

**Development of Chromatin Immunoprecipitation Microarray  
Technology for the Identification of Regulatory Elements in  
the Human Genome**

Shane Dillon  
St. Catharine's College

Thesis submitted for the degree of  
Doctor of Philosophy

University of Cambridge  
2008

## **Disclaimer**

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

This dissertation does not exceed the word limit prescribed by the Biology Degree Committee.

Shane Dillon

7<sup>th</sup> February 2008

## **Abstract**

### **Development of Chromatin Immunoprecipitation Microarray Technology for the Identification of Regulatory Elements in the Human Genome**

The recent development of methods which use chromatin immunoprecipitation in combination with genomic microarrays (ChIP-chip) have transformed the way in which the dynamics of chromatin and the regulation of gene expression are studied genome-wide. The aim of this thesis was to use both conventional and improved ChIP-chip approaches to characterize a variety of regulatory elements (promoters, enhancers, and putative insulators) across selected regions of the human genome in a number of cell types. Firstly, histone modifications which define promoter and enhancer elements have been used to map and characterise these elements in the K562 cell line. In parallel, the transcription factor CTCF and its known binding partners USF1, USF2, and mSin3a have been used to characterize putative insulator element. The findings of this work are discussed. Secondly, whilst ChIP-chip assays have been successfully used to map numerous DNA-protein interactions, there are limitations restricting their use in the study of cell populations that are rare and/or of limited availability. Therefore, this thesis describes the development of ChIP-chip assays that allow as few as 10,000 cells to be used per ChIP condition. These experiments do not necessitate the need to perform the ChIP reactions in the presence of carrier chromatin or the need to amplify the ChIP material prior to hybridization onto the microarray. The distribution of histone methyl and acetyl modifications with material derived from 10 000 cells was detected with a similar efficiency as that obtained from more conventional ChIP-chip approaches. This method has been applied in the study of a range of histone modifications across regions of the human genome using limited numbers of human monocytes and human embryonic stem cells. The results presented in this thesis demonstrate that developments in ChIP-chip technology can be used to accelerate our understanding of the general principles of gene regulation in the human genome.

## Acknowledgements

First and foremost I would like to thank my supervisor Dave Vetrie for his help, advice, and support throughout my studies. His commitment to my project kept me going through difficult stages! I am also grateful to the Wellcome Trust for providing me with the opportunity to study at the Sanger Institute.

I would also like to thank a number of individuals at the Sanger Institute who made valuable contributions to my studies. In particular I would like to thank Pawandeep Dhami who taught me how to perform ChIP-chip. Thanks to Christoph Koch, Gayle Clelland, Sarah Wilcox and Ian Dunham, who developed the Sanger Institute ENCODE microarray. Thanks to the Microarray facility for printing fantastic arrays; in particular I would like to thank Cordelia Langford and Peter Allis. Former team members Alex Bruce, Philippe Couttet, Jonathan Cooper, Amanda Hall, and Johanna Jim helped in many ways and were a pleasure to work with. Thanks to the media preparation team, glassware team, the library staff, and Bee Ling for flow-sorting. Rob Andrews, Keith James, and Gregory Lefebvre provided great help with computational analysis of data. I would like to thank Stephan Beck and Steve Jackson for their input during thesis committee meetings.

Thanks to Enrique Millan and Roger Pedersen (Cambridge Institute for Medical Research) for providing human embryonic stem cells. Nicola Foad and Nick Watkins at the Department of Haematology, University of Cambridge, provided monocyte cells and monocyte microarray gene expression data respectively. Thanks also to Ulas Karaoz (Boston University) for performing ROC analysis of ChIP-chip data.

Finally, a huge thanks to my family for their constant support and encouragement during my studies. Carol and Ella, this thesis is for you. Now I can spend more time with you, something you both deserve!

# Table of Contents

<b>Chapter 1 Introduction</b> .....	1
1.1. Coding regions in the human genome .....	1
1.2. Non-coding elements in the human genome .....	2
1.3. Regulation of gene expression.....	4
1.4. Types of non-coding regulatory elements involved in controlling gene expression.....	4
1.4.1. Core and Proximal Promoters .....	5
1.4.2. Enhancers.....	7
1.4.3. Silencers/repressors .....	8
1.4.4. Insulators .....	9
1.4.4.1 Enhancer blocking insulators .....	11
1.4.4.2 Barrier Insulators .....	12
1.4.4.3 CTCF is a multifunctional protein .....	13
1.4.5. Locus control regions.....	15
1.4.6. Matrix Attachment regions (MARs).....	15
1.5. Proteins involved in transcriptional regulation.....	16
1.5.1. RNA polymerase complexes .....	16
1.5.2. Sequence-specific transcription factors .....	17
1.5.3. Proteins involved in chromatin remodelling and modification.....	19
1.5.3.1. ATP-dependent chromatin remodelling enzymes.....	19
1.5.3.2. Histone modifying enzymes.....	20
1.6. Epigenetic regulation of transcription.....	22
1.6.1. The Nucleosome – the fundamental repeating unit of chromatin.....	22
1.6.2. Histone modifications and their functions .....	26
1.6.2.1 Acetylation.....	27
1.6.2.2 Methylation .....	28
1.6.2.3. Phosphorylation.....	31
1.6.2.4. Ubiquitination.....	31
1.6.2.5. Sumoylation.....	32
1.6.2.6. ADP-ribosylation.....	32
1.6.3. The histone code hypothesis .....	32
1.6.4. DNA methylation.....	34
1.6.5. Regulation of embryonic stem cell pluripotency .....	35
1.6.5.1. Transcription factors involved in regulating pluripotency.....	35
1.6.5.2. Epigenetic regulation of pluripotency.....	36
1.7. Identification and characterisation of non-coding regulatory elements.....	37
1.7.1. Classical or low-throughput methods.....	38
1.7.1.1. DNA footprinting.....	38
1.7.1.2. DNase I hypersensitive site mapping .....	38
1.7.1.3. Electromobility shift assays.....	39
1.7.1.4. PCR-based methods for detecting DNA-protein interactions .....	39
1.7.1.5. Reporter-gene assays which confer function .....	40
1.7.2. Computational detection of regulatory elements .....	43
1.7.2.1. Promoter location prediction .....	43
1.7.2.2. Prediction of transcription factor binding sites .....	44

1.7.2.3.	Comparative sequence analysis.....	44
1.7.3.	Applications of genomic DNA microarrays to identify regulatory elements.....	46
1.7.3.1.	Transcript profiling.....	47
1.7.3.2.	Replication timing.....	48
1.7.3.3.	DNase I hypersensitive site microarrays.....	49
1.7.3.4.	Matrix attachment regions microarrays.....	50
1.7.3.5.	Chromatin immunoprecipitation microarrays (ChIP-chip).....	50
1.7.4.	Other ChIP-based methods.....	54
1.8.	Genomic microarray platforms used in this study.....	55
1.8.1.	The SCL genomic tiling path microarray.....	55
1.8.2.	The ENCODE project.....	58
1.8.2.1.	A summary of the pilot phase findings.....	58
1.8.2.2.	The Sanger Institute ENCODE Microarray.....	60
1.9.	Aims of this thesis.....	60
<b>Chapter 2 Materials and Methods</b> .....		<b>61</b>
2.1.	Composition of solutions.....	61
2.2.	Reagents.....	68
2.3.	Cells and cell lines.....	69
2.4.	Tissue culture.....	70
2.4.1.	Culturing of cell lines.....	70
2.4.2.	Cell cryopreservation.....	71
2.4.3.	Chromatin immunoprecipitation (ChIP).....	71
2.5.	Construction of the SCL tiling path microarray and the ENCODE tiling path microarray....	77
2.5.1.	Generation of tiling path amplicons.....	77
2.5.2.	Microarray slide printing and processing.....	77
2.6.	Hybridisation of microarrays.....	78
2.6.1.	Labelling by random priming of DNA samples.....	78
2.6.2.	Purification of labelled DNA samples.....	79
2.6.3.	Array hybridisation set-up using the Tecan HS 4800 hybridisation station.....	79
2.6.4.	Scanning and processing of ChIP-chip data.....	83
2.7.	Analysis of ChIP-chip data with respect to genomic features.....	84
2.8.	Statistical analysis.....	84
2.9.	Gene expression analysis.....	85
2.10.	Receiver operator characteristic (ROC) curve analysis.....	86
2.11.	DNA sequence motif analysis.....	86
2.12.	Real-time PCR.....	87
2.13.	Western blotting procedure.....	88
<b>Chapter 3 Applying ChIP-chip to study regulatory elements in 1% of the human genome</b> .....		<b>92</b>
3.1.	Introduction.....	92
3.2.	Aims of this chapter.....	96
3.3.	Overall strategy.....	97
3.4.	Criteria used for performing chromatin immunoprecipitation assays and microarray hybridisation.....	99
3.5.	Creating histone modification profiles across the ENCODE regions in K562 cells.....	104
3.5.1.	Assessing the performance of the ENCODE array.....	104
3.5.2.	Constructing histone modification and nucleosome density profiles.....	108

3.6.	Analysing the distribution of histone modifications .....	110
3.6.1.	Defining statistically enriched regions.....	110
3.6.2.	The distribution of histone modifications with respect to gene features .....	111
3.6.3.	Analysis of the combinatorial nature of histone modifications .....	112
3.7.	Distinct histone modification signatures define transcription start sites and distal elements in the ENCODE regions .....	114
3.8.	Histone modifications and transcription .....	118
3.8.1.	The relationship between histone modifications and transcription status .....	118
3.8.2.	Histone modifications and predicting expression status.....	121
3.8.3.	Nucleosome density at transcription start sites and distal elements .....	123
3.9.	Discussion .....	125
3.9.1.	Histone modification signatures associated with promoters and distal elements .....	126
3.9.2.	Histone modifications and transcriptional activity .....	127
3.9.3.	Nucleosome depletion at regulatory elements .....	129
3.9.4.	Analysis and interpretation of CHIP-chip data .....	129
3.9.5.	Conclusions .....	131
<b>Chapter 4 Identification and characterisation of binding sites of the insulator binding protein CTCF in the human genome .....</b>		<b>131</b>
4.1.	Introduction.....	132
4.2.	Aims of this chapter.....	137
4.3.	Overall strategy .....	137
4.4.	Assessing the specificity of transcription factor antibodies in western blotting assays.....	138
4.5.	Developing a CHIP-chip assay to detect putative insulators at the SCL locus .....	139
4.6.	Mapping and characterising CTCF binding sites in the ENCODE regions .....	142
4.6.1.	Implementation and validation of the CTCF CHIP-chip assay .....	142
4.6.2.	Distribution of CTCF-binding sites in the ENCODE regions.....	147
4.6.3.	CTCF sites at individual genes and gene clusters .....	150
4.7.	A comparative and sequence based analysis of CTCF sites in different cell types .....	153
4.8.	Analysis of other transcription factors implicated in CTCF or insulator function.....	157
4.8.1.	Developing assays using the SCL locus as model system .....	157
4.8.2.	Mapping the distribution of mSin3a, USF1, and USF2 binding sites in the ENCODE regions.....	159
4.8.3.	Analysing interactions between CTCF and mSin3a, USF1, and USF2.....	161
4.8.4.	Transcription factor binding and gene expression status .....	165
4.8.5.	Motifs analysis of CTCF, mSin3a, USF1 and USF2 binding sites.....	167
4.9.	Chromatin structure at insulators .....	172
4.9.1.	CTCF binding sites are located in accessible chromatin domains .....	172
4.9.2.	CTCF is located at the boundary between active and inactive chromatin domains.....	175
4.10.	Discussion .....	181
4.10.1.	Widespread distribution of CTCF binding sites in the human genome .....	181
4.10.2.	Cross-study comparison of CTCF binding sites .....	182
4.10.3.	CTCF co-localisation with mSin3a, USF1 and USF2 binding sites.....	185
4.10.4.	mSin3a interaction at promoters is associated with active gene expression .....	185
4.10.5.	Identification of transcription factor consensus binding motifs .....	186
4.10.6.	CTCF and chromatin structure .....	187

<b>Chapter 5 Optimisation of conditions for ChIP-chip when using cell types that are limiting in number</b> .....	190
5.1. Introduction.....	190
5.2. Aims of this Chapter .....	198
5.3. Overall strategy .....	199
5.4. The effect of titrating cell numbers in ChIP-chip and its effect on the detection of regulatory elements at the human SCL tiling locus .....	199
5.5. The effect of titrating antibody levels on low cell number ChIP .....	204
5.5.1. Histone H3 K9/K14 acetylation experiments .....	204
5.5.2. Histone H3 K4 trimethylation experiments .....	211
5.6. Examining the effect of protein-G agarose concentration on ChIP-chip experiments performed with low cell numbers .....	220
5.7. Assessing reproducibility of the limited cell numbers ChIP-chip method.....	222
5.8. Histone H3K4me3 microarray data correlates with real-time quantitative PCR of ChIP material derived from 10 <sup>5</sup> and 10 <sup>4</sup> cells.....	223
5.9. Assessing the impact of normalising ChIP-chip data with normal rabbit IgG data .....	225
5.10. Testing of other histone modification antibodies with the modified ChIP-chip method.....	231
5.11. Detection of transcription factor interactions using low cell number ChIP-chip .....	238
5.12. Discussion .....	240
5.12.1. The development of a ChIP-chip method applicable to small cell populations .....	241
5.12.2. Further optimisation of the modified ChIP-chip method .....	242
5.12.3. Comparisons with other methods.....	243
5.13. Conclusions.....	244
<b>Chapter 6 Histone modification profiles of human embryonic stem cells and lineage-committed human monocytes</b> .....	245
6.1. Introduction.....	245
6.2. Aims of this Chapter .....	249
6.3. Creating chromatin maps in pluripotent and lineage committed cells.....	249
6.4. Analysis of chromatin state in pluripotent hESCs and lineage committed monocytes .....	250
6.4.1. Promoter chromatin state in hESCs and monocytes.....	250
6.4.2. The chromatin state of promoters reflects developmental potential.....	253
6.4.3. Bivalent promoters are associated with developmental genes in hESCs and monocytes .....	256
6.4.4. The Polycomb Repressive Complex 2 (PRC2) subunit SUZ12 co-localises with bivalent developmental genes in human embryonic stem cells .....	258
6.4.5. Histone H3K36 trimethylation is associated with primary protein-coding transcripts and non-coding RNAs in hESCs and monocytes.....	262
6.4.6. Histone H3K79 trimethylation: Possible link to transcription elongation in hESCs and monocytes .....	267
6.5. A detailed analysis of histone acetylation and methylation modifications in human monocytes .....	269
6.5.1. Histone acetylation modifications are associated with active gene expression .....	270
6.5.2. Histone H3 lysine methylation patterns in human monocytes .....	272
6.5.2.1. Histone H3K4 methylation patterns at active and inactive genes .....	272
6.5.2.2. The histone signature of distal enhancer/repressor elements .....	273
6.5.2.3. Histone H3K9 methylation is implicated at actively transcribed genes.....	275
6.5.2.4. Histone H3K27 methylation states are found at active and inactive genes .....	277



6.5.2.5. Analysis of histone H3K36 methylation states .....	278
6.5.2.6. Histone H3K79 methylation states are associated with highly transcribed genes..	280
6.6. Discussion .....	281
6.6.1. Bivalent chromatin structures are present in pluripotent hESCs and lineage committed monocytes.....	282
6.6.2. Histone H3K36me3 and H3K79me3 are associated with different stages of transcription in hESCs and monocytes .....	284
6.6.3. Identification of a consensus histone code for active and inactive genes in human monocytes .....	285
<b>Chapter 7 Summary and future work .....</b>	<b>289</b>
7.1. Summary of the work presented in this thesis .....	289
7.1.1. Application of microarray technology for large-scale characterisation of promoter and distal enhancer/repressor elements in the human genome (Chapter 3).....	289
7.1.2. ChIP-chip analysis of the insulator binding protein CTCF and its associated transcription factors (Chapter 4) .....	290
7.1.3. Development of a modified ChIP-chip method for use with fewer cells (Chapter 5).....	291
7.1.4. Application of the modified ChIP-chip method to study chromatin regulation during differentiation (Chapter 6).....	292
7.1.5. Identification of a consensus histone code in human monocytes (Chapter 6) .....	293
7.2. Future Work.....	293
7.2.1. Further development of ChIP-chip assays for characterising regulatory elements and the role of histone modifications in gene regulation .....	294
7.2.2. Applications of microarrays for characterising other features of the human genome ...	296
7.2.3. Testing the proposed histone code.....	298
7.2.4. Further characterisation of putative CTCF insulator elements .....	299
7.3. Final thoughts.....	300
<b>References .....</b>	<b>301</b>
<b>Appendix 1 .....</b>	<b>338</b>

# List of Figures

Figure 1.1: Genome organisation of the major model organisms .....	3
Figure 1.2: The typical regulatory features involved in the regulation of gene expression .....	5
Figure 1.3: REST-mediated repression of target genes.....	9
Figure 1.4: The two types of insulator activity .....	10
Figure 1.5: Genomic organisation of the chicken $\beta$ -globin locus.....	13
Figure 1.6: Sites of histone modifications.....	26
Figure 1.7: Functional plasmid-based reporter gene assays for the identification of regulatory elements .....	42
Figure 1.8: The basic principles of two-colour competitive hybridisation .....	47
Figure 1.9: Outline of the carrier CHIP (CCHIP) method.....	53
Figure 1.10: Microarray surface chemistry .....	55
Figure 1.11: The genomic region included on the human SCL tiling path microarray.....	56
Figure 1.12: Histone acetylation and methylation states define regulatory elements at the SCL locus .....	57
Figure 3.1: Schematic representation of the overall strategy used to map regulatory interactions across 1% of the human genome.....	98
Figure 3.2: Analysis of cell growth by flow-cytometry.....	100
Figure 3.3: Electrophoresis of CHIP DNAs .....	101
Figure 3.4: Labeling of DNA by a random priming method .....	102
Figure 3.5: Electrophoresis of fluorescently labeled DNAs.....	103
Figure 3.6: A composite image of the Sanger Institute ENCODE array.....	106
Figure 3.7: Assessing the performance of the ENCODE microarray in independent CHIP-chip assays.....	107
Figure 3.8: An example of the histone modification and nucleosome density maps generated across the ENCODE regions .....	109
Figure 3.9: Distribution of histone modifications in K562 cells.....	111
Figure 3.10: The relative occurrence of overlapping histone modification sites in K562 .....	112
Figure 3.11: The extent of overlapping H3K4 methylation sites in K562. The pie-chart illustrates the occurrence of the seven H3K4 methylation combinations. ....	113
Figure 3.12: The distribution of histone H3K4 methylation combinations relative to the nearest transcriptional start site (TSS) .....	115
Figure 3.13: Histone modification profiles at TSSs and distal sites.....	117
Figure 3.14: Histone modification profiles for non-expressed and expressed genes in K562 cells .....	119
Figure 3.15: Profiles showing H3K4me3 to H3K4me1 ratios at the promoters of active and inactive genes in K562 cells .....	120
Figure 3.16: Receiver operating characteristic (ROC) curves illustrating the predictive power of histone modifications in K562 cells.....	122
Figure 3.17: Nucleosome occupancy at transcription start sites and distal elements in K562 cells .....	124
Figure 3.18: Histone modification levels relative to nucleosome occupancy in K562 cells.....	125
Figure 4.1: CTCF enhancer-blocking insulator at the Igf2/H19 imprinted region.....	134
Figure 4.2: A 20-mer motif is recognised by CTCF .....	136
Figure 4.3: Western blot analysis of CTCF, mSin3a, USF1 and USF2 in K562 cells .....	139

Figure 4.4: ChIP-chip profile of CTCF binding across the SCL locus in K562 before and after Goat IgG normalisation.....	141
Figure 4.5: CTCF binding sites in the IGF2/H19 locus.....	143
Figure 4.6: CTCF binding sites at the $\beta$ -globin locus.....	144
Figure 4.7: Comparison of CTCF binding sites in the K562 cell with binding sites identified in primary human fibroblast, IMR90 cells.....	146
Figure 4.8: The binding intervals and distribution of CTCF sites relative to transcription start sites (TSSs) in K562.....	147
Figure 4.9: Correlation of CTCF binding events with gene density in the ENCODE regions.....	148
Figure 4.10: The distribution of CTCF binding sites.....	149
Figure 4.11: CTCF binding intervals indicate multiple CTCF sites can bind in close proximity...	150
Figure 4.12: CTCF binding sites flank a cluster of genes in ENCODE region Enm006.....	151
Figure 4.13: Genes can contain several CTCF binding sites.....	152
Figure 4.14: A comparison of CTCF binding sites in different cell types and between studies ...	156
Figure 4.15: ChIP-chip profile of mSin3a interactions across the SCL locus in K562 before and after normal rabbit IgG normalisation.....	158
Figure 4.16: The distribution of USF1, USF2, and mSin3a sites of interaction in the ENCODE regions in K562 cells.....	159
Figure 4.17: Distribution of mSin3a, USF1, USF2 interactions.....	161
Figure 4.18: CTCF binding sites that overlap with binding sites of mSin3a, USF1, and USF2 are found in diverse locations.....	163
Figure 4.19: Predictive power of transcription factor binding for gene expression in K562.....	166
Figure 4.20: Nucleosome density and FAIRE profile at CTCF binding sites in the ENCODE regions.....	173
Figure 4.21: CTCF binding sites co-localise with sites of DNase I hypersensitivity at the $\alpha$ - and $\beta$ -globin loci.....	175
Figure 4.22: ChIP-chip profile of Histone H3K27me3 across the SCL locus in K562.....	176
Figure 4.23: CTCF binding sites demarcate the boundary between chromatin regions associated with active and inactive histone modifications.....	178
Figure 5.1 Principles of ChIP.....	192
Figure 5.2: ChIP-chip profiles of H4 K9/K14 diacetylation across the human SCL locus in K562 for a range of cell numbers.....	201
Figure 5.3: Titration of H3K9/K14 diacetylation antibody in ChIP assays performed with $10^7$ K562 cells.....	205
Figure 5.4: Titration of H3K9/K14 diacetylation antibody in ChIP assays performed with $10^6$ K562 cells.....	206
Figure 5.5: Titration of H3K9/K14 diacetylation antibody in ChIP assays performed with $10^5$ K562 cells.....	207
Figure 5.6: Titration of H3K9/K14 diacetylation antibody in ChIP assays performed with $10^4$ K562 cells.....	208
Figure 5.7: Titration of H3K4me3 antibody in ChIP assays performed with $10^7$ K562 cells.....	212
Figure 5.8: Titration of H3K4me3 antibody in ChIP assays performed with $10^6$ K562 cells.....	213
Figure 5.9: Titration of H3K4me3 antibody in ChIP assays performed with $10^5$ K562 cells.....	214
Figure 5.10: Titration of H3K4me3 antibody in ChIP assays performed with $10^4$ K562 cells.....	215
Figure 5.11: A comparison of the optimal antibody concentrations for titration assays performed with H3 acetylation and H3K4me3 antibodies.....	219
Figure 5.12: Titration of protein-G agarose in ChIP assays performed with $10^5$ K562 cells and 0.5 $\mu$ g of H3 acetylation antibody.....	221

Figure 5.13: Biological replicates for ChIP-chip assays performed with $10^4$ cells.....	222
Figure 5.14: SYBR green real-time quantitative PCR of H3K4me3 ChIP material from reduced cell numbers .....	224
Figure 5.15: How normalisation with a normal rabbit IgG affects the spread of data for a ChIP-chip experiment performed with $10^5$ cells.....	226
Figure 5.16: Rabbit IgG normalisation of a ChIP-chip experiment performed with $10^5$ cells .....	227
Figure 5.17: Rabbit IgG normalisation of data derived from ChIP-chip experiments performed with $10^4$ cells .....	229
Figure 5.18: Testing the modified ChIP-chip method to detect regions associated with H3K27me3 .....	233
Figure 5.19: Testing the modified ChIP-chip method to detect regions associated with H3K36me3 .....	235
Figure 5.20: Testing the modified ChIP-chip method to detect regions associated with H3K79me3 .....	237
Figure 5.21: Detection of GATA-1 enrichments in reduced numbers of K562 cells.....	239
Figure 6.1: The relationship between SSEA3+ and SSEA3- H9 hESCs and CD14+ monocyte cells.....	248
Figure 6.2: Chromatin modification patterns of promoters in human embryonic stem cells and monocyte cells. ....	251
Figure 6.3: Distinct histone modification profiles at high and low CpG content promoters .....	253
Figure 6.4: Promoter chromatin state and developmental potential.....	254
Figure 6.5: Conserved and cell-type specific bivalent promoters.....	258
Figure 6.6: Representative examples of genes associated with a bivalent promoter structure and SUZ12 binding in human embryonic stem cells .....	262
Figure 6.7: H3K36me3 annotates transcribed genes in hESCs and monocytes .....	264
Figure 6.8: Histone H3K36me3 is associated with non-coding RNAs and putative novel transcripts .....	267
Figure 6.9: H3K79me3 association with the immediate transcribed region of active genes in hESCs and monocytes .....	269
Figure 6.10: Histone acetylation patterns at active and inactive genes .....	271
Figure 6.11: Histone H3K4 methylation patterns at active and inactive genes .....	273
Figure 6.12: Histone acetylation and methylation patterns at distal enhancer/repressor elements .....	274
Figure 6.13: Histone H3K9 methylation patterns at active and inactive genes .....	275
Figure 6.14: Histone H3K27 methylation states at active and inactive genes .....	277
Figure 6.15: Histone H3K36 methylation patterns at active and inactive genes .....	278
Figure 6.16: Histone H3K79 methylation patterns at active and inactive genes .....	280
Figure 6.17: A consensus histone code for regulatory and transcribed regions in human monocytes .....	285

## List of Tables

Table 1.1: The structure of metazoan core promoters .....	7
Table 1.2: Histone variants and their functions .....	25
Table 2.1: Summary of chromatin, antibody and buffer amounts used in reduced cell number ChIP experiments .....	74
Table 2.2: Wash steps for hybridisations performed on the Tecan .....	82
Table 3.1: Description of the ENCODE regions .....	95
Table 3.2: Descriptive statistics for Chipotle sites for K562 ChIP-chip data .....	110
Table 4.1: Gene clusters flanked by CTCF binding sites. 10 gene clusters (containing five or more genes) were flanked by CTCF sites .....	151
Table 4.2: ENCODE Genes containing multiple CTCF binding sites in K562 are membrane components .....	153
Table 4.3: Correspondence between ChIP-chip defined and predicted CTCF sites.....	154
Table 4.4: Overlapping combinations of CTCF, mSin3a, USF1, and USF2 binding sites in K562 .....	162
Table 4.5: Known motifs associated with CTCF binding sites .....	168
Table 4.6: De novo CTCF motif discovery .....	169
Table 4.7: Discovery of novel motifs associated with mSin3a interactions.....	170
Table 4.8: Searching for known motifs in USF1 and USF2 binding sites.....	171
Table 4.9: CTCF binding sites in K562 located at the boundary between active and inactive genomic regions defined by enriched levels of H3K27me1 and H3K27me3 respectively.....	180
Table 5.1: A summary of the performance of H3 K9/K14 diacetylation ChIP-chip assays performed with a standard range of cell numbers and 10 µg of antibody .....	203
Table 5.2: Statistical assessment of assay performance for the six antibody concentrations used in ChIP-chip assays performed with 10 <sup>7</sup> , 10 <sup>6</sup> , 10 <sup>5</sup> , and 10 <sup>4</sup> cells and H3 K9/K14 diacetylation antibody .....	210
Table 5.3: Statistical assessment of assay performance for the six antibody concentrations used in ChIP-chip assays performed with 10 <sup>7</sup> , 10 <sup>6</sup> , 10 <sup>5</sup> , and 10 <sup>4</sup> cells and H3K4me3 antibody.....	217
Table 5.4: Comparisons of 10 <sup>5</sup> and 10 <sup>4</sup> ChIP-chip data before and after rabbit IgG normalisation .....	230
Table 5.5: Assessing the impact of normalisation on the average titration series signal, average series standard deviation and average series signal:noise values for experiments performed with 10 <sup>5</sup> and 10 <sup>4</sup> cells.....	231
Table 6.1: Gene ontology analysis of genes associated with bivalent promoters.....	257
Table 6.2: SUZ12 is associated with bivalent genes in human embryonic stem cells .....	259
Table 6.3: Non-coding RNAs in the ENCODE regions are associated with H3K36me3 .....	265