O, Schwaiger M, Oakeley EJ, Lienert F, Beisel C, Stadler MB & Schubeler D 2010. Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nat Struct Mol Biol,* 17**,** 894-900.
Bell O, Tiwari VK, Thomä NH & Schübeler D 2011. Determinants and dynamics of genome accessibility. *Nat Rev Genet,* 12**,** 554-564.
Bernstein E, Duncan EM, Masui O, Gil J, Heard E & Allis CD 2006. Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. *Molecular and Cellular Biology,* 26**,** 2560-2569.
Bestor TH 2000. The DNA methyltransferases of mammals. *Human Molecular Genetics,* 9**,** 2395-2402.
Bhatt D, Pandya-Jones A, Tong A-J, Barozzi I, Lissner M, Natoli G, Black D & Smale S 2012. Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. *Cell,* 150**,** 279-290.
Biddie S, John S, Sabo P, Thurman R, Johnson T, Schiltz R, Miranda T, Sung M-H, Trump S, Lightman S*, et al.* 2011a. Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. *Molecular Cell,* 43**,** 145-155.
Biddie Simon C, John S, Sabo Pete J, Thurman Robert E, Johnson Thomas A, Schiltz RL, Miranda Tina B, Sung M-H, Trump S, Lightman Stafford L*, et al.* 2011b. Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. *Molecular Cell,* 43**,** 145-155.
Bird AP 1978. Use of restriction enzymes to study eukaryotic DNA methylation. *Journal of Molecular Biology,* 118**,** 49-60.
Bird AP 1986. CpG-rich islands and the function of DNA methylation. *Nature,* 321**,** 209-213.
Bird AP 2002. DNA methylation patterns and epigenetic memory. *Genes & Development,* 16**,** 6-21.
Bird AP & Wolffe AP 1999. Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell,* 99**,** 451-454.
Birke M, Schreiner S, García-Cuéllar M-P, Mahr K, Titgemeyer F & Slany RK 2002. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Research,* 30**,** 958-965.
Blackledge N, Farcas A, Kondo T, King H, McGouran J, Hanssen L, Ito S, Cooper S, Kondo K, Koseki Y*, et al.* 2014. Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell,* 157**,** 1445-1459.
Blackledge NP & Klose R 2011. CpG island chromatin. *Epigenetics,* 6**,** 147-152.
Blackledge NP, Long HK, Zhou JC, Kriaucionis S, Patient R & Klose RJ 2012. Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA. *Nucleic Acids Research,* 40**,** e32.
Blackledge NP, Rose NR & Klose RJ 2015. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol,* 16**,** 643-649.
Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ & Klose RJ 2010. CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. *Molecular Cell,* 38**,** 179-190.
Blattler A & Farnham PJ 2013. Cross-talk between Site-specific Transcription Factors and DNA Methylation States. *Journal of Biological Chemistry,* 288**,** 34287-34294.

Bledau AS, Schmidt K, Neumann K, Hill U, Ciotta G, Gupta A, Torres DC, Fu J, Kranz A, Stewart AF*, et al.* 2014. The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development,* 141**,** 1022-1035.

Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, Imakaev M, Mirny LA, Wu C-t & Zhuang X 2016. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature,* 529**,** 418-422.

Bösken CA, Farnung L, Hintermair C, Merzel Schachter M, Vogel-Bachmayr K, Blazek D, Anand K, Fisher RP, Eick D & Geyer M 2014. The structure and substrate specificity of human Cdk12/Cyclin K. *Nature Communications,* 5.

Bostick M, Kim JK, Estève P-O, Clark A, Pradhan S & Jacobsen SE 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science,* 317**,** 1760-1764.

Boulard M, Edwards JR & Bestor TH 2015. FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat Genet,* 47**,** 479-485.

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew J-L, Ruan Y, Wei C-L, Ng HH*, et al.* 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research,* 18**,** 1752-1762.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG*, et al.* 2005. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell,* 122**,** 947-956.

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK*, et al.* 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature,* 441**,** 349-353.

Boyes J & Bird A 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell,* 64**,** 1123-1134.

Boyes J & Bird A 1992. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.,* 11**,** 327-333.

Boyes J & Felsenfeld G 1996. Tissue-specific factors additively increase the probability of the all-or-none formation of a hypersensitive site. *The EMBO Journal,* 15**,** 2496-2507.

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS & Crawford GE 2008. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell,* 132**,** 311-322.

Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Names A, Temper V, Razin A & Cedar H 1994. Spl elements protect a CpG island from de novo methylation. *Nature,* 371**,** 435-438.

Brogaard K, Xi L, Wang J-P & Widom J 2012. A map of nucleosome positions in yeast at base-pair resolution. *Nature,* 486**,** 496-501.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods,* 10**,** 1213-1218.

Buenrostro JD, Wu B, Chang HY & Greenleaf WJ 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology,* 109**,** 21.29.1-21.29.9.

Buganim Y, Faddah DA & Jaenisch R 2013. Mechanisms and models of somatic cell reprogramming. *Nat Rev Genet,* 14**,** 427-439.

Bultman S, Gebuhr T, Yee D, La Mantia C, Nicholson J, Gilliam A, Randazzo F, Metzger D, Chambon P, Crabtree G*, et al.* 2000. A Brg1 Null Mutation in the Mouse Reveals Functional Differences among Mammalian SWI/SNF Complexes. *Molecular Cell,* 6**,** 1287-1295.

Bultman SJ, Gebuhr TC, Pan H, Svoboda P, Schultz RM & Magnuson T 2006. Maternal BRG1 regulates zygotic genome activation in the mouse. *Genes & Development,* 20**,** 1744-1754.

Calabrese JM, Sun W, Song L, Mugford Joshua W, Williams L, Yee D, Starmer J, Mieczkowski P, Crawford Gregory E & Magnuson T 2012. Site-Specific Silencing of Regulatory Elements as a Mechanism of X Inactivation. *Cell,* 151**,** 951-963.

Cardozo T & Pagano M 2004. The SCF ubiquitin ligase: insights into a molecular machine. *Nat Rev Mol Cell Biol,* 5**,** 739-751.

Carlone DL, Lee J-H, Young SRL, Dobrota E, Butler JS, Ruiz J & Skalnik DG 2005. Reduced Genomic Cytosine Methylation and Defective Cellular Differentiation in Embryonic Stem Cells Lacking CpG Binding Protein. *Molecular and Cellular Biology,* 25**,** 4881-4891.

Carlone DL & Skalnik DG 2001. CpG Binding Protein Is Crucial for Early Embryonic Development. *Molecular and Cellular Biology,* 21**,** 7601-7606.

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CAM, Taylor MS, Engstrom PG, Frith MC*, et al.* 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet,* 38**,** 626-635.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia W-J, Anderson S, Yates J, Washburn MP*, et al.* 2005. Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell,* 123**,** 581-592.

Ceballos-Chávez M, Subtil-Rodríguez A, Giannopoulou EG, Soronellas D, Vázquez-Chávez E, Vicent GP, Elemento O, Beato M & Reyes JC 2015. The Chromatin Remodeler CHD8 Is Required for Activation of Progesterone Receptor-Dependent Enhancers. *PLoS Genet,* 11**,** e1005174.

Chang P-Y, Hom RA, Musselman CA, Zhu L, Kuo A, Gozani O, Kutateladze TG & Cleary ML 2010. Binding of the MLL PHD3 Finger to Histone H3K4me3 Is Required for MLL-Dependent Gene Transcription. *Journal of Molecular Biology,* 400**,** 137-144.

Chao S-H & Price DH 2001. Flavopiridol Inactivates P-TEFb and Blocks Most RNA Polymerase II Transcription in Vivo. *Journal of Biological Chemistry,* 276**,** 31793-31799.

Chédin F, Lieber MR & Hsieh C-L 2002. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences,* 99**,** 16916-16921.

Chen J, Chen X, Li M, Liu X, Gao Y, Kou X, Zhao Y, Zheng W, Zhang X, Huo Y*, et al.* 2016. Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. *Cell Reports,* 14**,** 1540-1554.

Chen J, Zhang Z, Li L, Chen B-C, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E*, et al.* 2014a. Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. *Cell,* 156**,** 1274-1285.

Chen K, Xi Y, Pan X, Li Z, Kaestner K, Tyler J, Dent S, He X & Li W 2013. DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Research,* 23**,** 341-351.

Chen PB, Zhu LJ, Hainer SJ, McCannell KN & Fazzio TG 2014b. Unbiased chromatin accessibility profiling by RED-seq uncovers unique features of nucleosome variants in vivo. *BMC Genomics,* 15**,** 1-18.

Chen RA-J, Stempor P, Down TA, Zeiser E, Feuer SK & Ahringer J 2014c. Extreme HOT regions are CpG-dense promoters in C. elegans and humans. *Genome Research,* 24**,** 1138-1146.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J*, et al.* 2008. Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell,* 133**,** 1106-1117.

Cheutin T, McNairn AJ, Jenuwein T, Gilbert DM, Singh PB & Misteli T 2003. Maintenance of Stable Heterochromatin Domains by Dynamic HP1 Binding. *Science,* 299**,** 721-725.

Choi JK 2010. Contrasting chromatin organization of CpG islands and exons in the human genome. *Genome Biology,* 11**,** 1-8.

Chuang LS-H, Ian H-I, Koh T-W, Ng H-H, Xu G & Li BFL 1997. Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21. *Science,* 277**,** 1996-2000.

Chung JH, Whiteley M & Felsenfeld G 1993. A 5′ element of the chicken β-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell,* 74**,** 505-514.

Cieply B, Park Juw W, Nakauka-Ddamba A, Bebee Thomas W, Guo Y, Shang X, Lengner Christopher J, Xing Y & Carstens Russ P 2016. Multiphasic and Dynamic Changes in Alternative Splicing during Induction of Pluripotency Are Coordinated by Numerous RNA-Binding Proteins. *Cell Reports,* 15**,** 247-255.

Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M & Zaret KS 2002. Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Molecular Cell,* 9**,** 279-289.

Clapier CR & Cairns BR 2009. The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry,* 78**,** 273-304.

Clayton AL, Hazzalin CA & Mahadevan LC 2006. Enhanced Histone Acetylation and Transcription: A Dynamic Perspective. *Molecular Cell,* 23**,** 289-296.

Clouaire T, de las Heras JI, Merusi C & Stancheva I 2010. Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic Acids Research,* 38**,** 4620-4634.

Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, Lee J-H, Skalnik D & Bird A 2012. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & Development,* 26**,** 1714-1728.

Cockerill PN 2011. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS Journal,* 278**,** 2182-2210.

Collings CK, Waddell PJ & Anderson JN 2013. Effects of DNA methylation on nucleosome stability. *Nucleic Acids Research,* 41**,** 2918-2931.

Conroy A, Stockett DE, Walker D, Arkin MR, Hoch U, Fox JA & Hawtin RE 2009. SNS-032 is a potent and selective CDK 2, 7 and 9 inhibitor that drives target modulation in patient samples. *Cancer Chemotherapy and Pharmacology,* 64**,** 723-732.

Cooper S, Dienstbier M, Hassan R, Schermelleh L, Sharif J, Blackledge Neil P, De Marco V, Elderkin S, Koseki H, Klose R*, et al.* 2014. Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell Reports,* 7**,** 1456-1470.

Core LJ, Waterfall JJ & Lis JT 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science,* 322**,** 1845-1848.

Corona DFV & Tamkun JW 2004. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 1677**,** 113-119.

Costello JF, Fruhwald MC, Smiraglia DJ, Rush LJ, Robertson GP, Gao X, Wright FA, Feramisco JD, Peltomaki P, Lang JC*, et al.* 2000. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet,* 24**,** 132-138.

Côté J, Peterson CL & Workman JL 1998. Perturbation of nucleosome core structure by the SWI/SNF complex persists after its detachment, enhancing subsequent transcription factor binding. *Proceedings of the National Academy of Sciences,* 95**,** 4947-4952.

Coulondre C, Miller JH, Farabaugh PJ & Gilbert W 1978. Molecular basis of base substitution hotspots in Escherichia coli. *Nature,* 274**,** 775-780.

Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D*, et al.* 2006. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Research,* 16**,** 123-131.

Cross SH, Meehan RR, Nan X & Bird A 1997. A component of the transcriptional represser MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat Genet,* 16**,** 256-259.

de Dieuleveult M, Yen K, Hmitou I, Depaux A, Boussouar F, Dargham DB, Jounier S, Humbertclaude H, Ribierre F, Baulard C*, et al.* 2016. Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature,* 530**,** 113-116.

Deal RB & Henikoff S 2010. Capturing the dynamic epigenome. *Genome Biology,* 11**,** 1-8.

Deaton AM & Bird A 2011. CpG islands and the regulation of transcription. *Genes & Development,* 25**,** 1010-1022.

Deaton AM, Gómez-Rodríguez M, Mieczkowski J, Tolstorukov MY, Kundu S, Sadreyev RI, Jansen LET & Kingston RE 2016. Enhancer regions show high histone H3.3 turnover that changes during differentiation. *eLife,* 5**,** e15316.

Dellino GI, Schwartz YB, Farkas G, McCabe D, Elgin SCR & Pirrotta V 2004. Polycomb Silencing Blocks Transcription Initiation. *Molecular Cell,* 13**,** 887-893.

Deng T, Zhu ZI, Zhang S, Leng F, Cherukuri S, Hansen L, Mariño-Ramírez L, Meshorer E, Landsman D & Bustin M 2013. HMGN1 Modulates Nucleosome Occupancy and DNase I Hypersensitivity at the CpG Island Promoters of Embryonic Stem Cells. *Molecular and Cellular Biology,* 33**,** 3377-3389.

Denissov S, Hofemeister H, Marks H, Kranz A, Ciotta G, Singh S, Anastassiadis K, Stunnenberg HG & Stewart AF 2014. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development,* 141**,** 526-537.

Deshaies RJ 1999. SCF and Cullin/RING H2-Based Ubiquitin Ligases. *Annual Review of Cell and Developmental Biology,* 15**,** 435-467.

Di Croce L & Helin K 2013. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol,* 20**,** 1147-1155.

Ding J, Xu H, Faiola F, Ma'ayan A & Wang J 2012. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Research,* 22**,** 155-167.

Dingwall C, Lomonossoff GP & Laskey RA 1981. High sequence specificity of micrococcal nuclease. *Nucleic Acids Res.,* 9**,** 2659–2673.

Dion MF, Altschuler SJ, Wu LF & Rando OJ 2005. Genomic characterization reveals a simple histone H4 acetylation code. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 5501-5506.
Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.
Doerfler W 1983. DNA Methylation and Gene Activity. *Annual Review of Biochemistry,* 52**,** 93-124.
Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L & Schübeler D 2015. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature,* 528**,** 575-579.
Du Q, Luu P-L, Stirzaker C & Clark SJ 2015. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics,* 7**,** 1051-1073.
Duncan BK & Miller JH 1980. Mutagenic deamination of cytosine residues in DNA. *Nature,* 287**,** 560-561.
Eberl HC, Spruijt Cornelia G, Kelstrup Christian D, Vermeulen M & Mann M 2013. A Map of General and Specialized Chromatin Readers in Mouse Tissues Generated by Label-free Interaction Proteomics. *Molecular Cell,* 49**,** 368-378.
Endoh M, Endo TA, Endoh T, Fujimura Y-i, Ohara O, Toyoda T, Otte AP, Okano M, Brockdorff N, Vidal M*, et al.* 2008. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development,* 135**,** 1513-1524.
Erdel F & Rippe K 2012. Quantifying transient binding of ISWI chromatin remodelers in living cells by pixel-wise photobleaching profile evolution analysis. *Proceedings of the National Academy of Sciences,* 109**,** E3221–E3230.
Erdel F, Schubert T, Marth C, Längst G & Rippe K 2010. Human ISWI chromatin-remodeling complexes sample nucleosomes via transient binding reactions and become immobilized at active sites. *Proceedings of the National Academy of Sciences,* 107**,** 19873-19878.
Ernst P, Fisher JK, Avery W, Wade S, Foy D & Korsmeyer SJ 2004. Definitive Hematopoiesis Requires the Mixed-Lineage Leukemia Gene. *Developmental Cell,* 6**,** 437-443.
Ernst P & Vakoc CR 2012. WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Briefings in Functional Genomics,* 11**,** 217-226.
Esch D, Vahokoski J, Groves MR, Pogenberg V, Cojocaru V, vom Bruch H, Han D, Drexler HCA, Araúzo-Bravo MJ, Ng CKL*, et al.* 2013. A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol,* 15**,** 295-301.
Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultchi AI, Wutz A*, et al.* 2010. Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Molecular Cell,* 38**,** 452-464.
Espinosa JM & Emerson BM 2001. Transcriptional Regulation by p53 through Intrinsic DNA/Chromatin Binding and Site-Directed Cofactor Recruitment. *Molecular Cell,* 8**,** 57-69.
FANTOM, RIKEN PMI & CLST 2014. A promoter-level mammalian expression atlas. *Nature,* 507**,** 462-470.
Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW*, et al.* 2012. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife,* 1**,** e00205.
Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L & Schübeler D 2013. Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions. *PLoS Genet,* 9**,** e1003994.
Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME*, et al.* 2010. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences,* 107**,** 8689-8694.
Fenouil R, Cauchy P, Koch F, Descostes N, Cabeza JZ, Innocenti C, Ferrier P, Spicuglia S, Gut M, Gut I*, et al.* 2012. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Research,* 22**,** 2399-2408.
Fierz B 2016. Dynamic Chromatin Regulation from a Single Molecule Perspective. *ACS Chemical Biology,* 11**,** 609-620.
Filion GJP, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E & Defossez P-A 2006. A Family of Human Zinc Finger Proteins That Bind Methylated DNA and Repress Transcription. *Molecular and Cellular Biology,* 26**,** 169-181.
Fischle W, Wang Y & Allis CD 2003a. Histone and chromatin cross-talk. *Current Opinion in Cell Biology,* 15**,** 172-183.

Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD & Khorasanizadeh S 2003b. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & Development,* 17**,** 1870-1881.

Fitzgerald DP & Bender W 2001. Polycomb Group Repression Reduces DNA Accessibility. *Molecular and Cellular Biology,* 21**,** 6585-6597.

Flanagan JF, Mi L-Z, Chruszcz M, Cymborowski M, Clines KL, Kim Y, Minor W, Rastinejad F & Khorasanizadeh S 2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature,* 438**,** 1181-1185.

Fonseca JP, Steffen PA, Müller S, Lu J, Sawicka A, Seiser C & Ringrose L 2012. In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells. *Genes & Development,* 26**,** 857-871.

Foo Sun M, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien C-Y, Kirov N, Shvartsman Stanislav Y & Rushlow Christine A 2014. Zelda Potentiates Morphogen Activity by Increasing Chromatin Accessibility. *Current Biology,* 24**,** 1341-1346.

Francis NJ, Kingston RE & Woodcock CL 2004. Chromatin Compaction by a Polycomb Group Protein Complex. *Science,* 306**,** 1574-1577.

Frank CL, Manandhar D, Gordân R & Crawford GE 2016. HDAC inhibitors cause site-specific chromatin remodeling at PU.1-bound enhancers in K562 cells. *Epigenetics & Chromatin,* 9**,** 1-17.

Frauer C, Rottach A, Meilinger D, Bultmann S, Fellinger K, Hasenöder S, Wang M, Qin W, Söding J, Spada F*, et al.* 2011. Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1. *PLoS ONE,* 6**,** e16627.

Frescas D, Guardavaccaro D, Kuchay SM, Kato H, Poleshko A, Basrur V, Elenitoba-Johnson KS, Katz RA & Pagano M 2008. KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. *Cell Cycle,* 7**,** 3539-3547.

Friedel CC & Dolken L 2009. Metabolic tagging and purification of nascent RNA: implications for transcriptomics. *Molecular BioSystems,* 5**,** 1271-1278.

Fujita N, Takebayashi S, Okumura K, Kudo S, Chiba T, Saya H & Nakao M 1999. Methylation-Mediated Transcriptional Silencing in Euchromatin by Methyl-CpG Binding Protein MBD1 Isoforms. *Mol Cell Biol,* 19**,** 6415-6426.

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP & Kouzarides T 2003. The Methyl-CpG-binding Protein MeCP2 Links DNA Methylation to Histone Methylation. *Journal of Biological Chemistry,* 278**,** 4035-4040.

Fukuda T, Tokunaga A, Sakamoto R & Yoshida N 2011. Fbxl10/Kdm2b deficiency accelerates neural progenitor cell death and leads to exencephaly. *Molecular and Cellular Neuroscience,* 46**,** 614-624.

Gao H, Lukin K, Ramírez J, Fields S, Lopez D & Hagman J 2009. Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5. *Proceedings of the National Academy of Sciences,* 106**,** 11258-11263.

Garcia BA, Thomas CE, Kelleher NL & Mizzen CA 2008. Tissue-Specific Expression and Post-Translational Modification of Histone H3 Variants. *Journal of Proteome Research,* 7**,** 4225-4236.

Gardiner-Garden M & Frommer M 1987. CpG Islands in vertebrate genomes. *Journal of Molecular Biology,* 196**,** 261-282.

Gargiulo G, Levy S, Bucci G, Romanenghi M, Fornasari L, Beeson KY, Goldberg SM, Cesaroni M, Ballarini M, Santoro F*, et al.* 2009. NA-Seq: A Discovery Tool for the Analysis of Chromatin Structure and Dynamics during Differentiation. *Developmental Cell,* 16**,** 466-481.

Garrard WT 1991. Histone H1 and the conformation of transcriptlonally active chromatin. *BioEssays,* 13**,** 87-88.

Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, Heidersbach A, Ramalho-Santos J, McManus MT, Plath K, Meshorer E*, et al.* 2009. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature,* 460**,** 863-868.

Gaszner M & Felsenfeld G 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet,* 7**,** 703-713.

Giresi PG, Kim J, McDaniell RM, Iyer VR & Lieb JD 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research,* 17**,** 877-885.

Glaser S, Lubitz S, Loveland KL, Ohbo K, Robb L, Schwenk F, Seibler J, Roellig D, Kranz A, Anastassiadis K*, et al.* 2009. The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics & Chromatin,* 2**,** 1-16.

Glaser S, Schaft J, Lubitz S, Vintersten K, van der Hoeven F, Tufteland KR, Aasland R, Anastassiadis K, Ang S-L & Stewart AF 2006. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development,* 133**,** 1423-1432.

Goessling E, Kel-Margoulis OV, Kel AE & Wingender E 2001. MATCH - a tool for searching transcription factor binding sites in DNA sequences. http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi.

Göke J, Jung M, Behrens S, Chavez L, O'Keeffe S, Timmermann B, Lehrach H, Adjaye J & Vingron M 2011. Combinatorial Binding in Human and Mouse Embryonic Stem Cells Identifies Conserved Enhancers Active in Early Embryonic Development. *PLoS Comput Biol,* 7**,** e1002304.

Görisch SM, Wachsmuth M, Tóth KF, Lichter P & Rippe K 2005. Histone acetylation increases chromatin accessibility. *Journal of Cell Science,* 118**,** 5825-5834.

Gowher H, Leismann O & Jeltsch A 2000. DNA of Drosophila melanogaster contains 5-methylcytosine. *The EMBO Journal,* 19**,** 6918-6923.

Grau DJ, Chapman BA, Garlick JD, Borowsky M, Francis NJ & Kingston RE 2011. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes & Development,* 25**,** 2210-2221.

Greer Eric L, Blanco Mario A, Gu L, Sendinc E, Liu J, Aristizábal-Corrales D, Hsu C-H, Aravind L, He C & Shi Y 2015. DNA Methylation on N6-Adenine in C. elegans. *Cell,* 161**,** 868-878.

Grewal SIS & Jia S 2007. Heterochromatin revisited. *Nat Rev Genet,* 8**,** 35-46.

Grøntved L, John S, Baek S, Liu Y, Buckley JR, Vinson C, Aguilera G & Hager GL 2013. C/EBP maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements. *The EMBO Journal,* 32**,** 1568-1583.

Gruenbaum Y, Cedar H & Razin A 1982. Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature,* 295**,** 620-622.

Guccione E, Martinato F, Finocchiaro G, Luzi L, Tizzoni L, Dall' Olio V, Zardo G, Nervi C, Bernard L & Amati B 2006. Myc-binding-site recognition in the human genome is determined by chromatin context. *Nat Cell Biol,* 8**,** 764-770.

Guenther MG, Jenner RG, Chevalier B, Nakamura T, Croce CM, Canaani E & Young RA 2005. Global and Hox-specific roles for the MLL1 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 8603-8608.

Guenther MG, Levine SS, Boyer LA, Jaenisch R & Young RA 2007. A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell,* 130**,** 77-88.

Guertin MJ & Lis JT 2010. Chromatin Landscape Dictates HSF Binding to Target DNA Elements. *PLoS Genet,* 6**,** e1001114.

Guertin MJ & Lis JT 2013. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Current Opinion in Genetics & Development,* 23**,** 116-123.

Gupta S, Dennis J, Thurman RE, Kingston R, Stamatoyannopoulos JA & Noble WS 2008. Predicting Human Nucleosome Occupancy from Primary Sequence. *PLoS Comput Biol,* 4**,** e1000134.

Hahn M, Dambacher S, Dulev S, Kuznetsova AY, Eck S, Wörz S, Sadic D, Schulte M, Mallm J-P, Maiser A*, et al.* 2013. Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin. *Genes & Development,* 27**,** 859-872.

Hainer Sarah J & Fazzio Thomas G 2015. Regulation of Nucleosome Architecture and Factor Binding Revealed by Nuclease Footprinting of the ESC Genome. *Cell Reports,* 13**,** 61-69.

Hainer SJ, Gu W, Carone BR, Landry BD, Rando OJ, Mello CC & Fazzio TG 2015. Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF. *Genes & Development,* 29**,** 362-378.

Hallson G, Hollebakken RE, Li T, Syrzycka M, Kim I, Cotsworth S, Fitzpatrick KA, Sinclair DAR & Honda BM 2012. dSet1 Is the Main H3K4 Di- and Tri-Methyltransferase Throughout Drosophila Development. *Genetics,* 190**,** 91-100.

Han H, Cortez CC, Yang X, Nichols PW, Jones PA & Liang G 2011. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Human Molecular Genetics,* 20**,** 4299-4310.

Han XR, Zha Z, Yuan HX, Feng X, Xia YK, Lei QY, Guan KL & Xiong Y 2016. KDM2B/FBXL10 targets c-Fos for ubiquitylation and degradation in response to mitogenic stimulation. *Oncogene,* 35**,** 4179-4190.

Hargreaves DC & Crabtree GR 2011. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res,* 21**,** 396-420.

Harris M 1982. Induction of thymidine kinase in enzyme-deficient chinese hamster cells. *Cell,* 29**,** 483-492.
Hartono SR, Korf IF & Chédin F 2015. GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. *Nucleic Acids Research,* 43**,** 9729-9741.
Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE & Cheng X 2008. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature,* 455**,** 826-829.
Hashimoto K, Kokubun S, Itoi E & Roach H 2007. Improved quantification of DNA methylation using methylation-sensitive restriction enzymes and real-time PCR. *Epigenetics,* 2**,** 86-91.
Hashimoto TB, Sherwood R, Kang DD, Rajagopal N, Barkal AA, Zeng H, Emons BJ, Srinivasan S, Jaakkola T & Gifford D 2016. A synergistic DNA logic predicts genome-wide chromatin accessibility. *Genome Research*.
Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, Carrozza MJ & Workman JL 2002. Function and Selectivity of Bromodomains in Anchoring Chromatin-Modifying Complexes to Promoter Nucleosomes. *Cell,* 111**,** 369-379.
Hatta M & Cirillo LA 2007. Chromatin Opening and Stable Perturbation of Core Histone:DNA Contacts by FoxO1. *Journal of Biological Chemistry,* 282**,** 35583-35593.
Hay D, Hughes JR, Babbs C, Davies JOJ, Graham BJ, Hanssen LLP, Kassouf MT, Oudelaar AM, Sharpe JA, Suciu MC*, et al.* 2016. Genetic dissection of the α-globin super-enhancer in vivo. *Nat Genet,* 48**,** 895-903.
He HH, Meyer CA, Chen MW, Jordan VC, Brown M & Liu XS 2012. Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. *Genome Research,* 22**,** 1015-1025.
He HH, Meyer CA, Hu SeS, Chen M-W, Zang C, Liu Y, Rao PK, Fei T, Xu H, Long H*, et al.* 2014. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat Meth,* 11**,** 73-78.
He J, Kallin EM, Tsukada Y-i & Zhang Y 2008. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15Ink4b. *Nat Struct Mol Biol,* 15**,** 1169-1175.
He J, Shen L, Wan M, Taranova O, Wu H & Zhang Y 2013. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat Cell Biol,* 15**,** 373-384.
Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW*, et al.* 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature,* 459**,** 108-112.
Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H & Glass CK 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell,* 38**,** 576-589.
Hendrich B & Bird A 1998. Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Molecular and Cellular Biology,* 18**,** 6538-6547.
Henikoff S 2008. Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat Rev Genet,* 9**,** 15-26.
Henriques T, Gilchrist Daniel A, Nechaev S, Bern M, Muse Ginger W, Burkholder A, Fargo David C & Adelman K 2013. Stable Pausing by RNA Polymerase II Provides an Opportunity to Target and Integrate Regulatory Signals. *Molecular Cell,* 52**,** 517-528.
Higgs DR, Wood WG, Jarman AP, Sharpe J, Lida J, Pretorius IM & Ayyub H 1990. A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes & Development,* 4**,** 1588-1601.
Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F*, et al.* 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research,* 34**,** D590-D598.
Ho JWK, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn K-A, Minoda A, Tolstorukov MY, Appert A*, et al.* 2014. Comparative analysis of metazoan chromatin organization. *Nature,* 512**,** 449-452.
Ho L, Jothi R, Ronan JL, Cui K, Zhao K & Crabtree GR 2009. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proceedings of the National Academy of Sciences,* 106**,** 5187-5191.
Ho L, Miller EL, Ronan JL, Ho WQ, Jothi R & Crabtree GR 2011. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat Cell Biol,* 13**,** 903-913.

Höller M, Westin G, Jiricny J & Schaffner W 1988. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes & Development,* 2**,** 1127-1135.

Holliday R & Pugh J 1975. DNA modification mechanisms and gene activity during development. *Science,* 187**,** 226-232.

Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD & Ren B 2013. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet,* 45**,** 1198-1206.

Hong L, Schroth GP, Matthews HR, Yau P & Bradbury EM 1993. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *Journal of Biological Chemistry,* 268**,** 305-314.

Hosseini M, Goodstadt L, Hughes JR, Kowalczyk MS, de Gobbi M, Otto GW, Copley RR, Mott R, Higgs DR & Flint J 2013. Causes and Consequences of Chromatin Variation between Inbred Mice. *PLoS Genet,* 9**,** e1003570.

Hsia DA, Tepper CG, Pochampalli MR, Hsia EYC, Izumiya C, Huerta SB, Wright ME, Chen H-W, Kung H-J & Izumiya Y 2010. KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. *Proceedings of the National Academy of Sciences,* 107**,** 9671-9676.

Hsieh C 1994. Dependence of transcriptional repression on CpG methylation density. *Mol Cell Biol,* 14**,** 5487-5494.

Hu B, Petela N, Kurze A, Chan K-L, Chapard C & Nasmyth K 2015. Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Research,* 43**,** e132.

Hu G, Cui K, Northrup D, Liu C, Wang C, Tang Q, Ge K, Levens D, Crane-Robinson C & Zhao K 2013. H2A.Z Facilitates Access of Active and Repressive Complexes to Chromatin in Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell,* 12**,** 180-192.

Hu G, Schones DE, Cui K, Ybarra R, Northrup D, Tang Q, Gattinoni L, Restifo NP, Huang S & Zhao K 2011. Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Research,* 21**,** 1650-1658.

Hu J-L, Zhou BO, Zhang R-R, Zhang K-L, Zhou J-Q & Xu G-L 2009. The N-terminus of histone H3 is required for de novo DNA methylation in chromatin. *Proceedings of the National Academy of Sciences,* 106**,** 22187-22192.

Iguchi-Ariga SM & Schaffner W 1989. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes & Development,* 3**,** 612-619.

Iida T, Suetake I, Tajima S, Morioka H, Ohta S, Obuse C & Tsurimoto T 2002. PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. *Genes to Cells,* 7**,** 997-1007.

Illingworth RS & Bird AP 2009. CpG islands – 'A rough guide'. *FEBS Letters,* 583**,** 1713-1720.

Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R & Bird AP 2010. Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genet,* 6**,** e1001134.

Ishii H, Kadonaga JT & Ren B 2015. MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proceedings of the National Academy of Sciences,* 112**,** E3457-E3465.

Ishimura A, Minehata K-i, Terashima M, Kondoh G, Hara T & Suzuki T 2012. Jmjd5, an H3K36me2 histone demethylase, modulates embryonic cell proliferation through the regulation of Cdkn1a expression. *Development,* 139**,** 749-759.

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

Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts Jason A, Mahony S, Pugh BF, Lee D, Kaestner Klaus H & Zaret Kenneth S 2016. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Molecular Cell,* 62**,** 79-91.

Iwafuchi-Doi M & Zaret KS 2014. Pioneer transcription factors in cell reprogramming. *Genes & Development,* 28**,** 2679-2692.

Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR & Haussler D 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature,* 516**,** 242-245.

Jacques P-É, Jeyakani J & Bourque G 2013. The Majority of Primate-Specific Regulatory Sequences Are Derived from Transposable Elements. *PLoS Genet,* 9**,** e1003504.
Jaenisch R, Schnieke A & Harbers K 1985. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. *Proc Natl Acad Sci USA,* 82**,** 1451-1455.
Janzer A, Stamm K, Becker A, Zimmer A, Buettner R & Kirfel J 2012. The H3K4me3 Histone Demethylase Fbxl10 Is a Regulator of Chemokine Expression, Cellular Morphology, and the Metabolome of Fibroblasts. *Journal of Biological Chemistry,* 287**,** 30984-30992.
Jauch R, Aksoy I, Hutchins AP, Ng CKL, Tian XF, Chen J, Palasingam P, Robson P, Stanton LW & Kolatkar PR 2011. Conversion of Sox17 into a Pluripotency Reprogramming Factor by Reengineering Its Association with Oct4 on DNA. *Stem Cells,* 29**,** 940-951.
Jeong KW, Andreu-Vieyra C, You JS, Jones PA & Stallcup MR 2014. Establishment of active chromatin structure at enhancer elements by mixed-lineage leukemia 1 to initiate estrogen-dependent gene expression. *Nucleic Acids Research,* 42**,** 2245-2256.
Jerabek S, Merino F, Schöler HR & Cojocaru V 2014. OCT4: Dynamic DNA binding pioneers stem cell pluripotency. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1839**,** 138-154.
Jermann P, Hoerner L, Burger L & Schübeler D 2014. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proceedings of the National Academy of Sciences,* 111**,** E3415-E3421.
Jiang C & Pugh BF 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet,* 10**,** 161-172.
Jin C & Felsenfeld G 2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & Development,* 21**,** 1519-1529.
Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K & Felsenfeld G 2009. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet,* 41**,** 941-945.
Jin J, Cardozo T, Lovering RC, Elledge SJ, Pagano M & Harper JW 2004. Systematic analysis and nomenclature of mammalian F-box proteins. *Genes & Development,* 18**,** 2573-2580.
John S, Sabo PJ, Canfield TK, Lee K, Vong S, Weaver M, Wang H, Vierstra J, Reynolds AP, Thurman RE*, et al.* 2013. Genome-Scale Mapping of DNase I Hypersensitivity. *Current Protocols in Molecular Biology,* 103**,** 21.27.1-21.27.20.
John S, Sabo PJ, Thurman RE, Sung M-H, Biddie SC, Johnson TA, Hager GL & Stamatoyannopoulos JA 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet,* 43**,** 264-268.
Johnson SM, Tan FJ, McCullough HL, Riordan DP & Fire AZ 2006. Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin. *Genome Research,* 16**,** 000.
Jones PA 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet,* 13**,** 484-492.
Jones PL, Jan Veenstra GC, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J & Wolffe AP 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet,* 19**,** 187-191.
Jonkers I, Kwak H & Lis JT 2014. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife,* 3**,** e02407.
Kadam S & Emerson BM 2003. Transcriptional Specificity of Human SWI/SNF BRG1 and BRM Chromatin Remodeling Complexes. *Molecular Cell,* 11**,** 377-389.
Kadoch C & Crabtree GR 2015. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Science Advances,* 1**,** e1500447.
Kalb R, Latwiel S, Baymaz HI, Jansen PWTC, Müller CW, Vermeulen M & Müller J 2014. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol,* 21**,** 569-571.
Kamakaka RT & Thomas JO 1990. Chromatin structure of transcriptionally competent and repressed genes. *EMBO J,* 9**,** 3997-4006.
Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J*, et al.* 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature,* 458**,** 362-366.
Kass SU, Landsberger N & Wolffe AP 1997. DNA methylation directs a time-dependent repression of transcription initiation. *Current Biology,* 7**,** 157-165.

Katoh M & Katoh M 2004. Identification and characterization of FBXL19 gene in silico. *International Journal of Molecular Medicine,* 14**,** 1109-1114.
Kawakami E, Tokunaga A, Ozawa M, Sakamoto R & Yoshida N 2015. The histone demethylase Fbxl11/Kdm2a plays an essential role in embryonic development by repressing cell-cycle regulators. *Mechanisms of Development,* 135**,** 31-42.
Kellner WA, Bell JSK & Vertino PM 2015. GC skew defines distinct RNA polymerase pause sites in CpG island promoters. *Genome Research,* 25**,** 1600-1609.
Kennison JA 1995. The Polycomb and Trithorax Group Proteins of Drosophila: Trans-Regulators of Homeotic Gene Function. *Annual Review of Genetics,* 29**,** 289-303.
Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler & David 2002. The Human Genome Browser at UCSC. *Genome Research,* 12**,** 996-1006.
Keshet I, Lieman-Hurwitz J & Cedar H 1986. DNA methylation affects the formation of active chromatin. *Cell,* 44**,** 535-543.
Kidder BL, Palmer S & Knott JG 2009. SWI/SNF-Brg1 Regulates Self-Renewal and Occupies Core Pluripotency-Related Genes in Embryonic Stem Cells. *Stem Cells,* 27**,** 317-328.
Kim D-H, Tang Z, Shimada M, Fierz B, Houck-Loomis B, Bar-Dagen M, Lee S, Lee S-K, Muir TW, Roeder RG*, et al.* 2013a. Histone H3K27 Trimethylation Inhibits H3 Binding and Function of SET1-Like H3K4 Methyltransferase Complexes. *Molecular and Cellular Biology,* 33**,** 4936-4946.
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R & Salzberg SL 2013b. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology,* 14**,** 1-13.
Kim JB, Greber B, Arauzo-Bravo MJ, Meyer J, Park KI, Zaehres H & Scholer HR 2009. Direct reprogramming of human neural stem cells by OCT4. *Nature,* 461**,** 649-653.
Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S*, et al.* 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature,* 465**,** 182-187.
King IFG, Emmons RB, Francis NJ, Wild B, Müller J, Kingston RE & Wu C-t 2005. Analysis of a Polycomb Group Protein Defines Regions That Link Repressive Activity on Nucleosomal Templates to In Vivo Function. *Molecular and Cellular Biology,* 25**,** 6578-6591.
Kireeva ML, Hancock B, Cremona GH, Walter W, Studitsky VM & Kashlev M 2005. Nature of the Nucleosomal Barrier to RNA Polymerase II. *Molecular Cell,* 18**,** 97-108.
Klemm JD, Rould MA, Aurora R, Herr W & Pabo CO 1994. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell,* 77**,** 21-32.
Kloet SL, Makowski MM, Baymaz HI, van Voorthuijsen L, Karemaker ID, Santanach A, Jansen PWTC, Di Croce L & Vermeulen M 2016. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol,* 23**,** 682-690.
Klose RJ & Bird AP 2006. Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences,* 31**,** 89-97.
Klose RJ, Cooper S, Farcas AM, Blackledge NP & Brockdorff N 2013. Chromatin Sampling—An Emerging Perspective on Targeting Polycomb Repressor Proteins. *PLoS Genet,* 9**,** e1003717.
Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, Bernstein BE & Meissner A 2011. Reprogramming Factor Expression Initiates Widespread Targeted Chromatin Remodeling. *Cell Stem Cell,* 8**,** 96-105.
Koohy H, Down TA & Hubbard TJ 2013. Chromatin Accessibility Data Sets Show Bias Due to Sequence Specificity of the DNase I Enzyme. *PLoS ONE,* 8**,** e69853.
Kornberg RD & Lorch Y 1999. Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. *Cell,* 98**,** 285-294.
Kouzarides T 2007. Chromatin Modifications and Their Function. *Cell,* 128**,** 693-705.
Krebs AR, Dessus-Babus S, Burger L & Schübeler D 2014. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife,* 3**,** e04094.
Krishnakumar R & Kraus WL 2010. PARP-1 Regulates Chromatin Structure and Transcription through a KDM5B-Dependent Pathway. *Molecular Cell,* 39**,** 736-749.
Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, Richards DP, Beattie BK, Emili A, Boone C*, et al.* 2003. Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. *Molecular and Cellular Biology,* 23**,** 4207-4218.

Krueger F & Andrews SR 2011. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics,* 27**,** 1571-1572.
Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS*, et al.* 2008. Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *PLoS Genet,* 4**,** e1000242.
Kumar S & Subramanian S 2002. Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences,* 99**,** 803-808.
Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, 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 & Salzberg SL 2012. Fast gapped-read alignment with Bowtie 2. *Nat Meth,* 9**,** 357-359.
Laurette P, Strub T, Koludrovic D, Keime C, Le Gras S, Seberg H, Van Otterloo E, Imrichova H, Siddaway R, Aerts S*, et al.* 2015. Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. *eLife,* 4**,** e06857.
Law JA & Jacobsen SE 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet,* 11**,** 204-220.
Lay FD, Liu Y, Kelly TK, Witt H, Farnham PJ, Jones PA & Berman BP 2015. The role of DNA methylation in directing the functional organization of the cancer epigenome. *Genome Research*.
Lazarovici A, Zhou T, Shafer A, Dantas Machado AC, Riley TR, Sandstrom R, Sabo PJ, Lu Y, Rohs R, Stamatoyannopoulos JA*, et al.* 2013. Probing DNA shape and methylation state on a genomic scale with DNase I. *Proceedings of the National Academy of Sciences,* 110**,** 6376-6381.
Lee DY, Hayes JJ, Pruss D & Wolffe AP 1993. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell,* 72**,** 73-84.
Lee J-H & Skalnik DG 2005. CpG-binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. *Journal of Biological Chemistry,* 280**,** 41725-41731.
Lee J-H & Skalnik DG 2008. Wdr82 Is a C-Terminal Domain-Binding Protein That Recruits the Setd1A Histone H3-Lys4 Methyltransferase Complex to Transcription Start Sites of Transcribed Human Genes. *Molecular and Cellular Biology,* 28**,** 609-618.
Lee J-H, Tate CM, You J-S & Skalnik DG 2007a. Identification and Characterization of the Human Set1B Histone H3-Lys4 Methyltransferase Complex. *Journal of Biological Chemistry,* 282**,** 13419-13428.
Lee JY & Lee T-H 2012. Effects of DNA Methylation on the Structure of Nucleosomes. *Journal of the American Chemical Society,* 134**,** 173-175.
Lee MG, Villa R, Trojer P, Norman J, Yan K-P, Reinberg D, Di Croce L & Shiekhattar R 2007b. Demethylation of H3K27 Regulates Polycomb Recruitment and H2A Ubiquitination. *Science,* 318**,** 447-450.
Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR & Nislow C 2007c. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet,* 39**,** 1235-1244.
Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K & Wutz A 2010. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & Development,* 24**,** 265-276.
Lennartsson A, Arner E, Fagiolini M, Saxena A, Andersson R, Takahashi H, Noro Y, Sng J, Sandelin A, Hensch TK*, et al.* 2015. Remodeling of retrotransposon elements during epigenetic induction of adult visual cortical plasticity by HDAC inhibitors. *Epigenetics & Chromatin,* 8**,** 1-15.
Levasseur DN, Wang J, Dorschner MO, Stamatoyannopoulos JA & Orkin SH 2008. Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes & Development,* 22**,** 575-580.
Lewis EB 1978. A gene complex controlling segmentation in Drosophila. *Nature,* 276**,** 565-570.
Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F & Bird A 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. *Cell,* 69**,** 905-914.
Li B, Carey M & Workman JL 2007. The Role of Chromatin during Transcription. *Cell,* 128**,** 707-719.
Li B, Jackson J, Simon MD, Fleharty B, Gogol M, Seidel C, Workman JL & Shilatifard A 2009a. Histone H3 Lysine 36 Dimethylation (H3K36me2) Is Sufficient to Recruit the Rpd3s Histone

Deacetylase Complex and to Repress Spurious Transcription. *Journal of Biological Chemistry,* 284**,** 7970-7976.
Li E, Bestor TH & Jaenisch R 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69**,** 915-926.
Li G & Widom J 2004. Nucleosomes facilitate their own invasion. *Nat Struct Mol Biol,* 11**,** 763-769.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R & Subgroup GPDP 2009b. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.
Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD & Patel DJ 2006. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature,* 442**,** 91-95.
Li Y, Schulz VP, Deng C, Li G, Shen Y, Tusi BK, Ma G, Stees J, Qiu Y, Steiner LA*, et al.* 2016. Setd1a and NURF mediate chromatin dynamics and gene regulation during erythroid lineage commitment and differentiation. *Nucleic Acids Research,* 44**,** 7173-7188.
Li Y, Zhang Q, Yin X, Yang W, Du Y, Hou P, Ge J, Liu C, Zhang W, Zhang X*, et al.* 2011a. Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. *Cell Res,* 21**,** 196-204.
Li Z, Schug J, Tuteja G, White P & Kaestner KH 2011b. The nucleosome map of the mammalian liver. *Nat Struct Mol Biol,* 18**,** 742-746.
Liang G, He J & Zhang Y 2012. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nat Cell Biol,* 14**,** 457-466.
Lickert H, Takeuchi JK, von Both I, Walls JR, McAuliffe F, Lee Adamson S, Mark Henkelman R, Wrana JL, Rossant J & Bruneau BG 2004. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature,* 432**,** 107-112.
Lienert F, Wirbelauer C, Som I, Dean A, Mohn F & Schubeler D 2011. Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet,* 43**,** 1091-1097.
Linder B, Mentele E, Mansperger K, Straub T, Kremmer E & Rupp RAW 2007. CHD4/Mi-2β activity is required for the positioning of the mesoderm/neuroectoderm boundary in Xenopus. *Genes & Development,* 21**,** 973-983.
Ling G, Sugathan A, Mazor T, Fraenkel E & Waxman DJ 2010. Unbiased, Genome-Wide In Vivo Mapping of Transcriptional Regulatory Elements Reveals Sex Differences in Chromatin Structure Associated with Sex-Specific Liver Gene Expression. *Molecular and Cellular Biology,* 30**,** 5531-5544.
Llères D, James J, Swift S, Norman DG & Lamond AI 2009. Quantitative analysis of chromatin compaction in living cells using FLIM–FRET. *The Journal of Cell Biology,* 187**,** 481-496.
Lock LF, Takagi N & Martin GR 1987. Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell,* 48**,** 39-46.
Long HK, Blackledge NP & Klose RJ 2013a. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society Transactions,* 41**,** 727-740.
Long HK, King HW, Patient RK, Odom DT & Klose RJ 2016. Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Research,* 44**,** 6693-6706.
Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grützner F, Odom DT, Patient R, Ponting CP*, et al.* 2013b. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *eLife,* 2**,** e00348.
Love MI, Huber W & Anders S 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15**,** 1-21.
Lu F, Liu Y, Inoue A, Suzuki T, Zhao K & Zhang Y 2016. Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell,* 165**,** 1375-1388.
Ludwig AK, Zhang P & Cardoso MC 2016. Modifiers and Readers of DNA Modifications and Their Impact on Genome Structure, Expression, and Stability in Disease. *Frontiers in Genetics,* 7.
Lunter G & Hein J 2004. A nucleotide substitution model with nearest-neighbour interactions. *Bioinformatics,* 20**,** i216-i223.
Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze S-H, Chenoweth J, Schwartz P, Pevzner PA, Glass C, Mandel G*, et al.* 2002. Corepressor-Dependent Silencing of Chromosomal Regions Encoding Neuronal Genes. *Science,* 298**,** 1747-1752.

Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS & Brown M 2008. FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription. *Cell,* 132**,** 958-970.

Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F*, et al.* 2013. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci,* 16**,** 898-902.

Macleod D, Charlton J, Mullins J & Bird AP 1994. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes & Development,* 8**,** 2282-2292.

Magnani L, Eeckhoute J & Lupien M 2011. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends in Genetics,* 27**,** 465-474.

Mancini DN, Singh SM, Archer TK & Rodenhiser DI 1999. Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. *Oncogene,* 18**,** 4108-4119.

Marathe HG, Mehta G, Zhang X, Datar I, Mehrotra A, Yeung KC & de la Serna IL 2013. SWI/SNF Enzymes Promote SOX10- Mediated Activation of Myelin Gene Expression. *PLoS ONE,* 8**,** e69037.

Marfella CGA & Imbalzano AN 2007. The Chd family of chromatin remodelers. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 618**,** 30-40.

Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD & Reinberg D 2008. Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Molecular Cell,* 32**,** 503-518.

Marmorstein R & Berger SL 2001. Structure and function of bromodomains in chromatin-regulating complexes. *Gene,* 272**,** 1-9.

Marsman J & Horsfield JA 2012. Long distance relationships: Enhancer–promoter communication and dynamic gene transcription. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1819**,** 1217-1227.

Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC & Shinkai Y 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature,* 464**,** 927-931.

Maurano Matthew T, Wang H, John S, Shafer A, Canfield T, Lee K & Stamatoyannopoulos John A 2015. Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Reports,* 12**,** 1184-1195.

McArthur M, Gerum S & Stamatoyannopoulos G 2001. Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse β-globin LCR1. *Journal of Molecular Biology,* 313**,** 27-34.

McCall K & Bender W 1996. Probes of chromatin accessibility in the Drosophila bithorax complex respond differently to Polycomb-mediated repression. *The EMBO Journal,* 15**,** 569-580.

McCall M, Brown T & Kennard O 1985. The crystal structure of d(G-G-G-G-C-C-C-C) a model for poly(dG) · poly(dC). *Journal of Molecular Biology,* 183**,** 385-396.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM & Bejerano G 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotech,* 28**,** 495-501.

McNally JG, Müller WG, Walker D, Wolford R & Hager GL 2000. The Glucocorticoid Receptor: Rapid Exchange with Regulatory Sites in Living Cells. *Science,* 287**,** 1262-1265.

Medvedeva YA, Khamis AM, Kulakovskiy IV, Ba-Alawi W, Bhuyan MSI, Kawaji H, Lassmann T, Harbers M, Forrest AR & Bajic VB 2014. Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics,* 15**,** 1-12.

Meehan RR, Lewis JD, McKay S, Kleiner EL & Bird AP 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell,* 58**,** 499-507.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB*, et al.* 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature,* 454**,** 766-770.

Melcer S, Hezroni H, Rand E, Nissim-Rafinia M, Skoultchi A, Stewart CL, Bustin M & Meshorer E 2012. Histone modifications and lamin A regulate chromatin protein dynamics in early embryonic stem cell differentiation. *Nature Communications,* 3**,** 910.

Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M & Bernstein BE 2010. GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. *PLoS Genet,* 6**,** e1001244.

Meyer C, Kowarz E, Hofmann J, Renneville A, Zuna J, Trka J, Ben Abdelali R, Macintyre E, De Braekeleer E, De Braekeleer M*, et al.* 2009. New insights to the MLL recombinome of acute leukemias. *Leukemia,* 23**,** 1490-1499.

Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA, Larschan E, Park PJ*, et al.* 2016. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nature Communications,* 7**,** 11485.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP*, et al.* 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature,* 448**,** 553-560.

Miller JA & Widom J 2003. Collaborative Competition Mechanism for Gene Activation In Vivo. *Molecular and Cellular Biology,* 23**,** 1623-1632.

Milne TA, Dou Y, Martin ME, Brock HW, Roeder RG & Hess JL 2005. MLL associates specifically with a subset of transcriptionally active target genes. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 14765-14770.

Milne TA, Kim J, Wang GG, Stadler SC, Basrur V, Whitcomb SJ, Wang Z, Ruthenburg AJ, Elenitoba-Johnson KSJ, Roeder RG*, et al.* 2010. Multiple Interactions Recruit MLL1 and MLL1 Fusion Proteins to the HOXA9 Locus in Leukemogenesis. *Molecular Cell,* 38**,** 853-863.

Min J, Zhang Y & Xu R-M 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & Development,* 17**,** 1823-1828.

Mirny LA 2010. Nucleosome-mediated cooperativity between transcription factors. *Proceedings of the National Academy of Sciences,* 107**,** 22534-22539.

Mohan M, Lin C, Guest E & Shilatifard A 2010. Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nat Rev Cancer,* 10**,** 721-728.

Mohandas T, Sparkes R & Shapiro L 1981. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science,* 211**,** 393-396.

Mohrmann L & Verrijzer CP 2005. Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 1681**,** 59-73.

Morettini S, Tribus M, Zeilner A, Sebald J, Campo-Fernandez B, Scheran G, Wörle H, Podhraski V, Fyodorov DV & Lusser A 2011. The chromodomains of CHD1 are critical for enzymatic activity but less important for chromatin localization. *Nucleic Acids Research,* 39**,** 3103-3115.

Morey L, Brenner C, Fazi F, Villa R, Gutierrez A, Buschbeck M, Nervi C, Minucci S, Fuks F & Di Croce L 2008. MBD3, a Component of the NuRD Complex, Facilitates Chromatin Alteration and Deposition of Epigenetic Marks. *Molecular and Cellular Biology,* 28**,** 5912-5923.

Morey L, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah Salvador A & Di Croce L 2012. Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. *Cell Stem Cell,* 10**,** 47-62.

Morris SA, Baek S, Sung M-H, John S, Wiench M, Johnson TA, Schiltz RL & Hager GL 2014. Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nature Structural Molecular Biology,* 21**,** 73-81.

Morselli M, Pastor WA, Montanini B, Nee K, Ferrari R, Fu K, Bonora G, Rubbi L, Clark AT, Ottonello S*, et al.* 2015. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife,* 4**,** e06205.

Mucha M, Lisowska K, Goc A & Filipski J 2000. Nuclease-hypersensitive Chromatin Formed by a CpG Island in Human DNA Cloned as an Artificial Chromosome in Yeast. *Journal of Biological Chemistry,* 275**,** 1275-1278.

Mueller D, García-Cuéllar M-P, Bach C, Buhl S, Maethner E & Slany RK 2009. Misguided Transcriptional Elongation Causes Mixed Lineage Leukemia. *PLoS Biol,* 7**,** e1000249.

Muntean AG, Tan J, Sitwala K, Huang Y, Bronstein J, Connelly JA, Basrur V, Elenitoba-Johnson KSJ & Hess JL 2010. The PAF Complex Synergizes with MLL Fusion Proteins at HOX Loci to Promote Leukemogenesis. *Cancer Cell,* 17**,** 609-621.

Musselman CA, Ramírez J, Sims JK, Mansfield RE, Oliver SS, Denu JM, Mackay JP, Wade PA, Hagman J & Kutateladze TG 2012. Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proceedings of the National Academy of Sciences,* 109**,** 787-792.

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N & Yamanaka S 2008. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotech,* 26**,** 101-106.

Nakayama J-i, Rice JC, Strahl BD, Allis CD & Grewal SIS 2001. Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science,* 292**,** 110-113.

Nan X, Campoy FJ & Bird A 1997. MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. *Cell,* 88**,** 471-481.

Nan X, Ng H-H, Johnson CA, Laherty CD, Turner BM, Eisenman RN & Bird A 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature,* 393**,** 386-389.

Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, Vonrhein C, Moras D, Romier C, Bolognesi M*, et al.* 2013. Sequence-Specific Transcription Factor NF-Y Displays Histone-like DNA Binding and H2B-like Ubiquitination. *Cell,* 152**,** 132-143.

Narlikar G, Sundaramoorthy R & Owen-Hughes T 2013. Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell,* 154**,** 490-503.

Narlikar GJ, Phelan ML & Kingston RE 2001. Generation and Interconversion of Multiple Distinct Nucleosomal States as a Mechanism for Catalyzing Chromatin Fluidity. *Molecular Cell,* 8**,** 1219-1230.

Natarajan A, Yardımcı GG, Sheffield NC, Crawford GE & Ohler U 2012. Predicting cell-type–specific gene expression from regions of open chromatin. *Genome Research,* 22**,** 1711-1722.

Ng CKL, Li NX, Chee S, Prabhakar S, Kolatkar PR & Jauch R 2012. Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. *Nucleic Acids Research*.

Ng H-H, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D & Bird A 1999. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet,* 23**,** 58-61.

Ng HH, Robert F, Young RA & Struhl K 2003. Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. *Molecular Cell,* 11**,** 709-719.

Niwa H, Miyazaki J-i & Smith AG 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet,* 24**,** 372-376.

Noh K-M, Wang H, Kim Hyunjae R, Wenderski W, Fang F, Li Charles H, Dewell S, Hughes Stephen H, Melnick Ari M, Patel Dinshaw J*, et al.* 2015. Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs. *Molecular Cell,* 59**,** 89-103.

Okano M, Bell DW, Haber DA & Li E 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell,* 99**,** 247-257.

Olave I, Wang W, Xue Y, Kuo A & Crabtree GR 2002. Identification of a polymorphic, neuron-specific chromatin remodeling complex. *Genes & Development,* 16**,** 2509-2517.

Oldfield Andrew J, Yang P, Conway Amanda E, Cinghu S, Freudenberg Johannes M, Yellaboina S & Jothi R 2014. Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors. *Molecular Cell,* 55**,** 708-722.

Ooi L, Belyaev ND, Miyake K, Wood IC & Buckley NJ 2006. BRG1 Chromatin Remodeling Activity Is Required for Efficient Chromatin Binding by Repressor Element 1-silencing Transcription Factor (REST) and Facilitates REST-mediated Repression. *Journal of Biological Chemistry,* 281**,** 38974-38980.

Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin S-P, Allis CD*, et al.* 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature,* 448**,** 714-717.

Orlando D, Chen M, Brown V, Solanki S, Choi Y, Olson E, Fritz C, Bradner J & Guenther M 2014. Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome. *Cell Reports,* 9**,** 1163-1170.

Pajoro A, Madrigal P, Muiño JM, Matus JT, Jin J, Mecchia MA, Debernardi JM, Palatnik JF, Balazadeh S, Arif M*, et al.* 2014. Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biology,* 15**,** 1-19.

Palacios D, Summerbell D, Rigby PWJ & Boyes J 2010. Interplay between DNA Methylation and Transcription Factor Availability: Implications for Developmental Activation of the Mouse Myogenin Gene. *Molecular and Cellular Biology,* 30**,** 3805-3815.

Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM & Choudhary J 2010. An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. *Cell Stem Cell,* 6**,** 382-395.

Pattenden SG, Simon JM, Wali A, Jayakody CN, Troutman J, McFadden AW, Wooten J, Wood CC, Frye SV, Janzen WP*, et al.* 2016. High-throughput small molecule screen identifies inhibitors of aberrant chromatin accessibility. *Proceedings of the National Academy of Sciences,* 113**,** 3018-3023.

Peters AHFM, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AAHA, Perez-Burgos L, Kohlmaier A, Opravil S, Tachibana M, Shinkai Y*, et al.* 2003. Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. *Molecular Cell,* 12**,** 1577-1589.

Peterson CL & Laniel M-A 2004. Histones and histone modifications. *Current Biology,* 14**,** R546-R551.

Pfau R, Tzatsos A, Kampranis SC, Serebrennikova OB, Bear SE & Tsichlis PN 2008. Members of a family of JmjC domain-containing oncoproteins immortalize embryonic fibroblasts via a JmjC domain-dependent process. *Proceedings of the National Academy of Sciences,* 105**,** 1907-1912.

Picelli S, Björklund ÅK, Reinius B, Sagasser S, Winberg G & Sandberg R 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research,* 24**,** 2033-2040.

Polach KJ & Widom J 1995. Mechanism of Protein Access to Specific DNA Sequences in Chromatin: A Dynamic Equilibrium Model for Gene Regulation. *Journal of Molecular Biology,* 254**,** 130-149.

Polach KJ & Widom J 1996. A Model for the Cooperative Binding of Eukaryotic Regulatory Proteins to Nucleosomal Target Sites. *Journal of Molecular Biology,* 258**,** 800-812.

Portela A & Esteller M 2010. Epigenetic modifications and human disease. *Nat Biotech,* 28**,** 1057-1068.

Pray-Grant MG, Daniel JA, Schieltz D, Yates JR & Grant PA 2005. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature,* 433**,** 434-438.

Prokhortchouk A, Hendrich B, Jørgensen H, Ruzov A, Wilm M, Georgiev G, Bird A & Prokhortchouk E 2001. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & Development,* 15**,** 1613-1618.

Qin J, Whyte Warren A, Anderssen E, Apostolou E, Chen H-H, Akbarian S, Bronson Roderick T, Hochedlinger K, Ramaswamy S, Young Richard A*, et al.* 2012. The Polycomb Group Protein L3mbtl2 Assembles an Atypical PRC1-Family Complex that Is Essential in Pluripotent Stem Cells and Early Development. *Cell Stem Cell,* 11**,** 319-332.

Quinlan AR & Hall IM 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N*, et al.* 2011. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotech,* 29**,** 436-442.

Radman-Livaja M & Rando OJ 2010. Nucleosome positioning: How is it established, and why does it matter? *Developmental Biology,* 339**,** 258-266.

Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M & Smale ST 2009. A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. *Cell,* 138**,** 114-128.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F & Manke T 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research,* 44**,** W160-5.

Ramírez J, Dege C, Kutateladze TG & Hagman J 2012. MBD2 and Multiple Domains of CHD4 Are Required for Transcriptional Repression by Mi-2/NuRD Complexes. *Molecular and Cellular Biology,* 32**,** 5078-5088.

Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, Nguyen N, Paten B, Zweig AS, Karolchik D*, et al.* 2014. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics,* 30**,** 1003-1005.

Raposo A, Vasconcelos F, Drechsel D, Marie C, Johnston C, Dolle D, Bithell A, Gillotin S, van den Berg D, Ettwiller L*, et al.* 2015. Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis. *Cell Reports,* 10**,** 1544-1556.

Razin A & Cedar H 1977. Distribution of 5-methylcytosine in chromatin. *Proc Natl Acad Sci U S A.,* 74**,** 2725-2728.

Reik W, Dean W & Walter J 2001. Epigenetic Reprogramming in Mammalian Development. *Science,* 293**,** 1089-1093.
Reményi A, Tomilin A, Pohl E, Lins K, Philippsen A, Reinbold R, Schöler HR & Wilmanns M 2001. Differential Dimer Activities of the Transcription Factor Oct-1 by DNA-Induced Interface Swapping. *Molecular Cell,* 8**,** 569-580.
Reynolds N, Salmon-Divon M, Dvinge H, Hynes-Allen A, Balasooriya G, Leaford D, Behrens A, Bertone P & Hendrich B 2012. NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *The EMBO Journal,* 31**,** 593-605.
Reznikoff WS 2008. Transposon Tn5. *Annual Review of Genetics,* 42**,** 269-286.
Riising Eva M, Comet I, Leblanc B, Wu X, Johansen Jens V & Helin K 2014. Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Molecular Cell,* 55**,** 347-360.
Rincon-Arano H, Halow J, Delrow Jeffrey J, Parkhurst Susan M & Groudine M 2012. UpSET Recruits HDAC Complexes and Restricts Chromatin Accessibility and Acetylation at Promoter Regions. *Cell,* 151**,** 1214-1228.
Roberts S & Gavery M 2012. Is There a Relationship between DNA Methylation and Phenotypic Plasticity in Invertebrates? *Frontiers in Physiology,* 2.
Robertson KD & Wolffe AP 2000. DNA methylation in health and disease. *Nat Rev Genet,* 1**,** 11-19.
Robin P, Fritsch L, Philipot O, Svinarchuk F & Ait-Si-Ali S 2007. Post-translational modifications of histones H3 and H4 associated with the histone methyltransferases Suv39h1 and G9a. *Genome Biology,* 8**,** 1-10.
Rose NR, King HW, Blackledge NP, Fursova NA, Ember KJI, Fischer R, Kessler BM & Klose RJ 2016. RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. *eLife,* 5**,** e18591.
Rose NR & Klose RJ 2014. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1839**,** 1362-1372.
Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, Aktas T & Trono D 2013. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development,* 140**,** 519-529.
Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J*, et al.* 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature,* 463**,** 237-240.
Ruthenburg AJ, Allis CD & Wysocka J 2007. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Molecular Cell,* 25**,** 15-30.
Sabo PJ, Humbert R, Hawrylycz M, Wallace JC, Dorschner MO, McArthur M & Stamatoyannopoulos JA 2004. Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 4537-4542.
Saldanha AJ 2004. Java Treeview—extensible visualization of microarray data. *Bioinformatics,* 20**,** 3246-3248.
Santos-Rosa H, Schneider R, Bernstein BE, Karabetsou N, Morillon A, Weise C, Schreiber SL, Mellor J & Kouzarides T 2003. Methylation of Histone H3 K4 Mediates Association of the Isw1p ATPase with Chromatin. *Molecular Cell,* 12**,** 1325-1332.
Saxonov S, Berg P & Brutlag DL 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences,* 103**,** 1412-1417.
Scacheri PC, Davis S, Odom DT, Crawford GE, Perkins S, Halawi MJ, Agarwal SK, Marx SJ, Spiegel AM, Meltzer PS*, et al.* 2006. Genome-Wide Analysis of Menin Binding Provides Insights into MEN1 Tumorigenesis. *PLoS Genet,* 2**,** e51.
Schmerwitz UK, Sass G, Khandoga AG, Joore J, Mayer BA, Berberich N, Totzke F, Krombach F, Tiegs G, Zahler S*, et al.* 2011. Flavopiridol Protects Against Inflammation by Attenuating Leukocyte-Endothelial Interaction via Inhibition of Cyclin-Dependent Kinase 9. *Arteriosclerosis, Thrombosis, and Vascular Biology,* 31**,** 280-288.
Schmidt D, Schwalie Petra C, Wilson Michael D, Ballester B, Gonçalves Â, Kutter C, Brown Gordon D, Marshall A, Flicek P & Odom Duncan T 2012. Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. *Cell,* 148**,** 335-348.

Schmitges F, Prusty A, Faty M, Stützer A, Lingaraju G, Aiwazian J, Sack R, Hess D, Li L, Zhou S*, et al.* 2011. Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Molecular Cell,* 42**,** 330-341.
Schoenfelder S, Sugar R, Dimond A, Javierre B-M, Armstrong H, Mifsud B, Dimitrova E, Matheson L, Tavares-Cadete F, Furlan-Magaril M*, et al.* 2015. Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet,* 47**,** 1179-1186.
Schones DE, Cui K, Cuddapah S, Roh T-Y, Barski A, Wang Z, Wei G & Zhao K 2008. Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell,* 132**,** 887-898.
Schubeler D 2015. Function and information content of DNA methylation. *Nature,* 517**,** 321-326.
Schuettengruber B, Chourrout D, Vervoort M, Leblanc B & Cavalli G 2007. Genome Regulation by Polycomb and Trithorax Proteins. *Cell,* 128**,** 735-745.
Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, McKay DJ & Harrison MM 2015. Zelda is differentially required for chromatin accessibility, transcription-factor binding and gene expression in the early Drosophila embryo. *Genome Research,* 25**,** 1715-1726.
Schwaiger M, Schönauer A, Rendeiro AF, Pribitzer C, Schauer A, Gilles AF, Schinko JB, Renfer E, Fredman D & Technau U 2014. Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Research,* 24**,** 639-650.
Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W & Selbach M 2011. Global quantification of mammalian gene expression control. *Nature,* 473**,** 337-342.
Schwartz YB & Pirrotta V 2013. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat Rev Genet,* 14**,** 853-864.
Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang J-PZ & Widom J 2006. A genomic code for nucleosome positioning. *Nature,* 442**,** 772-778.
Segal E & Widom J 2009. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Current Opinion in Structural Biology,* 19**,** 65-71.
Sekinger EA, Moqtaderi Z & Struhl K 2005. Intrinsic Histone-DNA Interactions and Low Nucleosome Density Are Important for Preferential Accessibility of Promoter Regions in Yeast. *Molecular Cell,* 18**,** 735-748.
Shakya A, Callister C, Goren A, Yosef N, Garg N, Khoddami V, Nix D, Regev A & Tantin D 2015. Pluripotency Transcription Factor Oct4 Mediates Stepwise Nucleosome Demethylation and Depletion. *Molecular and Cellular Biology,* 35**,** 1014-1025.
Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, Crawford GE & Furey TS 2013. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Research,* 23**,** 777-788.
Shen J-C, Rideout WM & Jones PA 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Research,* 22**,** 972-976.
Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, Shu J, Chen X, Waterland RA & Issa J-PJ 2007. Genome-Wide Profiling of DNA Methylation Reveals a Class of Normally Methylated CpG Island Promoters. *PLoS Genet,* 3**,** e181.
Shlyueva D, Stampfel G & Stark A 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet,* 15**,** 272-286.
Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR & Peterson CL 2006. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science,* 311**,** 844-847.
Siggens L, Cordeddu L, Rönnerblad M, Lennartsson A & Ekwall K 2015. Transcription-coupled recruitment of human CHD1 and CHD2 influences chromatin accessibility and histone H3 and H3.3 occupancy at active chromatin regions. *Epigenetics & Chromatin,* 8**,** 1-14.
Simandi Z, Horvath A, Wright Lyndsey C, Cuaranta-Monroy I, De Luca I, Karolyi K, Sauer S, Deleuze J-F, Gudas Lorraine J, Cowley Shaun M*, et al.* 2016. OCT4 Acts as an Integrator of Pluripotency and Signal-Induced Differentiation. *Molecular Cell,* 63**,** 647-661.
Simic R, Lindstrom DL, Tran HG, Roinick KL, Costa PJ, Johnson AD, Hartzog GA & Arndt KM 2003. Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *The EMBO Journal,* 22**,** 1846-1856.
Simon JA & Kingston RE 2009. Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol,* 10**,** 697-708.
Simon Jeffrey A & Kingston Robert E 2013. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Molecular Cell,* 49**,** 808-824.

Simon JA & Tamkun JW 2002. Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. *Current Opinion in Genetics & Development,* 12**,** 210-218.
Simon JM, Giresi PG, Davis IJ & Lieb JD 2012. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat. Protocols,* 7**,** 256-267.
Sims RJ, Chen C-F, Santos-Rosa H, Kouzarides T, Patel SS & Reinberg D 2005. Human but Not Yeast CHD1 Binds Directly and Selectively to Histone H3 Methylated at Lysine 4 via Its Tandem Chromodomains. *Journal of Biological Chemistry,* 280**,** 41789-41792.
Singhal N, Esch D, Stehling M & Schöler HR 2014. BRG1 Is Required to Maintain Pluripotency of Murine Embryonic Stem Cells. *BioResearch Open Access,* 3**,** 1-8.
Singhal N, Graumann J, Wu G, Araúzo-Bravo MJ, Han DW, Greber B, Gentile L, Mann M & Schöler HR 2010. Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. *Cell,* 141**,** 943-955.
Skene PJ, Hernandez AE, Groudine M & Henikoff S 2014. The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *eLife,* 3**,** e02042.
Smale ST, Plevy SE, Weinmann AS, Zhou L, Ramirez-Carrozzi VR, Pope SD, Bhatt DM & Tong A-J 2013. Toward an Understanding of the Gene-Specific and Global Logic of Inducible Gene Transcription. *Cold Spring Harbor Symposia on Quantitative Biology,* 78**,** 61-68.
Smith ZD & Meissner A 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet,* 14**,** 204-220.
Smith ZD, Sindhu C & Meissner A 2016. Molecular features of cellular reprogramming and development. *Nat Rev Mol Cell Biol,* 17**,** 139-154.
Song L & Crawford GE 2010. DNase-seq: A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome from Mammalian Cells. *Cold Spring Harbor Protocols,* 2010**,** 1-11.
Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee B-K, Sheffield NC, Gräf S, Huss M, Keefe D*, et al.* 2011. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Research,* 21**,** 1757-1767.
Soufi A, Donahue G & Zaret Kenneth S 2012. Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell,* 151**,** 994-1004.
Soufi A, Garcia Meilin F, Jaroszewicz A, Osman N, Pellegrini M & Zaret Kenneth S 2015. Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell,* 161**,** 555-568.
Spitz F & Furlong EEM 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet,* 13**,** 613-626.
Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK*, et al.* 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature,* 480**,** 490-495.
Stalder J, Larsen A, Engel JD, Dolan M, Groudine M & Weintraub H 1980. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell,* 20**,** 451-460.
Stark R & Brown G (eds.) 2011. *DiffBind: differential binding analysis of ChIP-Seq peak data. ,* Bioconductor.
Steffen PA & Ringrose L 2014. What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat Rev Mol Cell Biol,* 15**,** 340-356.
Stein R, Gruenbaum Y, Pollack Y, Razin A & Cedar H 1982. Clonal inheritance of the pattern of DNA methylation in mouse cells. *Proc Natl Acad Sci U S A.,* 79**,** 61-65.
Struhl K 1999. Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. *Cell,* 98**,** 1-4.
Struhl K & Segal E 2013. Determinants of nucleosome positioning. *Nat Struct Mol Biol,* 20**,** 267-273.
Suetake I, Shinozaki F, Miyagawa J, Takeshima H & Tajima S 2004. DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. *Journal of Biological Chemistry,* 279**,** 27816-27823.
Sun B, Hong J, Zhang P, Dong X, Shen X, Lin D & Ding J 2008. Molecular Basis of the Interaction of Saccharomyces cerevisiae Eaf3 Chromo Domain with Methylated H3K36. *Journal of Biological Chemistry,* 283**,** 36504-36512.
Suter B, Schnappauf G & Thoma F 2000. Poly(dA·dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Research,* 28**,** 4083-4089.

Suzuki MM & Bird A 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet,* 9**,** 465-476.

Sved J & Bird A 1990. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proceedings of the National Academy of Sciences,* 87**,** 4692-4696.

Swinstead Erin E, Miranda Tina B, Paakinaho V, Baek S, Goldstein I, Hawkins M, Karpova Tatiana S, Ball D, Mazza D, Lavis Luke D*, et al.* 2016. Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell,* 165**,** 593-605.

Taberlay PC, Statham AL, Kelly TK, Clark SJ & Jones PA 2014. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome Research,* 24**,** 1421-1432.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L*, et al.* 2009. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science,* 324**,** 930-935.

Takahashi K & Yamanaka S 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell,* 126**,** 663-676.

Tapia N, MacCarthy C, Esch D, Gabriele Marthaler A, Tiemann U, Araúzo-Bravo MJ, Jauch R, Cojocaru V & Schöler HR 2015. Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Scientific Reports,* 5**,** 13533.

Tate CM, Lee J-H & Skalnik DG 2010. CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. *FEBS Journal,* 277**,** 210-223.

Tavares L, Dimitrova E, Oxley D, Webster J, Poot R, Demmers J, Bezstarosti K, Taylor S, Ura H, Koide H*, et al.* 2012. RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. *Cell,* 148**,** 664-678.

Taverna SD, Li H, Ruthenburg AJ, Allis CD & Patel DJ 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol,* 14**,** 1025-1040.

Tazi J & Bird A 1990. Alternative chromatin structure at CpG islands. *Cell,* 60**,** 909-920.

Teif VB, Vainshtein Y, Caudron-Herger M, Mallm J-P, Marth C, Höfer T & Rippe K 2012. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol,* 19**,** 1185-1192.

Thomas JH & Schneider S 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Research,* 21**,** 1800-1812.

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr ARW, Deaton A, Andrews R, James KD*, et al.* 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature,* 464**,** 1082-1086.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B*, et al.* 2012. The accessible chromatin landscape of the human genome. *Nature,* 489**,** 75-82.

Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, Diaz MO, Scacheri PC & Harte PJ 2009. CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development,* 136**,** 3131-3141.

Tippmann SC, Ivanek R, Gaidatzis D, Schöler A, Hoerner L, van Nimwegen E, Stadler PF, Stadler MB & Schübeler D 2012. Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels. *Molecular Systems Biology,* 8.

Titov DV, Gilman B, He Q-L, Bhat S, Low W-K, Dang Y, Smeaton M, Demain AL, Miller PS, Kugel JF*, et al.* 2011. XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat Chem Biol,* 7**,** 182-188.

Todeschini A-L, Georges A & Veitia RA 2014. Transcription factors: specific DNA binding and specific gene regulation. *Trends in Genetics,* 30**,** 211-219.

Tong JK, Hassig CA, Schnitzler GR, Kingston RE & Schreiber SL 1998. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature,* 395**,** 917-921.

Trojer P, Cao Alina R, Gao Z, Li Y, Zhang J, Xu X, Li G, Losson R, Erdjument-Bromage H, Tempst P*, et al.* 2011. L3MBTL2 Protein Acts in Concert with PcG Protein-Mediated Monoubiquitination of H2A to Establish a Repressive Chromatin Structure. *Molecular Cell,* 42**,** 438-450.

Trojer P, Li G, Sims Iii RJ, Vaquero A, Kalakonda N, Boccuni P, Lee D, Erdjument-Bromage H, Tempst P, Nimer SD*, et al.* 2007. L3MBTL1, a Histone-Methylation-Dependent Chromatin Lock. *Cell,* 129**,** 915-928.

Tsompana M & Buck MJ 2014. Chromatin accessibility: a window into the genome. *Epigenetics & Chromatin,* 7**,** 1-16.
Tsukada Y-i, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P & Zhang Y 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature,* 439**,** 811-816.
Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S-i, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR*, et al.* 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells,* 11**,** 805-814.
Tweedie S, Charlton J, Clark V & Bird A 1997. Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Molecular and Cellular Biology,* 17**,** 1469-75.
Ura H, Usuda M, Kinoshita K, Sun C, Mori K, Akagi T, Matsuda T, Koide H & Yokota T 2008. STAT3 and Oct-3/4 Control Histone Modification through Induction of Eed in Embryonic Stem Cells. *Journal of Biological Chemistry,* 283**,** 9713-9723.
van den Berg DLC, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I & Poot RA 2010. An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. *Cell Stem Cell,* 6**,** 369-381.
Vanyushin BF & Ashapkin VV 2011. DNA methylation in higher plants: Past, present and future. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1809**,** 360-368.
Vardimon L, Kressmann A, Cedar H, Maechler M & Doerfler W 1982. Expression of a cloned adenovirus gene is inhibited by in vitro methylation. *Proceedings of the National Academy of Sciences,* 79**,** 1073-1077.
Venkatesh S, Smolle M, Li H, Gogol MM, Saint M, Kumar S, Natarajan K & Workman JL 2012. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature,* 489**,** 452-455.
Verschure PJ, van der Kraan I, de Leeuw W, van der Vlag J, Carpenter AE, Belmont AS & van Driel R 2005. In Vivo HP1 Targeting Causes Large-Scale Chromatin Condensation and Enhanced Histone Lysine Methylation. *Molecular and Cellular Biology,* 25**,** 4552-4564.
Vierstra J, Wang H, John S, Sandstrom R & Stamatoyannopoulos JA 2014. Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nat Meth,* 11**,** 66-72.
Villar D, Flicek P & Odom DT 2014. Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nat Rev Genet,* 15**,** 221-233.
Voigt P, Tee W-W & Reinberg D 2013. A double take on bivalent promoters. *Genes & Development,* 27**,** 1318-1338.
Voo KS, Carlone DL, Jacobsen BM, Flodin A & Skalnik DG 2000. Cloning of a Mammalian Transcriptional Activator That Binds Unmethylated CpG Motifs and Shares a CXXC Domain with DNA Methyltransferase, Human Trithorax, and Methyl-CpG Binding Domain Protein 1. *Molecular and Cellular Biology,* 20**,** 2108-2121.
Voss T, Schiltz R, Sung M-H, Yen P, Stamatoyannopoulos J, Biddie S, Johnson T, Miranda T, John S & Hager G 2011. Dynamic Exchange at Regulatory Elements during Chromatin Remodeling Underlies Assisted Loading Mechanism. *Cell,* 146**,** 544-554.
Voss TC & Hager GL 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet,* 15**,** 69-81.
Wachter E, Quante T, Merusi C, Arczewska A, Stewart F, Webb S & Bird A 2014. Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. *eLife,* 3**,** e03397.
Wade PA, Jones PL, Vermaak D & Wolffe AP 1998. A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. *Current Biology,* 8**,** 843-848.
Waechter DE & Baserga R 1982. Effect of Methylation on Expression of Microinjected Genes. *Proceedings of the National Academy of Sciences of the United States of America,* 79**,** 1106-1110.
Walter M, Teissandier A, Pérez-Palacios R & Bourc'his D 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife,* 5**,** e11418.
Wan M, Liang J, Xiong Y, Shi F, Zhang Y, Lu W, He Q, Yang D, Chen R, Liu D*, et al.* 2013. The Trithorax Group Protein Ash2l Is Essential for Pluripotency and Maintaining Open Chromatin in Embryonic Stem Cells. *Journal of Biological Chemistry,* 288**,** 5039-5048.
Wang J, Liu X, Ni P, Gu Z & Fan Q 2010. SP1 is required for basal activation and chromatin accessibility of CD151 promoter in liver cancer cells. *Biochemical and Biophysical Research Communications,* 393**,** 291-296.

Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y*, et al.* 2012. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Research,* 22**,** 1798-1812.

Wang L, Brown JL, Cao R, Zhang Y, Kassis JA & Jones RS 2004. Hierarchical Recruitment of Polycomb Group Silencing Complexes. *Molecular Cell,* 14**,** 637-646.

Wang L, Du Y, Ward J, Shimbo T, Lackford B, Zheng X, Miao Y, Zhou B, Han L, Fargo D*, et al.* 2014. INO80 Facilitates Pluripotency Gene Activation in Embryonic Stem Cell Self-Renewal, Reprogramming, and Blastocyst Development. *Cell Stem Cell,* 14**,** 575-591.

Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK & Haussler D 2007. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proceedings of the National Academy of Sciences,* 104**,** 18613-18618.

Wang W, Xue Y, Zhou S, Kuo A, Cairns BR & Crabtree GR 1996. Diversity and specialization of mammalian SWI/SNF complexes. *Genes & Development,* 10**,** 2117-2130.

Wapinski Orly L, Vierbuchen T, Qu K, Lee Qian Y, Chanda S, Fuentes Daniel R, Giresi Paul G, Ng Yi H, Marro S, Neff Norma F*, et al.* 2013. Hierarchical Mechanisms for Direct Reprogramming of Fibroblasts to Neurons. *Cell,* 155**,** 621-635.

Ward Michelle C, Wilson Michael D, Barbosa-Morais Nuno L, Schmidt D, Stark R, Pan Q, Schwalie Petra C, Menon S, Lukk M, Watt S*, et al.* 2013. Latent Regulatory Potential of Human-Specific Repetitive Elements. *Molecular Cell,* 49**,** 262-272.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL & Schubeler D 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet,* 37**,** 853-862.

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M & Schubeler D 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet,* 39**,** 457-466.

Wei J, Mialki RK, Dong S, Khoo A, Mallampalli RK, Zhao Y & Zhao J 2013. A new mechanism of RhoA ubiquitination and degradation: Roles of SCF-FBXL19 E3 ligase and Erk2. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1833**,** 2757-2764.

Weider M, Küspert M, Bischof M, Vogl Michael R, Hornig J, Loy K, Kosian T, Müller J, Hillgärtner S, Tamm Ernst R*, et al.* 2012. Chromatin-Remodeling Factor Brg1 Is Required for Schwann Cell Differentiation and Myelination. *Developmental Cell,* 23**,** 193-201.

Weiner A, Hughes A, Yassour M, Rando OJ & Friedman N 2010. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Research,* 20**,** 90-100.

Weintraub H & Groudine M 1976. Chromosomal subunits in active genes have an altered conformation. *Science,* 193**,** 848-856.

Wenzel D, Palladino F & Jedrusik-Bode M 2011. Epigenetics in C. elegans: Facts and challenges. *genesis,* 49**,** 647-661.

Wigler M, Levy D & Perucho M 1981. The somatic replication of DNA methylation. *Cell,* 24**,** 33-40.

Williams K, Christensen J & Helin K 2012. DNA methylation: TET proteins—guardians of CpG islands? *EMBO reports,* 13**,** 28-35.

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PAC, Rappsilber J & Helin K 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature,* 473**,** 343-348.

Winter DR, Song L, Mukherjee S, Furey TS & Crawford GE 2013. DNase-seq predicts regions of rotational nucleosome stability across diverse human cell types. *Genome Research,* 23**,** 1118-1129.

Wolf D & Goff SP 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature,* 458**,** 1201-1204.

Wong S, Gearhart M, Taylor A, Nanyes D, Ha D, Robinson A, Artigas J, Lee O, Demeler B, Hart P*, et al.* 2016. KDM2B Recruitment of the Polycomb Group Complex, PRC1.1, Requires Cooperation between PCGF1 and BCORL1. *Structure*.

Wood IC, Belyaev ND, Bruce AW, Jones C, Mistry M, Roopra A & Buckley NJ 2003. Interaction of the Repressor Element 1-silencing Transcription Factor (REST) with Target Genes. *Journal of Molecular Biology,* 334**,** 863-874.

Woodcock CL, Skoultchi AI & Fan Y 2006. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Research,* 14**,** 17-25.

Wu C 1980. The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature,* 286**,** 854-860.
Wu G & Schöler HR 2014. Role of Oct4 in the early embryo development. *Cell Regeneration,* 3**,** 1-10.
Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Eve Sun Y & Zhang Y 2011a. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature,* 473**,** 389-393.
Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, Zhang B, Liu B, Wang Q, Xia W*, et al.* 2016. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature,* 534**,** 652-657.
Wu JI 2012. Diverse functions of ATP-dependent chromatin remodeling complexes in development and cancer. *Acta Biochimica et Biophysica Sinica,* 44**,** 54-69.
Wu JI, Lessard J, Olave IA, Qiu Z, Ghosh A, Graef IA & Crabtree GR 2007. Regulation of Dendritic Development by Neuron-Specific Chromatin Remodeling Complexes. *Neuron,* 56**,** 94-108.
Wu W, Cheng Y, Keller CA, Ernst J, Kumar SA, Mishra T, Morrissey C, Dorman CM, Chen K-B, Drautz D*, et al.* 2011b. Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. *Genome Research,* 21**,** 1659-1671.
Wu X, Johansen Jens V & Helin K 2013. Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. *Molecular Cell,* 49**,** 1134-1146.
Wunderlich Z & Mirny LA 2009. Different gene regulation strategies revealed by analysis of binding motifs. *Trends in Genetics,* 25**,** 434-440.
Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH & Allis CD 2005. WDR5 Associates with Histone H3 Methylated at K4 and Is Essential for H3 K4 Methylation and Vertebrate Development. *Cell,* 121**,** 859-872.
Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P*, et al.* 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature,* 442**,** 86-90.
Xi Y, Yao J, Chen R, Li W & He X 2011. Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. *Genome Research,* 21**,** 718-724.
Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, Barbera AJ, Zheng L, Zhang H, Huang S*, et al.* 2011. Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Molecular Cell,* 42**,** 451-464.
Xu Z, Chen H, Ling J, Yu D, Struffi P & Small S 2014. Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. *Genes & Development,* 28**,** 608-621.
Xue Y, Wong J, Moreno GT, Young MK, Côté J & Wang W 1998. NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities. *Molecular Cell,* 2**,** 851-861.
Yagi H, Deguchi K, Aono A, Tani Y, Kishimoto T & Komori T 1998. Growth Disturbance in Fetal Liver Hematopoiesis of Mll-Mutant Mice. *Blood,* 92**,** 108-117.
Yamada Y, Watanabe H, Miura F, Soejima H, Uchiyama M, Iwasaka T, Mukai T, Sakaki Y & Ito T 2004. A Comprehensive Analysis of Allelic Methylation Status of CpG Islands on Human Chromosome 21q. *Genome Research,* 14**,** 247-266.
Yang S-H, Kalkan T, Morissroe C, Marks H, Stunnenberg H, Smith A & Sharrocks A 2014a. Otx2 and Oct4 Drive Early Enhancer Activation during Embryonic Stem Cell Transition from Naive Pluripotency. *Cell Reports,* 7**,** 1968-1981.
Yang W, Lee Y-H, Jones AE, Woolnough JL, Zhou D, Dai Q, Wu Q, Giles KE, Townes TM & Wang H 2014b. The histone H2A deubiquitinase Usp16 regulates embryonic stem cell gene expression and lineage commitment. *Nature Communications,* 5**,** 3818.
Yang X, Noushmehr H, Han H, Andreu-Vieyra C, Liang G & Jones PA 2012. Gene Reactivation by 5-Aza-2′-Deoxycytidine–Induced Demethylation Requires SRCAP–Mediated H2A.Z Insertion to Establish Nucleosome Depleted Regions. *PLoS Genet,* 8**,** e1002604.
Yildirim O, Li R, Hung J-H, Chen P, Dong X, Ee L-S, Weng Z, Rando O & Fazzio T 2011. Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell,* 147**,** 1498-1510.
Yoon H-G, Chan DW, Reynolds AB, Qin J & Wong J 2003. N-CoR Mediates DNA Methylation-Dependent Repression through a Methyl CpG Binding Protein Kaiso. *Molecular Cell,* 12**,** 723-734.
You JS, Kelly TK, De Carvalho DD, Taberlay PC, Liang G & Jones PA 2011. OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic

regulation of its target genes. *Proceedings of the National Academy of Sciences,* 108**,** 14497-14502.
Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ & Rando OJ 2005. Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. *Science,* 309**,** 626-630.
Yuan W, Wu T, Fu H, Dai C, Wu H, Liu N, Li X, Xu M, Zhang Z, Niu T*, et al.* 2012. Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation. *Science,* 337**,** 971-975.
Zaret KS & Carroll JS 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development,* 25**,** 2227-2241.
Zemach A, McDaniel IE, Silva P & Zilberman D 2010. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science,* 328**,** 916-919.
Zentner GE & Henikoff S 2013. Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol,* 20**,** 259-266.
Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, Yin R, Zhang D, Zhang P, Liu J*, et al.* 2015. N6-Methyladenine DNA Modification in Drosophila. *Cell,* 161**,** 893-906.
Zhang X, Li B, Li W, Ma L, Zheng D, Li L, Yang W, Chu M, Chen W, Mailman Richard B*, et al.* 2014. Transcriptional Repression by the BRG1-SWI/SNF Complex Affects the Pluripotency of Human Embryonic Stem Cells. *Stem Cell Reports,* 3**,** 460-474.
Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W*, et al.* 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology,* 9**,** 1-9.
Zhang Z, Wippo CJ, Wal M, Ward E, Korber P & Pugh BF 2011. A Packing Mechanism for Nucleosome Organization Reconstituted Across a Eukaryotic Genome. *Science,* 332**,** 977-980.
Zhao J, Mialki RK, Wei J, Coon TA, Zou C, Chen BB, Mallampalli RK & Zhao Y 2013. SCF E3 ligase F-box protein complex SCFFBXL19 regulates cell migration by mediating Rac1 ubiquitination and degradation. *The FASEB Journal,* 27**,** 2611-2619.
Zhao J, Wei J, Mialki RK, Mallampalli DF, Chen BB, Coon T, Zou C, Mallampalli RK & Zhao Y 2012. F-box protein FBXL19-mediated ubiquitination and degradation of the receptor for IL-33 limits pulmonary inflammation. *Nat Immunol,* 13**,** 651-658.
Zhen CY, Duc HN, Kokotovic M, Phiel CJ & Ren X 2014. Cbx2 stably associates with mitotic chromosomes via a PRC2- or PRC1-independent mechanism and is needed for recruiting PRC1 complex to mitotic chromosomes. *Molecular Biology of the Cell,* 25**,** 3726-3739.
Zink D & Paro R 1995. Drosophila Polycomb-group regulated chromatin inhibits the accessibility of a trans-activator to its target DNA. *The EMBO Journal,* 14**,** 5660-5671.