Epigenetic programming defines stem cell identity and entry into the proliferative compartment in chronic myeloid leukaemia (CML)

Koorosh Korfi

Thesis submitted for the degree of Doctor of Philosophy

Institute of Cancer Sciences

College of Medical, Veterinary and Life Sciences

University of Glasgow

November 2012

Abstract

Chronic myeloid leukaemia (CML) is a haematological malignancy that is identified by the presence of a fusion oncogene, BCR-ABL1, which is a constitutive tyrosine kinase. The discovery of tyrosine kinase inhibitors (TKIs) over that past decade has resulted in significantly improved survival rates and disease management in CML patients. However, a subpopulation of BCR-ABL1$^+$ cells in the niche are found which exhibit stem cell-like features, such as self-renewal and quiescence. These CML stem cells (LSCs) are shown to be insensitive to TKI treatment and are capable of deriving the disease during the relapse. Consequently, the elimination of LSCs is a primary goal of current research. Therefore, the aim of this thesis was to obtain a global view of the cellular processes that maintain stem cell identity in CD34$^+$ CD38$^-$ LSCs as well as identify those processes which initiate the transition to proliferative CD34$^+$ CD38$^+$ CML progenitor cells (LPCs). A combined approach was exploited to investigate genome-wide gene expression profiles and histone modification signatures of normal HSCs and committed progenitors (HPCs), and their LSC and LPC counterparts. Despite having increased activity in pathways involved in cell division and proliferation, expression levels of the pathways involved in stem cell identity were not significantly different in LSCs to those found in HSCs. These pathways included Wnt, TGF-β signalling, and several novel neurotransmitter signalling pathways. By examining genome-wide histone modification patterns using ChIP-sequencing it was shown that the stem cell identities of HSCs and LSCs are programmed at the epigenetic level. All of the pathways which confer stem cell identity to both HSCs and LSCs are significantly enriched for bivalent gene promoters having both the H3K4me3 and H3K27me3 marks. These similarities were most evident in neurotransmitter signalling and it was demonstrated that these pathways are capable of promoting LSC maintenance *in vitro*. Intriguingly, although the stem cell entry into the proliferative state occurs through the repression of many of the same stem cell identity pathways in both HSCs and LSCs, it was shown that epigenetic reprogramming in CML mediates this repression via a different mechanism than in normal HSCs. Furthermore, abnormalities in levels of several chromatin enzymes were identified that are likely to be responsible for the epigenetic reprogramming of CML cells. The work presented in this thesis defines the chromatin landscape of a cancer stem cell for the first time and provides new therapeutic targets for the eradication of TKI resistant CML stem cells.

Table of contents

List of tables

List of figures

Acknowledgements

First and foremost, I would like to thank my primary supervisor David Vetrie for all the help, advice and guidance throughout this PhD. His commitment to this project kept me going through difficult stages. I would also like to express my deep gratitude to my second supervisor, Tessa Holyoake, for being a great mentor and giving me invaluable support and guidance during this project. I would like to thank my advisor Peter Adams for his valuable feedbacks and guidance throughout this project. I am also indebted to the University of Glasgow and the Wellcome Trust for funding this study.

I wish to thank a number of individuals at the Section of Epigenetics and the Paul O'Gorman Leukaemia Research Centre for their valuable contribution to this project. In particular, I would like to thank Peter Saffrey and Andrew Crossan for providing help with the bioinformatics analyses. Thanks to Heather Jorgensen, Alan Hair, Ashley Hamilton, and Jennifer Cassels for the processing the patient samples and flow-sorting. Thanks to the Glasgow Polyomics facility, especially Pawel Herzyk, Julie Galbraith, and Jing Wang, for performing the ChIP-seq and Affymetrix experiments. Thanks to Michelle Cruz for helping with the validation of ChIP-seq and Affymetrix analyses. Special thanks to Paolo Gallipoli for the help with the neurotransmitter part of this study. Former team members Yan Zhou, Kelly Chiang, Nicolas Bonhoure, Serdar Kasakyan, and Sreenivas Kurukuti helped in many ways and were a pleasure to work with.

Thanks to my amazing friends for their encouragements and for keeping my spirits up during the past four years, especially Susan, Gokula, Milica, Bhoomi, Behzad, Yan, Carolyn, and Elaine. You made my life in Glasgow fun, joyful and filled with excitement! My best friend Bahareh also deserves special thanks for keeping the phone lines busy and being a constant support over the past four years.

I want to express my deeply-felt thanks to Imogen who shared my highs and lows during the past few months and supported me in every way possible. Thank you for being very patient and helping me during the write-up period.

A big thanks to my wonderful parents, Felor and Nasser, who despite being so far away showered me with lots of love and encouragement and were the utmost source of inspiration for me to finish this PhD successfully. I dedicate this thesis to them. I could not have done this without your support!

Author's declaration

This thesis 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.

Definitions and abbreviations

3D	three-dimensional
5mC	5-methylcytosine
5HT	serotonin
ACH	acetylcholine
AGM	aortic-gonado-mesonephrons
ALL	Acute lymphoblastic leukaemia
AML	acute myeloid leukaemia
BC	blast crisis
BCL	B-cell CLL/lymphoma
BCR	breakpoint cluster region
BED	browser extensible data
BH	Benjamini Hochberg
BM	bone marrow
BMP	bone morphogenetic protein
bp	base pair
CCL	chemokine (C-C motif) ligand
CCND1	cyclin D1
CDK	cyclin-dependent kinase
CDKI	cyclin dependent kinase inhibitor
CDKN	cyclin-dependent kinase inhibitor
CEBPA	CCAAT/enhancer binding protein-alpha
CFC	colony-forming cell
CFSE	carboxyfluorescein diacetate succinimidyl ester
CFU-C	colony-forming unit in culture
ChIP	chromatin immunoprecipitation
ChIP-seq	ChIP-sequencing
CLB	cell lysis buffer
CLP	common lymphoid progenitor
CML	chronic myeloid leukaemia
CMP	common myeloid progenitor
CP	chronic phase
CXCL	chemokine (C-X-C motif) ligand
CXCR	chemokine (C-X-C motif) receptor
DMEM	Dulbecco's modified Eagle medium
DMSO	dimethyl sulfoxide
DNase I	deoxyribonuclease 1
DNMT	DNA methyltransferase
dNTP	deoxyribonucleotide triphoshphate
EDTA	ethylenediaminetetraacetic acid
EED	embryonic ectoderm development
ENCODE	the Encyclopaedia of DNA Elements
ERK	extracellular signal-regulated kinase
ES	embryonic stem
EZH	enhancer of zeste homolog
FACS	fluorescence-activated cell sorting
FBS	foetal bovine serum
FDR	false discovery rate
FISH	fluorescence *in situ* hybridisation
G-CSF	granulocyte colony-stimulating factor
GEF	guanine nucleotide exchange factor

GFI1	growth factor independent 1
GFP	green fluorescence protein
GLU	L-glutamate
GM-CSF	granulocyte-monocyte colony-stimulating factor
GMP	granulocyte/monocyte progenitor
GO	Gene Ontology
GPCR	G protein-coupled receptor
GRK	GPCR kinase
GSK	glycogen synthase kinase
GTF	general transcription factor
HAT	histone acetyltransferase
HDAC	histone deacetylase
HIF1α	hypoxia inducible factor 1α
HIS	histamine
HLF	hepatic leukaemia factor
HMT	histone methyltransferase
HPC	haemopoietic progenitor cell
HPLC	high-performance liquid chromatography
hrs	hours
HS	hybridisation station
HSC	haemopoietic stem cell
IL	interleukin
IMDM	Isocove's modified Dulbecco's medium
IP	immunoprecipitation
IPDB	IP dilution buffer
IPEB	IP elution buffer
IPWB	IP wash buffer
IVT	*in vitro* transcription
kb	kilobase
KDM	lysine demethylase
LCR	locus control region
LIF	leukaemia inhibitory factor
Lin	lineage
LPC	CML progenitor cell
LSC	CML stem cell
LSK	Lin$^-$ Sca-1$^+$ c-Kit$^+$
LT-HSC	long-term HSCs
LTC-IC	long-term culture-initiating cell
MAPK	mitogen activated protein kinase
me	methylation
MEP	megakaryocyte/erythroid progenitor
mES	mouse embryonic stem
mins	minutes
MLL	mixed-lineage leukaemia
MLP	multilymphoid progenitor
MNC	mononuclear cell
MPP	multipotent progenitor
MSC	mesenchymal stem cell
N-ChIP	native ChIP
ncRNA	non-coding RNA
NDC	no-drug control
NK	natural killer
NLB	nuclear lysis buffer

NOD	non-obese diabetic
NPC	non-pre-cleared
NUSE	Normalised Unscaled Standard Error
PB	peripheral blood
PBS	phosphate buffered saline
PC	pre-cleared
PCA	principal component analysis
PcG	Polycomb group
PCL	Polycomb-like
PCR	polymerase chain reaction
PDGF	platelet-derived growth factor
Ph	Philadelphia chromosome
PHD	plant homeodomain
PI3K	phosphatidylinositol 3-kinase
PIC	pre-initiation complex
PKA	protein kinase A
PKC	protein kinase C
PLCβ	phospholipase Cβ
PM	perfect match
PRC	Polycomb repressive complex
QC	quality control
qPCR	quantitative PCR
RBBP	retinoblastoma binding protein
RLE	Relative Log Expression
RMA	Robust Multi-array Averages
RNAPII	RNA polymerase II
RNase	ribonuclease
rpm	rounds per minute
RT-PCR	real time-PCR
RTm	room temperature
Scid	severe combined immune-deficient
SDS	sodium dodecyl sulphate
secs	seconds
SMO	Smoothened
STAT	signal transducer and activator of transcription
SUZ12	suppressor of zeste 12 homolog
TAE	Tris-acetate-EDTA
TAL1	T-cell acute lymphocytic leukaemia 1
TBE	Tris-borate-EDTA
TBP	TATA box binding protein
TCFs	T cell-specific transcription factors
TE	Tris-EDTA
TFBS	transcription factor binding site
TGF-β	Transforming growth factor-beta
TGFBR	TGF-β receptor
TKI	tyrosine kinase inhibitor
TPO	thrombopoietin
TRH	thyrotropin-releasing hormone
TrxG	Trithorax group
TSS	transcriptional start site
UCSC	University of California Santa Cruz
WT	Whole-Transcript
X-ChIP	crosslinked ChIP

1. Introduction

Haemopoiesis is a complex and hierarchical biological process that is responsible for generating one trillion blood cells daily, and therefore, is tightly regulated to maintain general homeostasis. However, abnormalities in the regulation of this process are responsible for the development of many haematological disorders and malignancies, such as leukaemias and lymphomas, which can significantly reduce life expectancy. Therefore, a better understanding of normal and malignant haemopoiesis from the biological point of view will provide opportunities to find new therapeutics that can rescue the normal phenotype and increase the survival rates of cancer sufferers. A comprehensive understanding of biology cannot be possible without studying the genome which consists of all the genetic material contained in a cell of an organism and contains all of the information necessary for life. The genome is comprised of three types of features - genes, regulatory elements and maintenance elements. Genes are the coding entities for the cellular proteome, while the spatial and temporal production of proteins are controlled by the regulatory elements, which include promoters, enhancers, insulators, as well as non-coding regulatory RNAs. Nevertheless, the epigenome regulates the function of the genome in an additional layer mediated by methylation of DNA and biochemical modifications of chromatin. DNA repair, replication and recombination are controlled by maintenance elements, such as centromeres, telomeres, origins of replication and recombination hotspots. After sequencing of the human genome was completed in 2004 (IHGS, 2004), the next challenge in the post-genome era is to elucidate the role of genomic and epigenomic regulation during normal development and disease states. Thus, this Chapter is going to review the knowns and unknowns of normal and malignant haemopoiesis and the role of epigenetics in these processes.

1.1 Systems biology approach

The complexity of eukaryotic systems and the rapid development of new technologies over the past decade to study the biology of cells and organisms have formed a new research territory, known as systems biology. The systems biology approach is, therefore, based on collecting all the information presented at the single cell or multicellular levels in order to understand various functions of the system and their regulation. In other words, the regulatory mechanisms and

biological pathways are so intensively interconnected that make the understanding of the whole system impossible without studying all the biological pathways and mechanisms involved. The systems biology research has been mainly driven by the information collected by genomics, epigenomics, transcriptomics, proteomics and metabolomics. Furthermore, sophisticated computational modelling is required to integrate the datasets collected by these five components (Alon, 2007).

The proteome is defined as the library of expressed proteins in an organism or a cell type under a defined condition. Thus, the proteomics task is to detect and quantify the levels, modifications, cellular localisation, and interactions of all expressed proteins. The characterisation of interacting partners of a single protein is by far the most challenging task of proteomics since such interactions are extremely fragile and transient. However, recent advances in the mass spectrometry technology has revolutionised the field of proteomics by allowing the analysis of protein-protein interactions as well identification of post-translational modifications, such as phosphorylation. The identification of cellular localisation is another important aspect of protein characterisation which can be achieved by cell imaging or fractionation approaches (Lovric, 2011).

Proteomics are closely interconnected with the transcriptomics that is the science of defining and quantifying all the various transcripts present in a cell, or population of cells, including mRNAs, non-coding RNAs (ncRNAs) and small RNAs. Transcriptomics allows to better understand the structure of genes and their transcripts; for instance their transcriptional start sites (TSSs), splicing patterns, 5′ and 3′ ends, and post-transcriptional modifications, in addition to measuring the expression changes of each transcript under various conditions. Nevertheless, the changes in mRNA levels are not always associated with changes at the level of proteomics as a result of splicing, post-transcriptional regulation and mRNA stability. However, transcriptomics approaches are more sensitive in the identification of less abundant transcripts and therefore, more sensitive tools for gene expression analysis, than proteomics approaches (Latchman, 2005, Lovric, 2011). Transcriptomics approaches are described in Table 1.1. As mentioned earlier, the transcriptomics studies are part of the systems biology approach which are closely interconnected with genomics and epigenomics in the control of gene regulation, which are discussed below.

	Technologies	Advantages	Disadvantages
Hybridisation-based	Tiling arrays	High throughput, high resolution, high genomic coverage, independent of gene annotations.	Previous knowledge of genome sequence, require plenty of RNA, require multiple microarrays for whole genome coverage, expensive, laborious.
	3'-IVT arrays	High throughput, covering most annotated genes.	Previous knowledge of genome, limited dynamic range of detection, cross-hybridisation, only transcripts with intact poly(A) tail, cannot distinguish alternative splicing or mutations.
	Gene/Exon arrays	High throughput, covering >99% of annotated genes, gene- or exon-level, alternative splicing, transcripts with unknown or truncated 3'-end, degraded RNA, random priming amplification, small amounts of RNA.	Require previous knowledge of genome, limited dynamic range of detection, cross-hybridisation.
Sequence-based	SAGE, CAGE, MPSS	High throughput, genome-wide, previous knowledge of genome is not required.	Expensive Sanger sequencing, require plenty of RNA, splice variants cannot be distinguished, only a portion of the transcript is analysed, laborious.
	RNA-seq	High throughput, genome-wide, previous knowledge of genome is not required, all types of transcripts, alternative splicing, chromosomal abnormalities and mutations, identification of novel transcripts.	Require plenty of RNA, relatively expensive, complex bioinformatics analysis, high sequencing depth for low-copy number transcripts.
Quantitative RT-PCR		Gene specific, high sensitivity, inexpensive.	Not genome-wide, previous knowledge of genome is required, normalisation varies depending on the reference genes.

Table 1.1: Various technologies used for gene expression profiling.
The summary of various transcriptomic technologies developed to date with their advantages and disadvantages is listed (Russell et al., 2009, Wang et al., 2009). IVT, *in vitro* transcription; SAGE, serial analysis of gene expression; CAGE, cap analysis of gene expression; MPSS; massively parallel signature sequencing; RT-PCR, real-time polymerase chain reaction

1.2 Transcribed regions of the human genome

By using comparative genomic and computational approaches after the completion of Human Genome Project, ~20,000-25,000 protein-coding genes were identified (IHGS, 2004). These protein-coding regions make up to less than 2% of the human genome. Nevertheless, it is believed that the majority of the genome is transcribed which is far beyond the scope of annotated genes. This phenomenon is known as pervasive transcription. This phenomenon was supported by hybridisation of the RNA content of various human tissues and cell lines on whole chromosome tiling arrays which identified 93% of the genomic bases in the transcripts of one cell type or another (Bertone et al., 2004, Cheng et al., 2005, Kapranov et al., 2002, Rinn et al., 2003). Therefore, it is suggested that a large proportion of the transcriptome contains novel ncRNAs.

The human transcriptomics analysis was part of the Encyclopaedia of DNA Elements (ENCODE) project which intended to identify and characterise the functional genomic elements within the human genome sequence (ENCODE,

2004). The pilot phase of study, which investigated 1% of the human genome, identified pervasive transcription in both gene-rich and gene-poor regions (Birney et al., 2007). Furthermore, in the second phase of the project, in which the complete human genome was examined across 15 cell lines, 62.1% and 74.7% of the human genome were found to be associated with either processed or primary transcripts, respectively (Djebali et al., 2012). This study also showed that genes have a tendency to express multiple isoforms simultaneously with an average of 10-12 isoforms per gene per cell type. Furthermore, coding transcripts were found to be significantly enriched in the cytoplasm, whereas the ncRNAs were predominantly found in the nucleus (Djebali et al., 2012). The expression of mRNAs and potentially ncRNAs are regulated by the non-transcribed regulatory elements in the genome, which provide sophisticated control for gene regulation.

1.3 Transcriptional regulatory elements

1.3.1 Core and proximal promoters

The core promoter is the region that is situated at the 5′-end of the gene that binds to the transcriptional machinery and includes the TSS. Many core promoters are characterised by the presence of the TATA box, which is the docking site for the TATA box binding protein (TBP) that is a subunit of the pre-initiation complex (PIC). The other less common genomic elements within the core promoters are the initiator element (Inr), downstream promoter element (DPE), downstream core element (DCE), general transcription factor IIB (GTF2B)-recognition element (BRE), and motif ten element (MTE), which serve as docking sites for various transcription machinery subunits (Gershenzon and Ioshikhes, 2005, Maston et al., 2006). Moreover, the proximal promoter is the upstream region adjacent to the core promoter. There is no universal consensus to distinguish the proximal promoters based on their length as they exhibit a huge degree of variation. These promoters serve as docking sites for various transcription factors (Maston et al., 2006).

The mammalian promoters are classified based on their sequence composition into promoters with high CpG dinucleotide frequency, also known as CpG-rich, and promoters with low CpG dinucleotide frequency, also known as CpG-poor (Carninci et al., 2006). The CpG-rich promoters contain multiple TSSs which are

referred to as 'broad' promoters. These promoters are typically found at the start of ubiquitously expressed and developmentally regulated loci (Carninci et al., 2006, Yamashita et al., 2005). On the contrary, the CpG-poor promoters are associated with a single TSS which are referred to as 'sharp' promoters. These promoters are associated with tissue-specific genes. Furthermore, the tissue-specific promoters harbour a TATA box, whereas the CpG-rich promoters lack TATA box (Carninci et al., 2006). Two subcategories of CpG-rich promoters have been identified: the promoters with one short CpG island that encompasses the TSS and the promoters with multiple large and gene body-extended CpG islands. The former class is associated with the ubiquitously expressed genes, while the latter is linked to the developmentally regulated genes (Akalin et al., 2009).

1.3.2 Enhancers

The discovery of enhancers was made by the observations that some genomic regions of SV40 tumour virus are capable of increasing the transcriptional activity of a human gene with a promoter (Banerji et al., 1981). The immunoglobulin heavy chain locus was the first human locus for which an enhancer was identified (Banerji et al., 1983). Enhancers typically contain transcription factor binding sites and therefore, are capable of transcriptional regulation which is independent of their orientation and distance with respect to promoter. A single promoter can be regulated by multiple enhancers which may operate at different cell types, in response to different external cues or at different times in the course of development (Maston et al., 2006, Remenyi et al., 2004). The functions of enhancers and proximal promoters are very similar, although the former can be located several hundred kilobase pairs upstream or downstream of the promoter or within the gene body (Blackwood and Kadonaga, 1998). The distally located enhancer is believed to be brought into the close proximity of the core promoter by forming a loop that is referred to as the DNA-looping model (Vilar and Saiz, 2005).

1.3.3 Silencers

Silencers are genomic elements that highly resemble the enhancers which induce transcriptional repression in their controlled locus. Silencers are also distally located and can function irrespective of their orientation or distance from promoter (Ogbourne and Antalis, 1998). Silencers serve as docking sites for a subset of transcription factors and cofactors that are referred to as repressors and

corepressors, respectively (Privalsky, 2004). Repressors can mediate their function in multiple ways; for instance by blocking the assembly of PIC (Chen and Widom, 2005), by inhibiting the binding of an activator to an adjacent enhancer (Harris et al., 2005) or by competing for the same site (Li et al., 2004).

1.3.4 Insulators

Insulators or boundary elements are important in blocking the function of neighbouring transcriptional regulatory elements which results in partitioning the genome into discreet transcriptional domains. Insulators may exhibit their function by blocking the enhancer-promoter interaction or the spread of repressive chromatin conformation. The activity of insulators is position-dependent, but orientation-independent (Maston et al., 2006, Recillas-Targa et al., 2002). The number of insulator elements in the human genome is not established, but the most well-studied insulator is the β-globin insulator in the human and chicken genomes (Felsenfeld et al., 2004, Li et al., 1999). Insulators serve as docking sites for the CCCTC-binding factor (CTCF) (Bell and Felsenfeld, 2000, Bell et al., 1999). It is believed that the insulators function as physical barriers between enhancers and promoters by forming chromatin loops as a result of interaction with other insulator elements in the genome which generates distinct domains of transcriptional activity (Maston et al., 2006).

1.3.5 Locus control regions (LCRs)

LCRs are responsible for controlling the transcriptional activity of a cluster of genes. LCRs consist of multiple genomic elements which include enhancers, silencers, insulators, and scaffold/matrix attachment regions (S/MAR). LCRs are key players in tissue specific transcriptional regulation which are copy number-dependent, but position- and distance-independent. LCRs act as binding sites for many transcription factors, repressors, coactivators and/or chromatin modifiers. Nevertheless, predominantly LCRs are associated with enhancer (Li et al., 2002, Maston et al., 2006). It is believed that LCRs also function via long-range looping mechanisms that bring them together with the core promoter (Tolhuis et al., 2002).

The summary of above transcriptional regulatory elements is schematically illustrated in Figure 1.1.
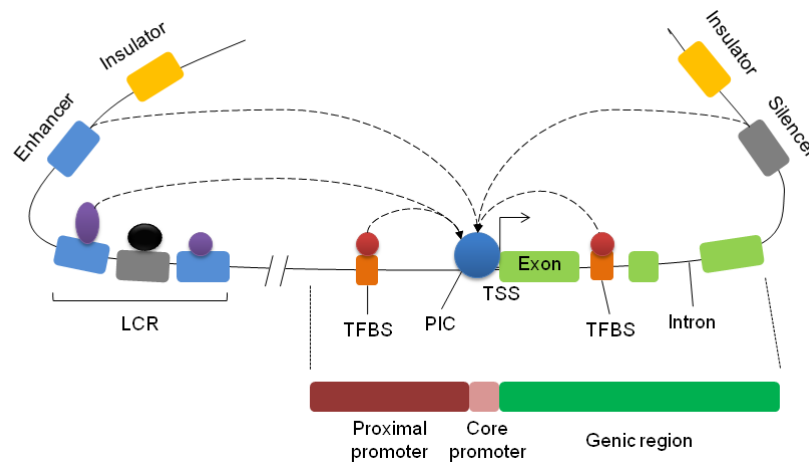
Figure 1.1: Transcriptional regulatory elements.
A schematic representation of the transcriptional regulatory elements. The binding of various transcriptional activators and repressors to the enhancers, silencers and LCRs regulate the assembly of PIC to the core promoter. The proximal promoter which is located adjacent to core promoter harbours transcriptional factor binding sites (TFBSs) which also induce PIC assembly upon transcription factor binding. The TFBSs can also be found downstream of the TSS.

1.4 Proteins involved in transcriptional regulation

As mentioned above, the transcriptional regulatory elements serve as docking sites for various *trans*-acting factors that induce the transcriptional activation or repression of the controlled locus. The proteins involved in transcriptional regulation can be classified into the following categories: (i) general transcription factors (GTFs) which are part of the basic transcriptional machinery and PIC; (ii) sequence-specific transcription factors; (iii) coactivators which are not sequence-specific and mediate the interaction between the sequence-specific transcription factors and the basic transcriptional machinery; and (iv) chromatin remodelling enzymes (discussed in Section 1.5).

1.4.1 GTFs

The GTFs are essential for the initiation of transcription from the core promoters. These factors include RNA polymerase II (RNAPII) and the GTF2 (also known as TFII) family of accessory proteins TBP, GTF2A, GTF2B, , GTF2E, GTF2F, and GF2H. The ordered assembly of these accessory subunits at the core promoters results in the formation of PIC which guides RNAPII to the TSS. PIC assembly initiates with the binding of TBP to the TATA box and interaction with a subset of TBP-associated factors (TAFs) which form a complex that is known as TFIID. Subsequently, RNAPII is recruited to the core complex and its carboxyl-terminal

domain undergoes phosphorylation that marks transcriptional elongation process. During transcriptional elongation, TFIID, GTF2E, GTF2H and a multi-subunit complex known as Mediator form a scaffold at the core promoter. In order to reinitiate transcription, GTF2F and GTF2B are required to recruit RNAPII to the core promoter (Hahn, 2004). The binding of general transcription factors can only induce low transcriptional activity or basal transcription, and requires the binding of sequence-specific transcription factors to the proximal promoter in order to stimulate transcription and also to mediate the recruitment and stabilisation of basic transcriptional machinery.

1.4.2 Sequence-specific transcription factors

It is estimated that there are ~1,400 genes coding for sequence-specific transcription factors (~6% of the protein coding genes) which are expressed in either all cell types or specific tissues (Vaquerizas et al., 2009). These sequence-specific factors are not part of the basic transcription machinery and also lack enzymatic activity, but can recruit or stabilise PIC and the subsequent initiation, elongation or reinitiation of transcription. Furthermore, these transcription factors can be important in recruiting the chromatin modifying machinery, as discussed below (de la Serna et al., 2005). The Gene Ontology (GO) analysis of the characterised transcription factors revealed that they are mainly associated with developmental regulation, cellular processes such as signalling, and response to stimulus such as immune response (Vaquerizas et al., 2009). Nevertheless, these GO classifications are fairly general and do not elucidate the precise function of these transcription factors. The classification can also be based on the structure of their DNA-binding domain which leads to the characterisation of unknown transcription factors based on their homology with the known transcription factors. This method of classification would allow understanding of DNA interaction and the binding site sequence of different clusters of transcription factors (Luscombe et al., 2000). The majority of sequence-specific transcription factors in human genome belong to the C_2H_2 zinc-finger, homeodomain, and helix-loop-helix families (Gray et al., 2004, Vaquerizas et al., 2009).

The comparison between the promoter sequences that are known to be bound by a specific transcription factor has resulted in the identification of consensus binding sites, which are referred to as motifs, for some transcription factors. The

consensus motifs are typically in the range of 6-12 basepairs (bp) (Wright and Funk, 1993). These consensus motifs are used in genome-wide bioinformatics approaches to identify all the potential binding sites for the transcription factor of interest (Elnitski et al., 2006, Morgan et al., 2007). However, the number of consensus motifs that can be identified by this approach is higher than the real binding sites for a transcription factor (Rabinovich et al., 2008). This study also suggested that despite the presence of motifs for some transcription factors such as RE1-silencing transcription factor (REST) and signal transducer and activator of transcription 1 (STAT1), the recruitment of some others is independent of a consensus motif such as the E2F family (Rabinovich et al., 2008, Robertson et al., 2007, Valouev et al., 2008). The transcription factors that are not linked to a consensus motif are suggested to bind DNA using alternative mechanisms such as forming a loop from a distal location to interact with another protein that is bound to its consensus motif which could be mediated by coactivators, binding to DNA through interactions with the DNA binding domain of another protein which is known as 'piggyback' binding, binding to the close proximity of another factor which binds to a consensus motif, or binding to a histone modification which can also be facilitated by coactivators, which is discussed in Section 1.5.3 (Farnham, 2009). Thus, the presence of consensus motifs is not a rule for DNA binding of transcription factors, and even the usage of consensus motifs at hypothetical binding sites should be verified in *in vitro* or *in vivo* assays.

1.5 The role of chromatin structure and modifications in transcriptional regulation

Human DNA is approximately 2 meters in length which has been packaged in a nucleus with only a few microns in diameter. This could only be achieved by the formation of a highly condensed and compact structure called chromatin. The chromatin structure is, therefore, responsible for providing access to the DNA-binding proteins such as transcription factors. As a result, all the metabolic activities involved in the maintenance of DNA and gene regulation, such as DNA replication, transcription, repair, recombination and chromosome maintenance, are regulated at the chromatin level. Therefore, the term epigenetics was first defined as the changes of gene function that can be heritably transmitted during mitosis and/or meiosis which are not associated with changes of DNA sequence (Russo et al., 1996, Turner, 1993). The significant technological advancements over the

recent years have enabled genome-wide analysis of changes in chromatin conformation and signature in different cell types and in the course of development and disease formation to better understand their role in gene regulation. Therefore, the roles of key chromatin modifiers and modifications are reviewed in this section (Figure 1.2).
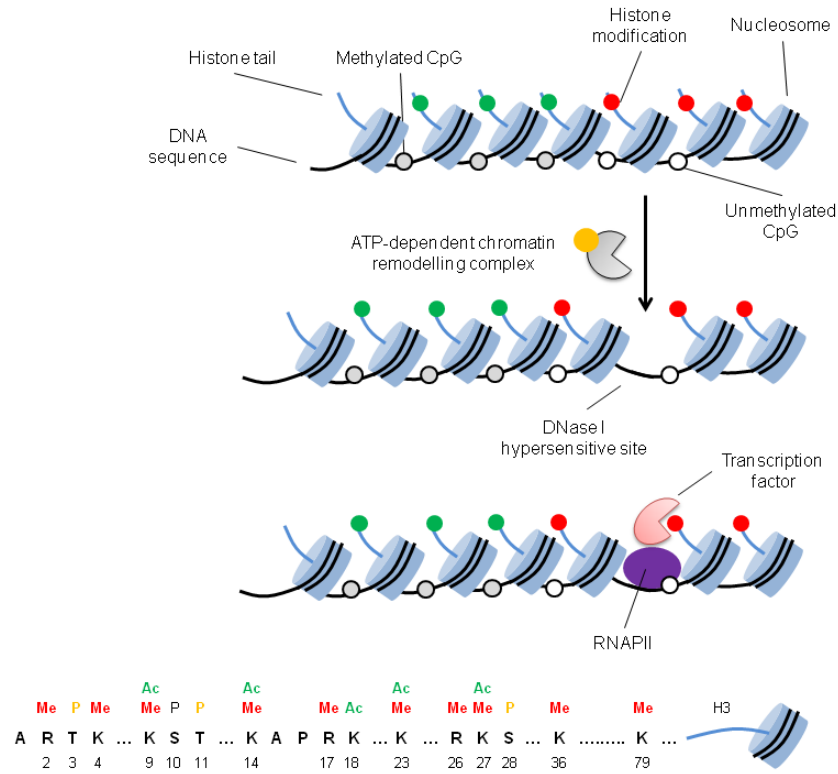
Figure 1.2: Different layers of chromatin modifications and organisation.
Different features of chromatin organisation are schematically illustrated. DNA methylation is associated with CpG dinucleotides and the histone modifications are biochemically added to the histone tails projecting out of the nucleosomes. The nucleosomes can be repositioned by the action of ATP-dependent chromatin remodelling complexes which provides accessibility for the binding of RNAPII and transcription factors. Bottom: the major histone H3 amino acid residues associated with methylation (Me), acetylation (Ac), and phosphorylation (P) marks are represented. The residues are also numbered. Some residues such as lysine 4 (K4) and lysine 27 (K27) can be associated with more than one type of modifications, but not simultaneously.

1.5.1 The structure of chromatin and nucleosomes

Nucleosomes are the primary units of chromatin structure which are composed of an octamer core of histone proteins that include H2A, H2B, H3, and H4, around which ~147 bp of left-handed DNA are wrapped in 1.65 turns (Luger et al., 1997). The core histones are highly conserved and are packed with positively charged arginine and lysine residues which enables them to favourably bind the negatively charged DNA template. Another histone protein, H1, is also present in the context of chromatin which serves as a linker histone that facilitates the compaction of

nucleosomes and chromatin structure. Core histones are composed of three major domains, which include the histone fold, the histone fold extension, and the histone tails. Nucleosomes are formed as a result of heterodimerisation of histones H2A/H2B and H3/H4 which is mediated by the hydrophobic interactions between the histone fold domains. The octamer core is, therefore, formed from a H3-H4 tetramer adjoined by two H2A-H2B dimers. On the other hand, the N-terminal tails of histones are not associated with a secondary structure and are found to be projected out of the nucleosome core and chromatin structure. The histone tails are between 16 to 44 amino acids in length and in some cases a C-terminal tail is also noticeable such as histone H2A (Luger et al., 1997). The projection of histone tails makes them accessible to a range of histone modifying enzymes that add post-translational modifications to various amino acid residues (Allfrey et al., 1964, Jenuwein and Allis, 2001, Kouzarides, 2007).

The chromatin structure comprises two major functional domains: euchromatin and heterochromatin (Henikoff, 2000, Felsenfeld and Groudine, 2003). Previous biochemical, cytological, and genetic studies have demonstrated that euchromatin is involved in active DNA processes, whereas heterochromatin is associated with repression or silencing as a result of blocking DNA accessiblity to transcription factors. The properties of heterochromatin are defined as compact and inaccessible, gene poor, late replicating, associated with large regions of repetitive DNA, and poorly associated with recombination events (Katan-Khaykovich and Struhl, 2005). The heterochromatic regions are mainly found at centromere, pericentromeric regions, and telomere (Karpen and Allshire, 1997). Heterochromatin formation is implied in various biological phenomena such as X chromosome inactivation in mammals (Lee and Jaenisch, 1997).

The chromatin accessibility can be studied by two major techniques: DNase I treatment and formaldehyde-assisted isolation of regulatory elements (FAIRE). It has been shown that the less compact euchromatin and the genomic regions that were subjected to nucleosomal remodelling and displacement can be digested by small amounts of DNase I (Stalder et al., 1980). DNase I-treated chromatin can be analysed by hybridisation to tiling microarrays (DNase-chip) or high throughput DNA sequencing (DNase-seq) which will identify genome-wide DNase I hypersensitive sites (Crawford et al., 2006b, Crawford et al., 2006a, Boyle et al., 2008). On the contrary, the FAIRE approach requires formaldehyde crosslinking of

the cells followed by sonication and phenol/chloroform extraction that would release the protein-unbound DNA fragments which can be further analysed by sequencing or microarray analysis (Giresi et al., 2007). Both of these methods are demonstrated to be efficient in identifying the open chromatin regions.

1.5.1.1 DNase I hypersensitive sites

DNase I hypersensitive sites are associated with transcriptional regulatory elements such as promoters, enhancers, silencers, insulators, and LCRs (Felsenfeld and Groudine, 2003, Gross and Garrard, 1988, Keene et al., 1981, McGhee et al., 1981). The pilot phase of ENOCDE project also revealed that TSSs, TFBSs, histone modification sites, and regions of early replication are associated with these hypersensitive sites (Birney et al., 2007). The above results indicate that the transcriptional regulatory elements are depleted from nucleosoms which therefore, can be used as markers for identifying such elements. The incorporation of the genome-wide analysis of hypersensitive sites and the histone modification signatures of the flanking nucleosomes could further elucidate the function of the identified regulatory elements. The completion of second phase of ENCODE project, which investigated the entire human genome, has identified approximately 2.89 million DNase I hypersensitive sites across 125 human cell and tissue types using DNase I-seq approach. The majority of these hypersensitive sites (95%) are located in intergenic or intronic locations and only 3% were found at the annotated TSSs (Thurman et al., 2012). The results also indicated that 95% of the bases in the human genome are in 8 kb proximity of DNase I hypersensitive sites, which was also shown to contain 99% of the known transcription factor binding motifs, as opposed to 1.2% of bases in the genome that are found within the protein-coding exons (Dunham et al., 2012). This data suggested that more genomic information is required for gene regulation than for biochemical activities.

1.5.2 Chromatin remodelling complexes

The access of transcription factors and basic transcription machinery to the regulatory elements requires conformational changes in chromatin structure, which can be achieved by physical repositioning, exclusion or inclusion of nucleosomes. This is achieved by the function of ATP-dependent chromatin remodelling complexes (Almer et al., 1986, Narlikar et al., 2001).

Five classes of remodelers have been identified which have differences in their mechanism of action and composition, which include the SWI/SNF complex, the ISWI complex, Mi-2/NURD complex, the INO80 complex and the SWR1 complex (Laurent, 2006). The SWI/SNF complex was first identified in yeast which was shown to be associated with gene activation (Winston and Carlson, 1992). This complex was also demonstrated to be important in human chromatin remodelling as well as being involved in transcriptional elongation, DNA repair and chromosome stability (Chai et al., 2005, Martens and Winston, 2003). These remodelers are capable of repositioning DNA around the nucleosome which in turn provides accessibility for transcription factors to bind DNA, as verified by the changes in the pattern of DNase I hypersensitivity (Cote et al., 1994, Kwon et al., 1994). The ISWI complex is composed of NURF (nucleosome remodelling factor), CHRAC (chromatin accessibility complex), and ACF (ATP-dependent chromatin assembly and remodelling factor) (Ito et al., 1997, Tsukiyama and Wu, 1995). ISWI family is involved predominantly in chromatin assembly as well as transcriptional elongation (Corona and Tamkun, 2004). Similar to SWI/SNF complex, the pattern of DNase I hypersensitivity can be altered by the function of ISWI complex which is indicative of chromatin accessibility; however, an increase in the frequency of transcription factor binding could not be established (Langst et al., 1999). Both SWI/SNF and ISWI complexes function by sliding nucleosomes along DNA instead of complete dissociation/reassociation of nucleosomes (Langst et al., 1999, Whitehouse et al., 1999).

1.5.3 DNA methylation

The methylation of DNA at cytosine C5 residues was first reported in vertebrates (Holliday and Pugh, 1975, Riggs, 1975), which brought the focus of research to understand the role of DNA methylation in biological processes and the mechanisms of maintenance and regulation of these marks. The mammalian DNA methylation has been mainly studied in the context of 5-methylcytosines (5mC) which are predominantly found within CpG sequences, although recent studies have confirmed the presence of non-CpG methylation with an unknown function (Lister et al., 2009).

CpG sequences are not evenly distributed in mammalian genomes. There are ~1 kb CpG-rich domains, also known as the CpG islands, which are associated with

more than half of the genes in the genome, while the rest of the genome is not associated with CpGs, also known as the CpG-poor regions. The role of DNA methylation in transcriptional regulation has been extensively investigated and it is believed that their position relative to transcriptional regulatory elements is key in determining their role in gene regulation (Jones, 2012). Except its essential role in transcriptional regulation, DNA methylation is also involved in genomic stability, the suppression of transposons, and DNA repair (Karpf and Matsui, 2005, Walsh et al., 1998, Walsh and Xu, 2006).

The analysis of DNA methylation at TSSs, which has been the primary focus of DNA methylation studies, showed that the methylated CpG islands suppress transcriptional initiation and therefore, DNA methylation can be regarded as a silencing epigenetic mechanism at the promoters (Hashimshony et al., 2003, Kass et al., 1997, Venolia and Gartler, 1983). With respect to the structure of promoters that was described earlier, most of the promoters associated with CpG islands are unmethylated in somatic cells and it was suggested that the transcriptional repression at these promoters are mediated by the Polycomb complex (Section 1.5.4.8) (Illingworth and Bird, 2009, Taberlay et al., 2011). Less than 10% of these promoters can become DNA methylated to confer a stable silenced state to the repressed genes such as imprinted and germline loci (Illingworth and Bird, 2009). However, tissue-specific changes of DNA methylation have been observed at the DNA sequences surrounding these CpG islands, also known as 'shores' (Irizarry et al., 2009). On the other hand, the CpG-poor promoters which are associated with tissue-specific genes are found to be methylated in the tissues and cell types that do not require their expression (Han et al., 2011).

Despite the extensive investigation of DNA methylation patterns at the TSSs, they can also be found at the gene bodies, which are predominantly CpG-poor. The DNA methylation patterns over gene bodies is known to be associated with active transcription (Hellman and Chess, 2007, Wolf et al., 1984). Therefore, promoter DNA methylation is inversely correlated with transcriptional activity, while the methylation over gene bodies is positively correlated with gene expression (Jones, 1999). Thus, as mentioned earlier, the position of DNA methylation with respect to the transcriptional regulatory elements is the key factor in determining the outcome of gene regulation.

DNA methylation is added and maintained by DNA methyltransferases (DNMTs), which include DNMT1, DNMT3A, and DNMT3B. The initial DNA methylation pattern during development is established by the *de novo* activity of DNMT3A and DNMT3B. DNMT1, on the other hand, favours hemi-methylated DNA which is an outcome of DNA replication and repair and therefore, is suggested to be involved in the maintenance of DNA methylation. However, it is now believed that the function of DNMT3A and DNMT3B is also crucial for the maintenance of DNA methyalation (Jones and Liang, 2009). The function of all three DNMTs is essential for embryonic development (Okano et al., 1999, Li et al., 1992) and the function of somatic cells (Jackson-Grusby et al., 2001); however, the absence of DNMTs in ES cells does not affect their self-renewal capacities (Tsumura et al., 2006). It is now understood that the mechanism of *de novo* methylation requires the presence of two molecules each of DNMT3-like protein (DMNT3L), DNMT3A and a nucleosome (Ooi et al., 2007). This study explained the reason behind why DNA methylation is not the initial silencing mechanism and normally confers a more stable silencing to a repressed promoter. Active TSSs are not occupied by a nucleosome, as described earlier, and therefore, cannot be targeted for *de novo* methylation. This observation was also supported by a recent study on the mechanism of DNA methylation at NANOG promoter in the course of embryonic differentiation (You et al., 2011).

The removal of DNA methylation from CpG dinucleotides is mediated by active or passive mechanisms. Active DNA demethylation is often associated with DNA base excision repair mechanisms to remove the 5mC residue. Demethylation enzymes are found in three major classes: ten-eleven translocation (TET) methylcytosine dioxygenase, activation-induced cytidine deaminase (AID), and thymine DNA glycosylase (TDG) (Cortellino et al., 2011, He et al., 2011, Popp et al., 2010). DNA demethylation by TET1 enzyme involves the formation of intermediate oxidation substrates, including 5-hydroxymethylcytosine (5hmC), 5-formylcytosince (5-fC), and 5 carboxylcytosine (5-caC). The importance of these demethylation derivatives in biological processes is largely unknown, except 5hmC which is shown to be associated with both transcriptional activating and silencing mechanisms. 5hmC is found at the promoters of transcriptionally poised genes, enhancers, and insulators (Wu and Zhang, 2011).

Several methods have been developed for the analysis of DNA methylation such as sodium bisulphite sequencing, methylated DNA immunoprecipitation (MeDIP), MBD-affinity column (MAC), and methylated CpG island recovery assay (MIRA). However, bisulphite sequencing and MeDIP are the two most frequently used methods. Sodium bisulphite treatment deaminates cytosine to uracil, unless the cytosine residue is methylated (Clark et al., 1994, Frommer et al., 1992). The combination of bisuphite treatment and deep sequencing has enabled genome-wide analysis of DNA methylations at single base resolution (Lister et al., 2009). MeDIP method, on the other hand, is an immunoprecipitation (IP) based method that uses an antibody which specifically identifies 5mC residues on the sonicated DNA fragments (Weber et al., 2005). MeDIP-purified DNA fragments can be subsequently coupled to microarray hybridisation (MeDIP-chip) or deep sequencing (MeDIP-seq), although single base resolution cannot be achieved.

1.5.4 Histone modifications

The histone tails are the main target of histone modifying enzymes which covalently add biochemical groups to the amino acid residues. 130 amino acid residues in the core histones are found to be associated with a histone modification (Tan et al., 2011). There are 14 known histone modifications, 10 of which have been further investigated to elucidate their biological function (Table 1.2) (Dawson and Kouzarides, 2012). Chromatin immunoprecipitation (ChIP) is the major method for the analysis of histone modification patterns that uses an antibody which specifically identifies the histone modification of interest. The ChIP-purified DNA fragments can then be subjected to microarray hybridisation (ChIP-chip) or deep sequencing (ChIP-seq), which are discussed in more details in Chapters 4 and 5. In this section, the biological role of 5 histone modifications in gene regulation will be discussed, which include acetylation, methylation, phosphorylation, ubiquitylation, and sumoylation.

Histone modification	Nomenclature	Biological function
Acetylation	K-ac	Transcription, repair, replication, and compaction
Methylation (lysine)	K-me1, K-me2, K-me3	Transcription and repair
Methylation (arginine)	R-me1, R-me2s, R-me2a	Transcription
Phosphorylation (serine and threonine)	S-ph, T-ph	Transcription, repair, compaction
Phosphorylation (tyrosine)	Y-ph	Transcription and repair
Ubiquitylation	K-ub	Transcription and repair
Sumoylation	K-su	Transcription and repair
ADP ribosylation	E-ar	Transcription and repair
Deimination	R→Cit	Transcription and decondensation
Proline isomerisation	P-cis↔P-trans	Transcription
Crotonylation	K-cr	Transcription
Propionylation	K-pr	Unknown
Butyrylation	K-bu	Unknown
Formylation	K-fo	Unknown
Hydroxylation	Y-oh	Unknown
O-GlcNAcylation (serine and threonine)	S-GlcNAc, T-GlcNAc	Transcription

Table 1.2: Histone modifications and their biological function.
14 different types of histone modifications are listed with their attributed biological functions (Dawson and Kouzarides, 2012, Tan et al., 2011). Me1, monomethylation; me2: dimethylation; me3, trimethylation; me2s, symmetrical dimethylation; me2a, asymmetrical dimethylation; Cit, citrulline.

Histone modifications can affect gene regulation via two major mechanisms: (i) through affecting the nucleosome-DNA and the nucleosome-nucleosome interactions as a result of altering histone charges or their physical entity that may cause spatial hindrance, and/or (ii) serve as docking sites for histone binding proteins or the 'readers' that orchestrate various biological outcomes. Nevertheless, transcriptional regulation is not the only biological implication of histone modifications, but also a variety of DNA-associated mechanisms such as replication, repair and condensation can also be regulated by these epigenetic marks (Kouzarides, 2007). Histone modifications are dynamic epigenetic marks which can represent different patterns in different cell types or in response to various stimuli and external conditions which therefore, provide a complex picture for their regulation and function. Histone modifications can be interpreted in a combinatorial manner in which the presence of different histone modifications is required for implementing a biological outcome such as transcriptional regulation. Furthermore, some histone modifications can affect the deposition and removal of other modifications or induce the activity of chromatin remodelling complexes. This

phenomenon is known as "histone crosstalk", also referred to as the "histone code", which has a significant biological implication on gene regulation and other biological processes (Jenuwein and Allis, 2001, Lee et al., 2010).

1.5.4.1 Sumoylation

Small ubiquitin-related modifier (SUMO) protein can covalently attach to the lysine residues of histone tails. It has been reported that the binding of SUMO will result in the recruitment of histone deacetylases (HDACs) as well as heterochromain protein 1 (HP1) which in turn results in the removal of acetylation and ubiquitylation marks that are associated with transcriptional activation (described below). Thus, sumoylation confers a repressed chromatin state and transcriptional suppression (Nathan et al., 2006, Shiio and Eisenman, 2003).

1.5.4.2 Ubiquitylation

Similar to sumoylation, ubiquitylation is also involved in the covalent attachment of ubiquitin protein to the lysine residues of histones. Monoubiquitylation results in changes of protein function, whereas polyubiquitylation targets the protein for degradation. Histones H2A, H2B, H3 and H1 are found to be monoubiquitylated (Goldknopf et al., 1975, West and Bonner, 1980). H2A ubiquitylation is well-characterised in the context of transcriptional regulation. Polycomb repressive complex 1 (PRC1) is responsible for mediating the ubiquitylation of H2A in response to the methylation of H3 lysine 27 (H3K27), which is a repressive mark (discussed below), to stably silence gene expression (Cao et al., 2005, Wang et al., 2004). On the other hand, the ubiquitylation of H2B is shown to be associated with gene activation (Minsky et al., 2008, Nickel et al., 1989). H2B ubiquitylation is associated with both TSSs and gene bodies and therefore, its role in transcriptional elongation is more profound than transcriptional initiation. The ubiquitylation of H2B is believed to be important in the recruitment of H3K4 and H3K79 methylases, which are associated with active transcription (discussed below) (Dover et al., 2002, McGinty et al., 2008). Thus, dependent on the position of ubiquitylated residues, this histone modification can be associated with both transcriptional activation and repression.

1.5.4.3 Phosphorylation

The phosphorylation of histones is catalysed by protein kinases which add a phosphate group (PO_4) to a serine, threonine, or tyrosine. The removal of phosphorylation is mediated by protein phosphatases. One of the well-characterised histone phosphorylations occurs at the histone H3 serine 10 (H3S10) residue which is linked to transcriptional activation (Nowak and Corces, 2000), whereas the removal of this mark by protein phosphatase 2A (PP2A) is associated with transcriptional repression (Nowak and Corces, 2004). The kinase signal transduction is therefore, not only important in post-translationally modifying the cytoplasmic proteins, but can also affect histone proteins, which indicates an important role for the signalling pathways that function via kinase activation cascades in regulating chromatin structure at the promoters and subsequently activating gene expression. The serine and threonine phosphorylations are bound by 14-3-3 family of proteins (Macdonald et al., 2005).

1.5.4.4 Acetylation

Histone acetylation is mediated by the addition of an acetyl group from acetyl coenzyme A (Acetyl-CoA) to the lysine residues of histone tails. This modification is catalysed by histone acetyltransferases (HATs) and can be removed by the function of HDACs. Lysine acetylation has been detected on the tails of all core histones. Histone acetylation is positively correlated with gene expression and recent genome-wide screenings have identified histone acetylation enrichments at the promoters and enhancers of transcribed genes (Heintzman et al., 2007, Wang et al., 2008).

One of the mechanisms through which histone acetylation can induce transcriptional activation is mediated by the changes of chromatin structure. The deposition of acetyl group on histone tails results in the neutralisation of histone charge that in turn weakens the interaction between histone and DNA template. As a result, the chromatin structure becomes less condensed and provides accessibility for the transcription factors and transcriptional machinery (Workman and Kingston, 1998). Additionally, histone acetylation can act as docking sites for transcriptional regulators that contain a bromodomain or tandem plant homeodomain (PHD) fingers, which include transcriptional coactivators, HATs, histone methylatransferases (HMTs), and chromatin remodelers (Chung and

Witherington, 2011, Taverna et al., 2007). The association of various histone modifiers with histone acetylation is indicative of their role in histone crosstalk which can be used as an additional mechanism of transcriptional regulation. two examples of this crosstalk are reported between phosphorylated H3S10 which recruits lysine acetyltransferase 2A (KAT2A/GCN5) in order to acetylate the H3K14 residue (Cheung et al., 2000, Lo et al., 2000), and between H3K14 acetylation and H3K4 trimethylation (H3K4me3), which is also an activating histone mark (discussed below) (Nakanishi et al., 2008).

HATs are classified based on their structure, which include GCN5-related acetyltransferases (GNATs), MYST, and E1A binding protein p300 (EP300)/CREB binding protein (CBP) (Sterner and Berger, 2000). EP300 and CBP are also two well-characterised transcription factors that are associated with the regulation of apoptosis and G1 arrest, induction of cell cycle inhibitors cyclin-dependent kinase inhibitor 1A (CDKN1A/p21) and CDKN1B/p27, and controlling differentiation (Kawasaki et al., 1998, Yuan et al., 1999). EP300 and CBP are linked to the acetylation of several histone residues, which include H2BK12, H2BK15, H3K14, H3K18, H3K56, H4K5, and H4K8 (Schiltz et al., 1999). The $Cbp^{+/-}$ mice were deficient in haemopoietic lineage differentiation, which indicated the essential role of CBP is haemopoietic development (Kung et al., 2000). It can be, therefore, proposed that abnormalities in the deposition of histone acetylation marks impair transcriptional activation of the genes that are required for lineage development and differentiation as was observed in haemopoietic development.

1.5.4.5 Methylation

Histone methylation is the result of covalently adding a methyl group from S-adenosylmethionine (SAM) substrates to the nitrogen atoms on the side chains of lysine and arginine residues. Lysine residues can be mono- (me1), di- (me2) or trimethylated (me3), whereas ariginine residues can only be mono- or dimethylated (Kouzarides, 2007). Histone lysine methylations can be associated with both transcriptional activation, such as H3K4, H3K36 and H3K79 methylations, and repression, such as H3K9, H3K27 and H4K20 methylations, in a context-dependent manner. Similarly, arginine monomethylation can be associated with both transcriptional activation and repression, whereas symmetric

arginine dimethylation is linked to repression and asymmetric arginine dimethylation is associated with activation (Di Lorenzo and Bedford, 2011).

The enzymes that catalyse histone methylations are referred to as HMTs which are classified into three categories: (i) SET domain-containing HMTs which specifically methylate lysine residues K4, K9, K27 and K36 of histone H3 as well as K20 of histone H4; (ii) HMTs that do not contain a SET domain which are linked to the methylation of H3K79 residue; (iii) arginine methylating HMTs which target R2, R17 and R26 residues of histone H3 as well as the R3 residue of histone H4 (Fyodorov and Kadonaga, 2001, Marmorstein, 2001). The histone methylations can be removed by the function of histone demethylases. Lysine demethylase 1A (KDM1A/LSD1) was the first identified lysine demethylase which can demethylate H3K4 and H3K9 (Shi et al., 2004). Other lysine demethylases are classified into two categories of Jumonji AT rich interactive domain 1 (JARID1) and Jumonji C-domain containing proteins (JMD2) families. The latter group consists of KDM4A and KDM4C, which demethylate H3K9 and H3K36; KDM4B and KDM4D, which demethylate H3K9; KDM2A, which demethylates H3K36; KDM2B, which demethylates H3K4 and H3K36; KDM6A/UTX and KDM6B, which demethylate H3K27 (Tsukada et al., 2006, Whetstine et al., 2006, Agger et al., 2007). The JARID1 family is composed of KDM5A, KDM5B, KDM5C and KDM5D, which are specific H3K4 demethylases (Christensen et al., 2007, Iwase et al., 2007). On the other hand, the demethylation of arginine residues is facilitated by histone deimination (Table 1.2), which is mediated by PADI4 deiminase (Cuthbert et al., 2004).

H3K36 methylation – H3K36me2/me3 are predominantly linked to active transcription and the levels of H3K36me3 marks are positively correlated with gene expression levels. The pattern of H3K36 methylation enrichment is found to be progressively moving from monomethylation at the promoter to trimethylation at the 3′-end of the gene with noticeable H3K36me3 enrichments over the expressed exons (Bannister et al., 2005, Kolasinska-Zwierz et al., 2009). H3K36me3 is demonstrated to be significantly associated with transcriptional elongation and the observations in yeast and human demonstrated an association between the H3K36 methylase, Set2, and elongating RNAPII (Kizer et al., 2005).

H3K79 methylation – H3K79me2/me3 are also enriched over the gene bodies of transcribed genes. H3K79 methylation distribution pattern over the active gene bodies shows significant enrichments at the 5′-end of the gene which is progressively reduced towards the 3′-end, as opposed to the H3K36me3 methylation pattern. H3K79 methylation is catalysed by DOT1L HMT (Barski et al., 2007, Steger et al., 2008).

H3K9 methylation – H3K9me3 marks have been linked to the formation of heterochromatin regions and therefore gene silencing (Richards and Elgin, 2002). This methylation is deposited by suppressor of variegation 3-9 homolog 1 (SUV39H1) and SUV39H2 enzymes (Tschiersch et al., 1994). The formation of heterochromatin is thought to be mediated by the binding of HP1 protein to H3K9me2/me3 marks. In addition to the formation of heterochromatin domains, the H3K9me3 enrichments at the gene level can be interpreted in a context-dependent manner in which the promoter enrichments are associated with transcriptional repression, whereas the gene body enrichments are linked to transcriptional activities (Vakoc et al., 2005). Furthermore, H3K9me1 marks are found at the promoters of active genes (Barski et al., 2007).

H3K4 methylation – The early studies of histone methylation distributions identified H3K4me3 marks at the promoter regions and TSSs of transcribed genes (Barski et al., 2007, Kim et al., 2005, Santos-Rosa et al., 2002). Moreover, H3K4me2 marks were mainly identified within the active gene bodies (Barski et al., 2007, Bernstein et al., 2002), and H3K4me1 marks were detected on active enhancers and active gene bodies (Barski et al., 2007, Heintzman et al., 2009). However, a more stringent analysis of H3K4me3 distribution at the genome-wide level indicated that this mark is associated with CpG-rich promoters regardless of their transcriptional activity (Guenther et al., 2007, Mikkelsen et al., 2007). H3K4me3 sites were also marked with histone acetylation and DNase I hypersensitive sites, which are involved in chromatin accessibility and transcriptional regulatory elements (Ernst and Kellis, 2010, Wang et al., 2008). Furthermore, H3K4me2/me3 sites were found to be inversely correlated with DNA methylation at the promoters (Meissner et al., 2008, Ooi et al., 2007).

Although the precise role of H3K4 methylation marks in transcriptional activation is not fully established, it has been proposed that GTFs, such as TAF3, which

contain a PHD domain can recognise H3K4 methylation at the promoters that in turn serve as a docking site for the assembly of PIC (Vermeulen et al., 2007). This has been further supported by the observation of RNAPII binding at the H3K4me3-asscoiated promoters (Guenther et al., 2007). Therefore, the H3K4me3-marked CpG-rich promoters are positively correlated with transcriptional initiation, but not necessarily with transcriptional elongation. Recent studies have demonstrated that the RNAPII at some H3K4me3-promoters are kept at a paused state which are released in response to activation signals that in turn results in transcriptional elongation (Hargreaves et al., 2009, Rahl et al., 2010). Thus, the H3K4me3-marked promoters can be classified as actively transcribed, which are associated with elongating RNAPII, or primed to be activated, which are associated with a stalled RNAPII at their TSSs (Zhou et al., 2011). H3K4me2 enrichments, on the other hand, have been found at the CpG-poor promoters of haemopoietic lineage-specific genes (Orford et al., 2008). These lineage-specific genes are repressed in HSCs and early progenitor cells, but can become activated in the course of differentiation which is associated with a switch from H3K4me2 to H3K4me3 at their promoters. Thus, H3K4me2 marks are suggested to be important in marking the CpG-poor tissue-specific promoters that can be induced in terminally differentiated cell types.

H3K4 methyltransferases – H3K4 methylation marks are deposited by the Trithorax group (TrxG) proteins. The TrxG complex responsible for H3K4 methylation in mammals is known as the complex proteins associated with Set1 (COMPASS)-like complex. Six non-redundant and essential COMPASS-related H3K4 methylases have been identified in mammals, which include SETD1A, SETD1B, mixed-lineage leukaemia (MLL), MLL2, MLL3, and MLL4 (Shilatifard, 2008). SETD1A and SETD1B proteins are associated with the majority of H3K4 methylation activity in the genome and therefore, they have an important role in global gene activation (Wu et al., 2008). In addition to the role of H3K4me3 in recruiting GTFs and inducing transcriptional initiation, the COMPASS-like complex can affect transcriptional regulation by recruiting other histone modifying enzymes that deposit the activating histone marks or remove the repressive modifications. For instance, MLL/MLL2 can recruit a HAT, lysine acetyltransferase 8 (KAT8), which acetylates H4K16 and induces open chromatin conformation (Dou et al., 2005). Moreover, MLL3/MLL4 can recruit KDM6A/UTX H3K27 demethylase to counteract with the repressive H3K27 methylation marks and induce

transcriptional activation (Agger et al., 2007). The above examples are therefore indicative of a histone crosstalk as an additional mechanisms of transcriptional activation by H3K4me3 marks (Figure 1.3).

The recruitment of TrxG complexes to chromatin in mammals is mediated by the CXXC domain of MLL family proteins or the CXXC1 subunit, which is associated with SETD1A and SETD1B HMTs (Thomson et al., 2010). The CXXC DNA binding domain identifies the unmethylated CpG dinucleotides which explains the association between H3K4me3 marks and the CpG-rich promoters.

TrxG proteins and ES cells – Several members of TrxG complex have been implicated in the maintenance of embryonic stem (ES) cells and during embryonic development, although it has been a challenging task to fully elucidate the role of six H3K4 methyltransferases in ES cells due to their partial redundancies. Nevertheless, the role of MLL has been widely investigated in embryonic and adult stem cells. MLL-knockout mice are embryonic lethal and are associated with reduced numbers of haemopoietic stem cells (HSCs) in their foetal livers (Gan et al., 2010, McMahon et al., 2007). Furthermore, MLL is essential for the self-renewal of haemopoiesis (Ernst et al., 2004) and its abnormal expression as a result of chromosomal translocations is associated with increased HSC self-renewal which in turn results in the formation of leukaemia (Krivtsov et al., 2006). A core component of the COMPASS-like complex, WDR5, has also been reported to have an essential role in ES cell self-renewal (Ang et al., 2011). WDR5 interacts with POU5F1/OCT4 in ES cells and targets the promoters of self-renewal genes. Thus, TrxG proteins have an essential role in stem cell maintenance through H3K4 methylation activities and/or engaging in a crosstalk with other histone modifications such as acetylation and H3K27 methylation, which is discussed below.

H3K27 methylation – H3K27me2/me3 marks are linked to transcriptional repression (Lee et al., 2006, Boyer et al., 2006). H3K27me3 marks are found at the promoters of repressed genes and are absent from the active promoters and transcribed gene bodies. On the contrary, H3K27me1 marks are found over the body of active genes with higher enrichments towards the 5′-end (Barski et al., 2007).

H3K27 methyltransferases – H3K27 methylations are deposited by the activity of Polycomb group (PcG) proteins that were first characterised in *Drosophila* as transcriptional factors involved in the suppression of Hox genes (Struhl, 1981). The PcG proteins are classified into two complexes known as PRC1 and PRC2. The H3K27 methylation is mediated by PRC2 complex, while PRC1 is important in the maintenance of repression in the target genes, which are mainly the developmentally regulated loci and transcription factors (Boyer et al., 2006, Bracken et al., 2006, Lee et al., 2006). The catalytic SET domain-containing subunits of PRC2 complex are enhancer of zeste homolog 1 (EZH1) and EZH2 which are activated through interactions with suppressor of zeste 12 homolog (SUZ12) and embryonic ectoderm development (EED) proteins (Cao and Zhang, 2004, Pasini et al., 2004). EZH1 and EZH2 subunits are not found in the same PRC2 complex. The PRC2-EZH1 complex is mainly expressed in adult stem cells and non-dividing cells, whereas the PRC2-EZH2 complex is predominantly expressed in proliferative cells and in the course of embryogenesis (Margueron et al., 2008, Shen et al., 2008). Furthermore, the PRC2-EZH1 complex possesses less H3K27 methyltransferase activity than the PRC2-EZH2 complex and can mediate transcriptional repression mainly by compacting the nucleosomes and altering chromatin conformation (Margueron et al., 2008). The PRC2 complex also contains other subunits which are involved in PRC2 stimulation, such as the Polycomb-like proteins (PCLs) (Nekrasov et al., 2007, Sarma et al., 2008), and histone binding, such as retinoblastoma binding protein 4 (RBBP4) and RBBP7 (Song et al., 2008).

The mechanism of PcG targeting in mammals is not fully understood; however, previous studies in *Drosophila* have reported the presence of Polycomb response elements (PREs) and recruiters that recognise PREs and subsequently recruit PcG complexes (Ringrose and Paro, 2007). PREs have not been identified in mammalian genomes, although several candidate recruiters have been proposed, such as yin and yang 1 (YY1) (Caretti et al., 2004) and POU5F1/OCT4 in embryonic stem cells (Endoh et al., 2008). Further investigation of the PRC1 and PRC2 binding sites identified 97% of the PcG targets to be located within the CpG-rich sequences (Ku et al., 2008), which are mainly associated with developmentally regulated promoters (Section 1.3.1). The identification of an EZH2 interacting partner, JARID2, and its association with CpG-rich domains suggested a potential role for JARID2 in the recruitment of PRC2 complex to the

CpG-rich promoters (Li et al., 2010). This hypothesis was further strengthened by the observation of significant overlap between the genome-wide binding regions of JARID2 and PRC2 complex (Pasini et al., 2010).

PRC2 roles in transcriptional repression – In addition to the methyltransferase activity, the PRC2 complex can also recruit a H3K4 demethylase, KDM5A (Pasini et al., 2008). This would in turn combine H3K4 demethylation with H3K27 methylation to confer a more repressed chromatin state. Additionally, the PRC2 complex is capable of recruiting DNMTs which can confer a more stable silencing to certain targets (Vire et al., 2006). HDAC enzymes have also been purified with PRC complexes that facilitate the removal of acetylation marks (Sparmann and van Lohuizen, 2006). Overall, the above observations demonstrate different mechanisms that the H3K27 methylation and PcG complexes exploit to confer transcriptional repression (Figure 1.3).

PRC1 roles in gene silencing – The PRC1 complex in mammals is associated with a significant diversity in terms of its composition and function. However, two core proteins have been identified in all human PRC1 complexes, which include BMI1 and ring finger protein 2 (RNF2). These core proteins are associated with histone H2A monoubiquitylation which is believed to be the major mechanism of suppression conferred by the PRC1 complex (Cao et al., 2005, Wang et al., 2004). This H2A monoubiquitylating complex comprised of other subunits, which include ring finger protein 1 (RING1), BCL6 corepressor (BCOR), Polycomb group ring finger 1 (PCGF1), and KDM2B (Gearhart et al., 2006, Sanchez et al., 2007). The PRC1-bound and H2A monoubiquitylated loci are associated with a paused RNAPII over the gene bodies, which indicate a block on transcriptional elongation. However, upon RING1 and RNF2 knockdown in ES cells, the H2A monoubiquitylation sites were depleted and the stalled RNAPII molecules resumed transcription (Stock et al., 2007).

The H3K27me3 marks provide docking sites for the PRC1 complex (Fischle et al., 2003, Min et al., 2003). Studies in *Drosophila* have suggested a PRC1-dependent mechanism for the maintenance and expansion of H3K27me3 marks which in turn results in transcriptional repression. In this model, the PRC1 interactions between both H3K27me3 marks and the PRE-bound PRC2 complex form a loop that brings the gene body and the promoter of the repressed gene to a close proximity which

in turn induces chromatin compaction (Schwartz and Pirrotta, 2007). The association of PRC1 and H3K27me3 marks in mammalian cells have also been demonstrated *in vivo* (Lee et al., 2007b, Mujtaba et al., 2008). These studies, therefore, indicate a mechanism of H3K27me3 repression via recruitment of PRC1 complex *in vivo*. Nevertheless, the PRC1 recruitment to chromatin is not dependent on PRC2 or H3K27 methylation (Schoeftner et al., 2006) and not all H3K27me3 domains recruit the PRC1 complex (Ku et al., 2008). Thus, the PRC1 interaction with H3K27me3 marks in maintaining a stable repression is target-dependent.

PcG proteins and ES cells – The role of PRC2 in the maintenance of ES cells is proposed in two different models. PRC2 complex can be required for the maintenance of pluripotency through suppressing developmentally regulated genes (Boyer et al., 2006), or it can promote differentiation by suppressing the pluripotency genes (Chamberlain et al., 2008, Pasini et al., 2007, Shen et al., 2008). There have been more evidence favouring the second model by demonstrating that the knockdown of PRC2 subunits does not affect the pluripotent state (Chamberlain et al., 2008, Pasini et al., 2007), whereas the knockdown of SUZ12 or JARID2 is linked to inefficient repression of pluripotency genes, NANOG and POU5F1/OCT4 (Landeira et al., 2010, Pasini et al., 2007). The *Eed$^{-/-}$* and *Ezh2$^{-/-}$* mouse ES (mES) cells displayed impaired mesoendodermal differentiation, while *Suz12$^{-/-}$* mES cells could not establish an endodermal layer upon differentiation (Pasini et al., 2007, Shen et al., 2008). Therefore, it is plausible that the PRC2 complex is required for lineage commitment and terminal differentiation. Furthermore, the knockdown of RNF2 and RING1 in ES cells was reported to impair ES cell maintenance and differentiation (Endoh et al., 2008). This data indicates that PRC1 also has an essential role in the maintenance of pluripotency and lineage commitment which can be independent of PRC2 function.
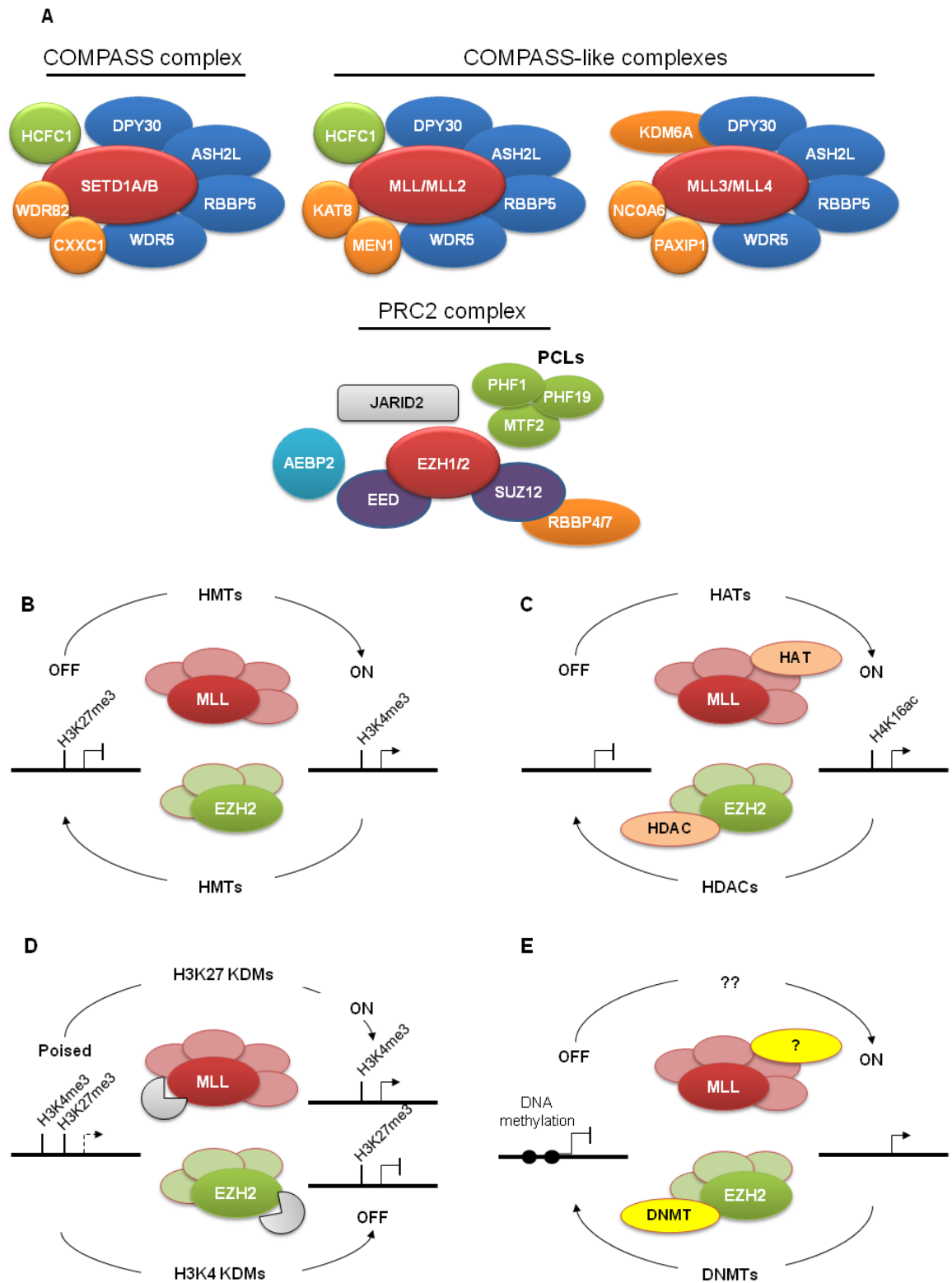
Figure 1.3: The composition of PcG and TrxG complexes and the mechanisms of transcriptional regulation by these complexes.
(A) The subunits of TrxG COMPASS and COMPASS-like and the PcG PRC2 complexes. The SET-domain containing catalytic subunits are illustrated in red. The common subunits between the COMPASS-like complexes are illustrated in blue. The mechanisms of gene regulation at the promoters by the PRC2 and TrxG complexes mediated via direct and indirect changes of chromatin modifications are illustrated. These mechanisms are not mutually exclusive. (B) Direct H3K27 methylation by EZH2, and direct H3K4 methylation by MLL; (C) the indirect regulation of histone acetylation by recruiting HDACs and HATs; (D) the indirect removal of H3K4 and H3K27 methylations by recruiting KDMs; (E) direct recruitment of DNMTs by PRC2 to DNA methylate certain targets, whereas TrxG targets and DNA methylation are antagonising.

1.5.4.6 The bivalent chromatin state

The genome-wide analyses of histone modifications in ES cells have identified H3K27me3-enriched gene promoters that are also associated with H3K4me3 marks and are, therefore, called bivalent domains (Azuara et al., 2006, Bernstein et al., 2006). The genes that have bivalent promoters in ES cells are associated with the developmentally regulated loci which are kept repressed in the pluripotent cells, but their fates are determined in the course of development. This can be achieved by resolving into a monovalent structure in the course of differentiation in which the resolution to H3K4me3 mark results in transcriptional activation and the resolution to H3K27me3 mark results in a more stable repression. Therefore, the chromatin conformation of bivalent promoters keeps them poised for fate determination at the later stages of differentiation (Bernstein et al., 2006, Mikkelsen et al., 2007). Intriguingly, the bivalent promoters are CpG-rich which can be explained by the preferential recruitment of TrxG and PRC2 complexes to the CpG-rich sequences, as discussed above (Mikkelsen et al., 2007, Pan et al., 2007, Zhao et al., 2007b). Bivalent promoters are found to be associated with RNAPII despite being transcriptionally inactive or expressed at very low levels which is indicative of transcriptional initiation block (Guenther et al., 2007, Stock et al., 2007). The expression of bivalent promoters is, therefore, under a balanced regulation mediated by both TrxG and PRC2 complexes, although the exact mechanisms of fate determination at these loci are not fully understood.

The analysis of chromatin conformation changes during *in vitro* neural differentiation from ES cells demonstrated a significant association between neural gene promoters and bivalent domains in ES cells. However, in neural progenitor cells the bivalent promoters of neural genes lost their H3K27me3 marks, but maintained their H3K4me3 marks which resulted in their transcriptional activation (Mikkelsen et al., 2007). This was proposed to be linked to KDM6B-mediated H3K27me3 demethylation at the bivalent loci, downregulation of EZH2 and changes of PRC2 targets in neural progenitor cells (Burgold et al., 2008, Lee et al., 2007a, Mohn et al., 2008). Furthermore, two accessory subunits of COMPASS-like complexes, RBBP5 and DPY30, were shown to be required for the bivalency resolution in the course of ES cell differentiation, whose knockdown abolished neural differentiation as a result of significant H3K4me3 reduction at the bivalent loci (Jiang et al., 2011). Therefore, in addition to the important role of PRC2

complex, TrxG complex also have an essential role in bivalency resolution via their core subunits during differentiation.

The different mechanisms of epigenetic regulation of transcription at the promoters that have been reviewed are schematically illustrated in Figure 1.4.
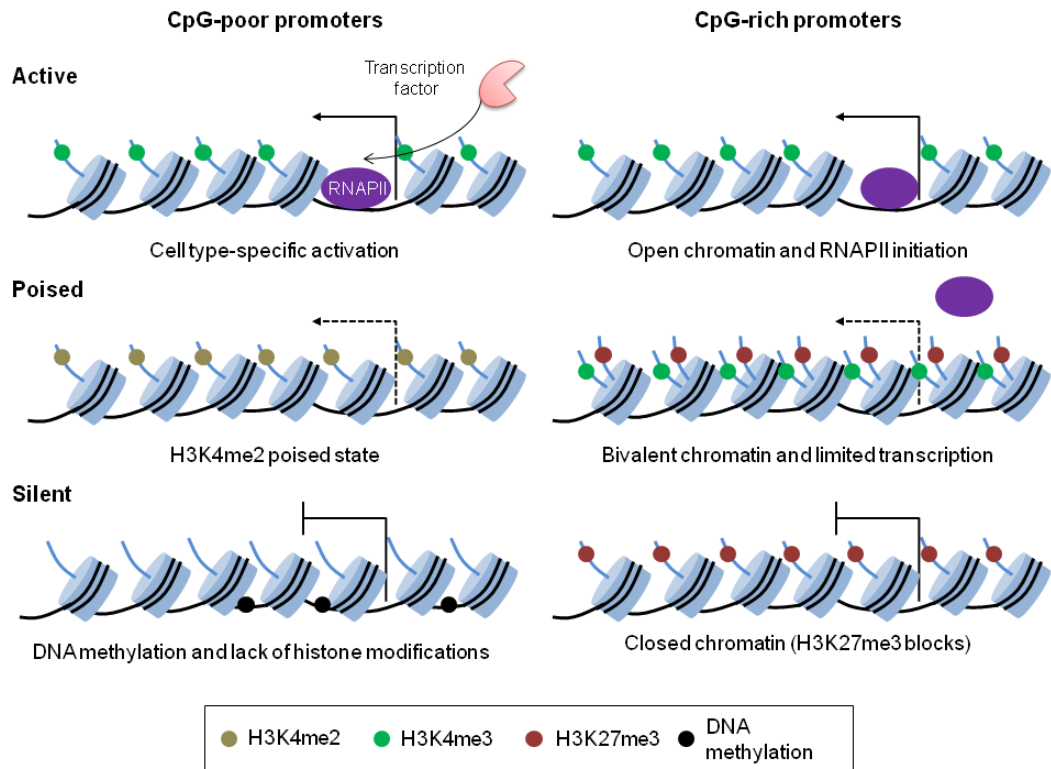
Figure 1.4: The epigenetic regulation CpG-rich and CpG-poor promoters.
A large subset of the CpG-rich promoters are ubiquitously expressed that are in open chromatin conformation at all times, which is defined by the presence of H3K4me3 marks and RNAPII at the TSS. Poised CpG-rich promoters are characterised by bivalent chromatin. RNAPII can be found paused at these promoters with basal transcriptional activity. Silenced CpG-rich promoters are marked with repressive H3K27me3 marks and are inaccessible to transcriptional machinery. The CpG-poor promoters can be activated in a cell type-specific manner and are marked with H3K4me3, whereas their poised promoters are inactive and marked with H3K4me2. The inactive CpG-poor promoters are marked with DNA methylation in the absence of histone marks.

1.6 Normal haemopoiesis

The systems biology approach can be applied to better elucidate the functional network of complex biological processes such as haemopoiesis. Haemopoiesis, the process of blood formation, is a hierarchical process that includes a defined origin, which is the HSC, and multiple intermediate levels of lineage potency. This process is precisely programmed to generate terminally differentiated lineage cells when needed. Thus, understanding of the molecular mechanisms that underlie the loss of pluripotency and subsequent lineage commitment and differentiation has

been the goal of ongoing studies in the field of haemopoiesis. The incorporation of genome-wide transcriptomics, proteomics and, recently, epigenomics have identified a series of key biological pathways and developmental regulators in haemopoiesis which are discussed below.

1.6.1 The haemopoietic hierarchy

The adult human bone marrow (BM) is capable of producing approximately 10^{12} cells daily and is therefore, one of the most regenerative tissues found in the human body. It has long been established that BM haemopoiesis is initiated from a common precursor, known as HSC, which gives rise to a hierarchy of various cell types. The existence of multipotent HSCs was initially reported by clonal *in vivo* repopulation assays (Becker et al., 1963, Till and Mc, 1961). HSCs are characterised by their long-term self-renewal capacities, but can also undergo differentiating cell divisions which give rise to the hierarchy of blood lineages. By this process a series of progenitor cells are formed in the intermediate stages before committing fate decisions in becoming mature blood cells. The haemopoietic hierarchy is comprised of two major branches: lymphoid and myeloid. The former is composed of T, B, natural killer (NK) and dendritic cells that possess adaptive and innate immune functions. Granulocytes (neutrophils, basophils, eosinophils, and mast cells), monocytes, megakaryocytes and erythrocytes are the short-lived terminally differentiated cells of the myeloid lineage. The earliest commitment decisions are made downstream of multipotent progenitors (MPPs), which only possess limited repopulation capacities (Adolfsson et al., 2001, Morrison et al., 1997). The commitment of MPPs results in the formation of common myeloid progenitors (CMPs) that initiate the myeloid branch and multilymphoid progenitors (MLPs) which initiate the lymphoid branch. MLPs have recently been defined as progenitors that can differentiate to all lymphoid lineages but may or may not exhibit other myeloid potentials (Doulatov et al., 2010), as opposed to the previously assigned common lymphoid progenitors (CLPs). CMPs then give rise to granulocyte/monocyte progenitors (GMPs), which in turn differentiate to granulocytes and monocytes and megakaryocyte/erythroid progenitors (MEPs), which give rise to erythrocytes and megakaryocytes that produce platelets. In the lymphoid branch, MLPs give rise to early thymic progenitors (ETPs) and B- and NK-cell precursors. This model is known as the

classical model of haemopoiesis (Figure 1.5) (Akashi et al., 2000, Kondo et al., 1997).
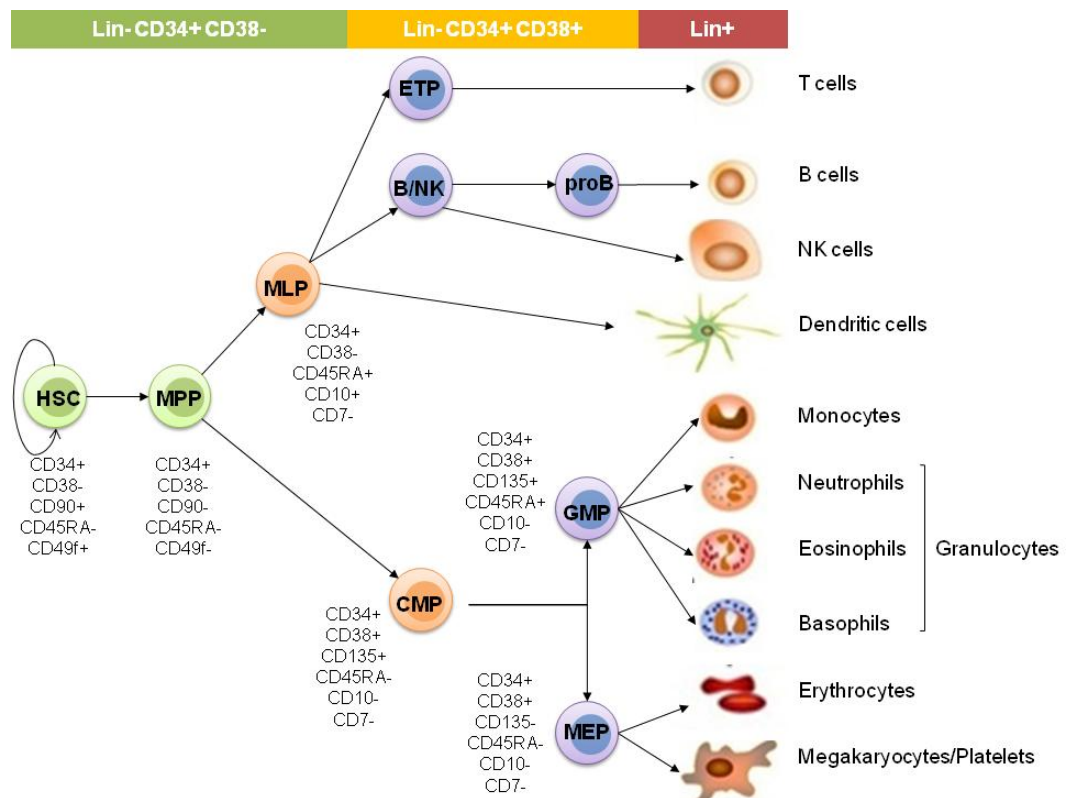
Figure 1.5: The lineage determination model in the adult human haemopoietic hierarchy.
The major subsets of haemopoietic cells in the adult human hierarchy are schematically illustrated. HSCs are the common precursors of haemopoiesis which are capable of self-renewal as well as giving rise to intermediate progenitors which drive the terminally differentiated mature blood cells, as listed on the right. HSCs, MPPs, and MLPs are characterised as lineage$^-$ (Lin$^-$) CD34$^+$ CD38$^-$ in addition to the other listed cell-specific surface markers (Doulatov et al., 2010, Doulatov et al., 2012). CMPs, GMPs, MEPs, pre-B and pre-T cells are characterised as Lin$^-$ CD34$^+$ CD38$^+$ in addition to other listed cell-specific markers. The terminally differentiated cells are defined as Lin$^+$ subsets. B/NK, B- and NK-cells precursor.

1.6.2 The origin of HSCs

In order to better understand the hierarchical model of haemopoiesis, it is essential to elucidate the origin of adult HSCs. During embryonic development, haemangioblasts act as the common precursors of blood and endothelium development (Choi et al., 1998). These cells are located within the mesodermal region of yolk sac, which is the first site of haemopoiesis in embryo (Huber et al., 2004). Intriguingly, although the haemopoiesis hierarchy in embryo is different from the adult mechanism, they are both originated from an FLK1$^+$ mesodermal population, which indicates common developmental mechanisms (Lugus et al., 2009). Upon transferring the haemopoiesis site to the aortic-gonado-mesonephrons (AGM) region in early embryogenesis, the haemogenic

endothelium cells, which express vascular endothelial (VE)-cadherin, become the new cellular origin of haemopoiesis (Boisset et al., 2010, de Bruijn et al., 2002). The haemogenic endothelial cells migrate to the foetal liver and subsequently to the BM where they repopulate and exhibit self-renewal and multilineage differentiation capacities (Zovein et al., 2008). This AGM origin of adult HSCs is confirmed by cell-fate tracing experiments in a murine model (Ivanovs et al., 2011). However, a recent report demonstrated both mesenchymal and endothelial origins for HSCs (Rybtsov et al., 2011). These observations, therefore, hypothesise two models of origin for HSCs. The first model suggests that mesenchymal cells migrate through the endothelial layer of AGM in order to form haemopoietic clusters, whereas the second model proposes that mesenchymal cells undergo an intermediate endothelial phase before developing HSCs (Lancrin et al., 2009). Thus, the embryonic origin of adult HSCs is the underlying reason for their multipotential capacities which can give rise to different haemopoietic lineages upon receiving external cues.

1.6.3 Characterisation and isolation of HSCs

The current knowledge of haemopoiesis is largely drawn from functional repopulation assays performed in mice due to the obvious limitations of analysing human HSCs. Nevertheless, the development of *in vitro* clonal assays, xenotransplantation methods and precise sorting strategies have significantly improved the understanding of human haemopoiesis (Doulatov et al., 2012).

1.6.3.1 Characterisation of HSCs

The initial studies in understanding haemopoiesis were directed towards the generation of colony-forming progenitors *in vitro* by colony-forming unit in culture (CFU-C) assays, in which stromal feeder layers were used to promote the formation of CFU-C colonies from BM cells (Sutherland et al., 1989). The precursor of CFU-Cs was identified as long-term culture-initiating cells (LTC-ICs). By using cytokine-producing stroma, the LTC-IC assays could promote multilineage differentiation over longer periods (Sutherland et al., 1991). Although LTC-ICs were multipotent precursors, they exhibited differences repopulation potential in comparison with HSCs.

The requirement for a more robust *in vivo* model to study haemopoiesis resulted in the development of immune-deficient mice that could be used for human haemopoietic cell engraftment assays. Severe combined immune-deficient (*Scid*) mice were the first humanised mouse models that were immune-compromised due to the absence of B and T cells (Bosma et al., 1983, Fulop and Phillips, 1990). The sublethally irradiated *Scid* mice were transplanted with human BM in the presence of interleukin (IL)-3, granulocyte-monocyte colony-stimulating factor (GM-CSF) and stem cell factor (SCF) (Kamel-Reid and Dick, 1988, Lapidot et al., 1992). The generation of myeloid progenitors and the production of B cells revealed successful long-term and multipotent engraftment of HSCs. This technique was the first quantitative *in vivo* assay that could prove that human HSC had been isolated and measure their activity. Further advancements in the success rate of engraftment assays were achieved by developing nonobese diabetic (NOD)-*scid* mice that were defective in their innate immunity (Shultz et al., 1995). In recent years, a new generation of *in vivo* models have been introduced which are mutated in their IL-2 receptor (IL-2R) gene that in turn results in complete loss of adaptive and innate immunity and, consequently, a 5-fold increase in $CD34^+$ cell engraftment (Ito et al., 2002, Shultz et al., 2005). Overall, the development of robust *in vivo* models has improved the characterisation of human HSCs in the recent years.

1.6.3.2 Isolation of murine HSCs

Murine HSCs were initially characterised as a lineage-negative (Lin^-), $Sca-1^+$, c-Kit^+ (LSK) population (Ikuta and Weissman, 1992, Spangrude et al., 1988). The $CD34^-$ subset of this population is associated with long-term self-renewal and multilineage development (Osawa et al., 1996). One in three $CD34^-$ LSK cells is identified as a Long-term HSC (LT-HSC) which is characterised by the expression of $CD150^+$ $CD48^-$ signalling lymphocytic activation molecule (SLAM) phenotype (Kiel et al., 2005).

1.6.3.3 Isolation of human HSCs

The isolation and purification of human HSCs is another obstacle in the path of their characterisation. In human BM only 1 in 10^6 cells is a transplantable HSC, which is surrounded by a pool of differentiated cells (Wang et al., 1997). LT-HSCs are defined as the cells that are capable of repopulation beyond 12 weeks in the

mouse model, whereas short-term HSCs (ST-HSCs) and MPPs are defined as the populations that can give rise to all lineages but possess transient engraftment capacities (Doulatov et al., 2012).

CD34 was identified as a key marker of human HSCs and progenitors, which is expressed in less than 5% of all blood cells (Civin et al., 1984, Vogel et al., 2000). Human HSCs do not express CD150, unlike their murine counterparts, but express the Fms-like tyrosine kinase-3 (FLT3) receptor which is absent in mouse (Larochelle et al., 2011, Sitnicka et al., 2003); thus, suggesting a fundamental difference between the haemopoietic markers in humans and mice. Since the CD34 antigen was also found on the surface of proliferating cells, more primitive HSCs were further characterised as $CD34^+CD38^-CD90^+CD45RA^-$ (Bhatia et al., 1997, Lansdorp et al., 1990, Murray et al., 1995). The loss of CD90 expression in HSCs was shown to be associated with cells involved in short-term engraftment or MPPs (Majeti et al., 2007), although they were still capable of serial transplantation and thus, were not fully resolved from HSCs. The cell-anchoring integrin molecules have recently been used as additional markers to distinguish true HSCs from their less primitive progeny. The CD49f integrin marker is found on 50% of the $CD90^+$ and 25% of the $CD90^-$ cells. Moreover, the $CD49f^+$ cells in both subsets were demonstrated as the population with long-term stem cell function, while the $CD49f^-$ cells represented transiently repopulating MPPs (Notta et al., 2011).

1.6.4 The intra- and extracellular regulators of HSCs self-renewal and quiescence

An HSC can undergo symmetrical or asymmetrical cell divisions with potentially three outcomes: two stem cells (self-renewal), a stem cell and a differentiated cell (homeostasis), or two differentiated cells (stem cell depletion). Such cell fate decisions are generally driven by gene expression changes which are regulated by transcription factors in response to external signals; however, the fate of early divisions in stem cells may be determined simply by the asymmetric distribution of fate determinants in the absence of transcriptional changes (Neumuller and Knoblich, 2009).

By using bromodeoxyuridine (BrdU) labelling in mice, it was estimated that LT-HSCs divide once every 30-50 days and are therefore, very quiescent (Cheshier et

al., 1999, Kiel et al., 2007). Nevertheless, even more quiescent HSCs have been identified in mice that only divide approximately five times in their lifetime (Foudi et al., 2009). These quiescent cells were identified by tracking cell divisions for long periods using green fluorescence protein (GFP)-tagged histone H2B as a chromatin marker, which revealed that the highest repopulation capacity resides with the cells that did not divide for over 200 days. Nevertheless, the highly quiescent cells, which constitute ~15% of the LT-HSC population, are believed to supply haemopoiesis in the case of injury, but are not thought to be involved in the daily production of blood lineages (Wilson et al., 2008). The presence of such dormant HSC populations in human BM could not be verified directly due to limitations in the application of label-retaining-cell methods, but flow cytometry analysis has demonstrated that the bulk of human HSCs reside in G_0 (Doulatov et al., 2012). The evaluation of mean telomere length or X chromosome inactivation ratios in mature blood cells has been used to estimate the cell division rate of human HSCs as once in every 175-350 days (Catlin et al., 2011, Shepherd et al., 2004). These observations in humans and mice have revealed less frequent HSC divisions in human BM, whereas the total number of cell divisions throughout the lifespan of both organisms appears to be similar. Therefore, HSCs reduce the risk of accumulating DNA damage as a result of oxidative and replicative stress by infrequent cycling (Doulatov et al., 2012). Therefore, in this section the role of various intrinsic and extrinsic factors in the maintenance of self-renewal and quiescence of HSCs will be reviewed.

1.6.4.1 Cyclin-dependent kinase inhibitors

The most intrinsic regulators of quiescence are the cyclin dependent kinase inhibitors (CDKIs) which act as checkpoints that monitor the passage of cells through different stages of the cell cycle. The two major subgroups of CDKIs are known as the Cip/Kip and the Ink4 families. The former consists of CDKN1A/p21, CDKN1B/p27, and CDKN1C/p57 proteins, whereas the latter comprises CDKN2A/p16, CDKN2B/p15, CDKN2C/p18 and CDKN2D/p19. The subfamilies are distinguished as a result of specific preferences for certain cyclin-dependent kinases (CDKs) they suppress. Cip/Kip family specifically target CDK2, whereas CDK4 and CDK6 are inhibited by the Ink4 family (Harper et al., 1995, Sherr and Roberts, 1995). Although the CDKIs are regulated at the post-transcriptional level, CDKN1A/p21 and CDKN1B/p27 are also under transcriptional regulation (Bloom

et al., 2003, Gartel and Tyner, 1999, Pagano et al., 1995). Elevated levels of CDKN1A/p21 are specifically associated with primitive, quiescent HSCs and deletion of CDKN1A/p21 in murine BM results in the proliferation of HSCs (Cheng et al., 2000). Similarly, the knockdown of CDKN1A/p21 in CD34$^+$ and CD34$^+$ CD38$^-$ cord blood cells resulted in their release from the quiescent state and subsequent expansion, both *in vitro* and *in vivo* (Stier et al., 2003). Similar outcomes are observed with the knockdown of CDKN1B/p27 (Dao et al., 1998). Moreover, two recent studies provided evidence for the involvement of CDKN1C/p57 in the maintenance of HSCs and their quiescence as the knockout mice showed reduced number of HSCs with impaired repopulation capacities upon transplantation, as well as a reduction in the G_0 population. The CDKN1A/p21 and CDKN1C/p57 double-knockout mice displayed more severely affected haemopoiesis than the CDKN1C/p57 knockout alone. The abnormalities in CDKN1C/p57 knockout mice could be corrected by the overexpression of CDKN1B/p27; thus, proposing functional similarities between CDKN1A/p21, CDKN1B/p27 and CDKN1C/p57 and the importance of the Cip/Kip CDKIs in the maintenance of HSCs self-renewal and quiescence (Matsumoto et al., 2011, Zou et al., 2011).

1.6.4.2 Cytokine growth factors

The intracellular activities of CDKIs and other transcriptional regulators are controlled by cell surface molecules responsible for transferring external signals. Cytokines, such as erythropoietin (EPO), granulocyte colony-stimulating factor (G-CSF), and GM-CSF were the very first group of external signals that were identified to stimulate the colony-formation potentials of haemopoietic cells (Bernstein et al., 1991, Haylock et al., 1992, Ozer et al., 2000). Nevertheless, cytokines displayed a pleiotropic effect on HSCs, largely in favour of differentiation rather than self-renewal and even in the presence of IL-11, SCF, and FLT3 ligand cocktail only a small expansion of murine HSCs was achievable (Attar and Scadden, 2004, Miller and Eaves, 1997).

Repopulating murine HSCs express the receptors for SCF and thrombopoietin (TPO), c-Mpl and c-Kit, respectively (Ikuta and Weissman, 1992, Kimura et al., 1998). Mice with genetic abnormalities in TPO and c-Mpl showed a significant decrease in their HSC numbers (Solar et al., 1998, Kimura et al., 1998).

Furthermore, TPO increased the survival of HSCs through reduction of apoptosis (Pestina et al., 2001). Overall, TPO and SCF are believed to be the most important of the classic haemopoietic cytokines in the regulation of stem cell features in HSCs.

1.6.4.3 Notch signalling pathway

Notch signalling is crucial in the fate determination of HSCs and in lymphopoiesis where it promotes the development of T cells (Radtke et al., 1999). Notch is cleaved upon binding to its extracellular ligands, Delta and Jagged, and its intracellular domain translocates to the nucleus where it functions as a transcription factor (Figure 1.6A) (Pajcini et al., 2011). Primitive HSCs and the surrounding niche cells express both Notch and Jagged (Karanu et al., 2000, Milner et al., 1994). The presence of Delta and Jagged resulted in the increased expansion of HSCs *in vitro* (Karanu et al., 2000, Karanu et al., 2001). One of the Notch targets, hairy and enhancer of split 1 (HES1), was found to be expressed in the $CD34^+$ $CD38^-$ population (Shojaei et al., 2005) and its overexpression was associated with increased repopulation capacities of HSCs. An extracellular ligand of Notch, nephroblastoma overexpressed gene (NOV), was shown to increase the expression of HES1, which in turn enhanced the stem cell activities of HSCs (Sakamoto et al., 2002).

Osteoblasts are reported to increase the expression of Jagged in response to activation of parathyroid hormone (PTH) receptor and subsequently induce the self-renewal capacities of HSCs as measured by the LTC-IC and repopulation assays (Calvi et al., 2003). Thus, indicating an important role for the Jagged/Notch signalling pathway in the regulation of HSC self-renewal. However, the normal repopulation of *Notch1*$^{-/-}$ HSCs into a *Jagged1*$^{-/-}$ microenvironment suggested a dispensable role for Notch signalling *in vivo* (Mancini et al., 2005), perhaps due to redundancy between the various Notch receptors and ligands.

1.6.4.4 Wnt signalling pathway

The action of canonical Wnt signalling is mediated by activation of its G protein-coupled receptor (GPCR), frizzled, which subsequently stabilises its intracellular effector, β-catenin, causing the nuclear translocation of β-catenin that forms a transcriptional complex with T cell-specific transcription factors (TCFs) and

lymphoid enhancer-binding factor 1 (LEF1) to target certain genes (Figure 1.6B) (Bienz, 1998). Canonical Wnt signalling was identified in LSK cells and β-catenin-transduced LSKs showed improved survival and 100-1,000-fold expansion *in vitro* (Reya et al., 2003). The same study also demonstrated elevated levels of self-renewal regulators Notch1 and HoxB4. Furthermore, through downregulation of Wnt signalling by using the soluble form of frizzled that competes with the membrane-bound receptor or by the overexpression of axin, which degrades β-catenin, HSC growth was significantly affected both *in vitro* and *in vivo,* indicating an essential role for Wnt signalling in the maintenance of HSCs (Reya et al., 2003). Nevertheless, it was reported that the overexpression or deletion of β-catenin *in vivo* did not result in a significant change of phenotype with regard to HSC number and function (Cobas et al., 2004, Kirstetter et al., 2006). The controversial findings may propose a redundancy between canonical and non-canonical Wnt signalling pathways in the maintenance of HSCs *in vivo.*

The self-renewal and repopulation potential of adult murine HSCs can be induced upon Wnt3a treatment, both *in vitro* and *in vivo* (Willert et al., 2003). This observation was further supported by showing that the Wnt3a stimulation can transcriptionally activate Notch target genes. Therefore, it is plausible to propose a synergistic mechanism of action between Wnt and Notch signalling pathways (Duncan et al., 2005). Both signalling pathways induce self-renewal of HSCs, but through different mechanisms, in which Wnt ligands induce the survival and expansion of LSK cells and Notch signalling maintains the immature state of stem cells. Finally, blocking the activity of Wnt intracellular effector, glycogen synthase kinase (GSK) 3β, impaired HSC self-renewal through misregulation of both Notch and Wnt target genes (Trowbridge et al., 2006).
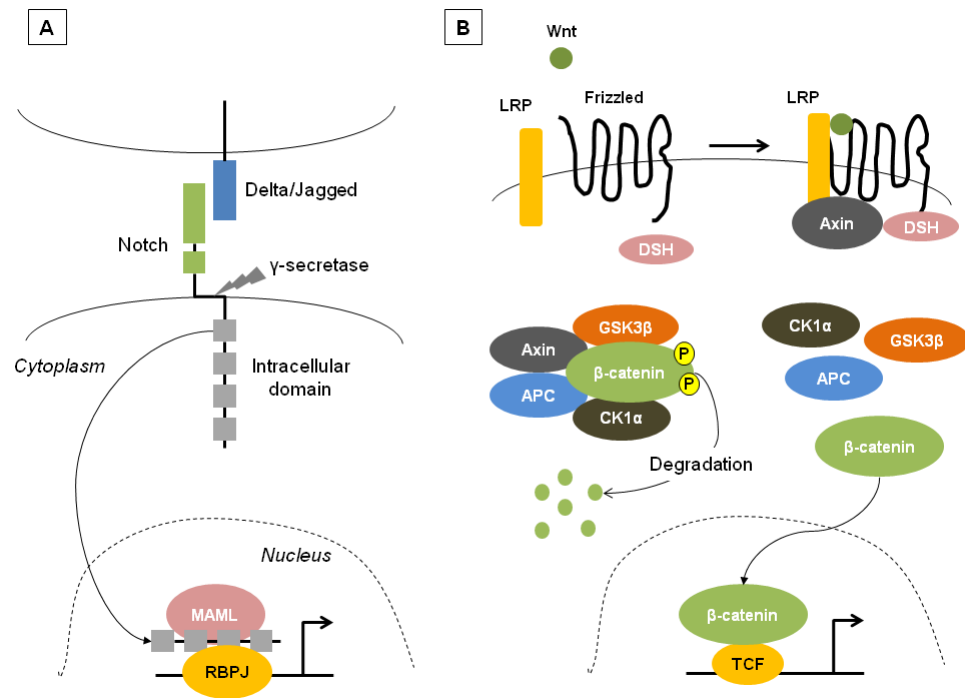
Figure 1.6: Notch and Wnt signalling pathways.
(A) Schematic mechanism of action in the Notch signalling pathway. The membrane-bound Notch receptor is cleaved by γ-secretase upon binding to its membrane-bound ligands, Delta (Delta 1-3) and Jagged (Jagged 1-2), which releases the intracellular domain of Notch that can translocate to the nucleus and activate target genes in the presence of Mastermind-like (MAML) cofactors and RBPJ transcription factor (Blank et al., 2008). (B) Schematic mechanism of action in the canonical Wnt signalling pathway. In the absence of Wnt-frizzled interaction intracellular β-catenin signalling is inactive due to its degradation by a complex of proteins including axin, GSK3β, adenomatous polyposis coli (APC), and casein kinase 1-α1 (CK1α). The binding of Wnt ligands to frizzled results in the recruitment of lipoprotein receptor-related protein (LRP) and subsequent recruitment of axin and dishevelled (DSH) to the cell membrane. Consequently, β-catenin is no longer degraded and can activate target genes in the presence of TCFs and LEF1 (Clevers, 2006). P indicates phosphorylation.

1.6.4.5 Transforming growth factor-beta (TGF-β) and Smad signalling pathways

The Smad pathway is an intracellular signal transduction network downstream of TGF-β, bone morphogenetic proteins (BMPs), activins, and Nodal superfamily of ligands and receptors (Figure 1.7), which regulate many biological processes during development. TGF-β is the master regulator of HSC quiescence (Batard et al., 2000, Sitnicka et al., 1996) and inhibition of TGF-β signalling triggers proliferation of HSCs (Hatzfeld et al., 1991). Nevertheless, the deletion of TGF-β receptor type I (TGFBR1) did not impair HSC self-renewal and repopulation capacities *in vivo*, even under stress conditions (Larsson et al., 2003, Larsson et al., 2005). This observation might be explained by the existing redundancy between the ligands and receptors of the TGF-β superfamily, all of which can activate the downstream Smad signalling pathway.

The activated type I receptor phosphorylate and activate SMAD2/SMAD3 intracellular effectors that can make a complex with SMAD4 proteins. Subsequently, this complex translocates to the nucleus and act as a transcription factor that regulates the expression of certain target genes (Heldin et al., 1997). The growth inhibitory function of TGF-β is mediated by the upregulation of CDKIs, such as CDKN1A/p21 and CDKN1C/p57, or by changing the expression of cytokine receptors (Dubois et al., 1994, Ducos et al., 2000, Jacobsen et al., 1991, Scandura et al., 2004). A novel mechanism of action for TGF-β/Smad signalling has been recently suggested involving transcriptional intermediary factor 1γ (TIF1γ) that can compete with Smad4 for binding Smad2/Smad3. Upon TGF-β stimulation of progenitors, the TIF1γ/Smad2/Smad3 complex can induce erythroid differentiation, while the Smad4/Smad2/Smad3 complex causes growth arrest, suggesting the balance between Smad4 and TIF1γ may predict the outcome of TGF-β stimulation (He et al., 2006).
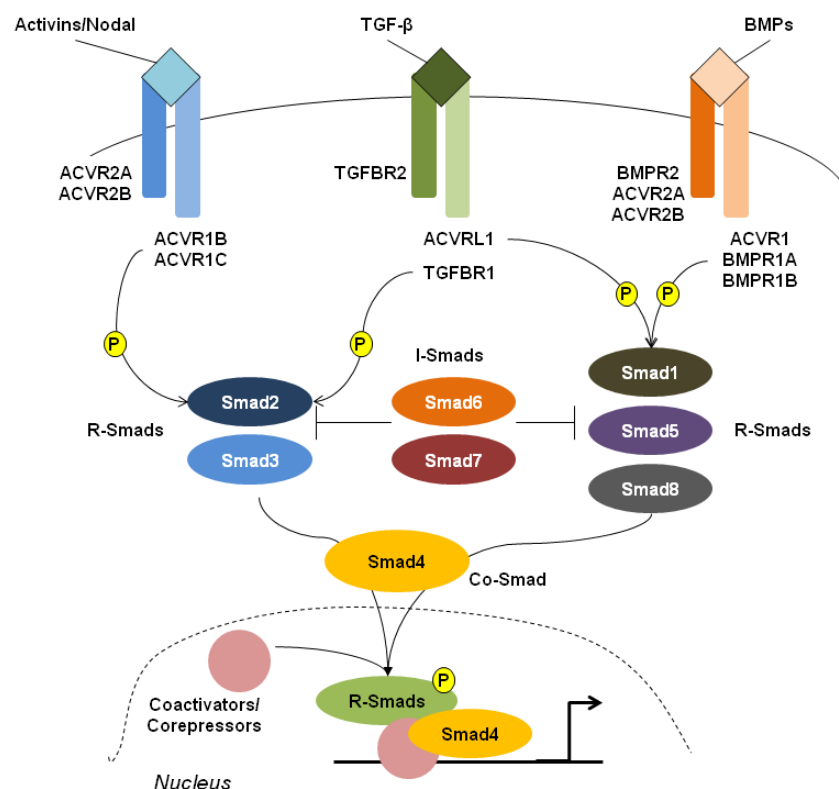
Figure 1.7: TGF-β and Smad signalling pathways.
The schematic diagram depicts the mechanism of signal transduction from the TGF-β superfamily of ligands and receptors to the nucleus via the intracellular Smad signalling pathway. Three major subclasses of TGF-β ligands are identified as activins and Nodal, TGF-β, and BMPs. Upon binding of ligands to type 2 receptors, a heterodimer of type 1/type 2 receptors is formed. Type 1 receptors subsequently phosphorylate receptor-activated Smads (R-Smads), which form a complex with co-Smad, Smad4, and translocate to the nucleus to activate or suppress target genes in the presence of coactivators or corepressors (Heldin et al., 1997). Inhibitory Smads (I-Smads) are capable of blocking R-Smads and therefore, act in a negative feedback loop (Nakao et al., 1997). Overall,

TGF-β, activins and Nodal use Smad2/Smad3 as their R-Smads, while BMPs activate Smad1/Smad5/Smad8 R-Smads. P indicates phosphorylation. ACVR1, activin A receptor type 1; ACVR1B, activin A receptor type 1B; ACVR1C, activin A receptor type 1C; ACVR2A, activin A receptor type 2A; ACVR2B, activin A receptor type 2B; ACVRL1, activin A receptor type 2-like 1; TGFBR2, TGF-β receptor type 2; BMPR2, BMP receptor type 2.

In addition to TGF-β, BMP4 can also maintain the self-renewal capacity of HSCs *in vitro*, whilst at lower concentrations it favours the proliferation of early progenitors in adult human haemopoiesis (Bhatia et al., 1999). BMP4 functions via activation of intracellular Smad5 protein. The deletion of Smad5 protein in the cells derived from yolk sac or embryoid bodies resulted in enhanced proliferation and colony forming potential (Liu et al., 2003); however, similar results could not be achieved in adult mouse HSCs (Singbrant et al., 2006). Thus, it is plausible to propose a role for BMP4/Smad5 signalling in early haemopoiesis. Nonetheless, the observation of normal phenotype upon Smad5 deletion in adult HSCs could be due to redundancy between the Smad signalling proteins.

It has been proposed that Smad signalling is a negative regulator of self-renewal, since the overexpression of inhibitory Smad protein, Smad7, in murine HSCs significantly induced self-renewal (Blank et al., 2006). Furthermore, Smad signalling may have a role during fate determination. The SMAD7 overexpressing HSCs favour myeloid differentiation significantly more than lymphoid. Thus, the activation of Smad signalling results in the reduction of myeloid differentiation and a shift towards lymphoid commitment (Chadwick et al., 2005). Overall, TGF-β ligands maintain the quiescence of HSCs through the activation of intracellular Smad signalling which in turn may inhibit self-renewal or induce lymphopoiesis in early progenitors. Thus, the biological outcome of TGF-β stimulation is dependent on the differentiation state of the haemopoietic cells (Ruscetti et al., 2005).

1.6.4.6 Hedgehog signalling pathway

The hedgehog signalling pathway has been demonstrated to have an important role in the development of the early mesoderm layer through the morphogenic activity of its ligands, Sonic hedgehog (SHH), Indian hedgehog (IHH), and desert hedgehog (DHH), which are cell surface proteins that can bind directly to their receptor Patched (PTC) on neighbouring cells or can be cleaved and diffuse to distal locations. The ligand-receptor interaction results in the activation of another transmembrane protein Smoothened (SMO) and subsequent activation of downstream effectors (Ruiz i Altaba, 1997). Human HSCs express SHH, PTC, and

SMO, while the inhibition of SHH by blocking antibodies resulted in reduced HSC expansion that was rescued by the addition of soluble SHH in NOD-*scid* mice. Furthermore, the SHH effect was blocked in the presence of Noggin, which is an inhibitor of bone BMP4, therefore, suggesting a potential link between SHH and BMP4 in the survival of HSCs (Bhardwaj et al., 2001).

1.6.4.7 HSC niche: acellular factors versus cellular support

Hypoxia and reactive oxygen species (ROS) – HSCs are believed to be located in a hypoxic compartment of the BM niche as confirmed by pimonidazole staining, which is a hypoxia marker (Parmar et al., 2007). Furthermore, Hoechst staining of the niche revealed that LT-HSCs and osteoblasts co-reside in hypoxia, whereas mesenchymal stem cells (MSCs) and vascular endothelial cells are located in normoxia (Winkler et al., 2010). Overall, these observations indicated that LT-HSCs and osteoblasts reside relatively far from the oxygenated circulation. However, recent observations indicate that the apparent hypoxic behaviour of HSCs could be regulated intrinsically, rather than dictated by the niche, as hypoxic HSCs could also be located in close proximity to the vasculature and to other non-hypoxic haemopoietic cells (Wang and Wagers, 2011). The regulation of hypoxia has been shown to be essential for HSC maintenance and can be tracked by the elevated levels of hypoxia inducible factor 1α (HIF1α) under low-oxygen level conditions (Takubo et al., 2010). This upregulation of HIF1α results in a switch from mitochondrial oxidative phosphorylation to glycolysis, which in turn is a pro-survival mechanism for HSCs in hypoxic conditions (Simsek et al., 2010). HIF1α-deficient HSCs were shown to be highly proliferative which resulted in their exhaustion in repopulation assays, especially under haemopoietic stress (Takubo et al., 2010). The same study also showed that the overexpression of HIF1α was associated with loss of HSCs, indicating a tight range of activity for the optimum functionality of this factor.

ROS can also lead to HSC exhaustion as a result of activating the p38/mitogen activated protein kinase (MAPK) pathway which induces proliferation (Ito et al., 2006). The forkhead box O (FoxO) family of transcription factors are important in protecting HSCs from oxidative stress by activating genes involved in the detoxification process. Murine HSCs with triple knockouts of FoxO1, FoxO3 and

FoxO4 displayed impaired repopulation potential, increased cell division and apoptosis, in addition to elevated levels of ROS (Tothova et al., 2007).

Osteoblasts – They are the best characterised cells that reside in contact with HSCs in the BM niche. An increase in their number was directly associated with an increase in the number of HSCs (Calvi et al., 2003) and the loss of osteoblasts from the niche resulted in the loss of HSCs (Visnjic et al., 2004). Osteoblasts express and secrete G-CSF, GM-CSF, IL-6, angiopoietin, TPO, Wnt, Notch, N-cadherin, and osteopontin, which have impact on both HSC maintenance and proliferation (Fleming et al., 2008, Haug et al., 2008, Nilsson et al., 2005, Taichman et al., 1996). Recent visualisation of the niche by confocal microscopy has revealed the close proximity of LT-HSCs and osteoblasts, while the more mature progenitor cells are distally located (Lo Celso et al., 2009).

The perivascular niche – A proportion of the BM niche is in close proximity of vasculature and is therefore known as the perivascular space. The perivascular space provides the means for migration, homing and engraftment of haemopoietic cells and for them to receive differentiation signals and various other stimuli. MSCs, endothelial cells, chemokine (C-X-C motif) ligand 12 (CXCL12)-abundant reticular (CAR) cells and neural cells are the major types of cells in this niche (Wang and Wagers, 2011). MSCs are required for the maintenance of LT-HSCs and their loss resulted in reduction of LT-HSCs by 50% (Mendez-Ferrer et al., 2010). MSCs also express CXCL12 that is involved in HSC migration (Tzeng et al., 2011). The influence of MSCs on haemopoiesis can be altered through change of global gene expression signatures as a result of physiological stimuli. A 50% loss of LT-HSCs was also observed with the loss of CAR cells (Omatsu et al., 2010). CAR cells are characterised by expressing N-cadherin and Jagged ligand on their cell surface. Moreover, endothelial cells are characterised by the expression of vascular endothelial growth factor receptor 2 (VEGFR2), VEGFR3, platelet endothelial cell adhesion molecule 1 (PECAM1) and cadherin 5 type 2 (CDH5), which are required for the maintenance of self-renewal and repopulation capacities of HSCs (Butler et al., 2010). Thus, the above observations indicate a crucial role for the cells in the BM niche in the maintenance of HSCs.

Neural regulation – BM is noticeably innervated by sympathetic and parasympathetic nerves in close proximity of bones and arterial smooth muscle

cells (Mignini et al., 2003). The neural cells in the BM niche can regulate HSC function directly and indirectly. Several neurotransmitter receptors are identified on the surface of HSCs, such as dopaminergic, adrenergic, serotonin, adenosine, opioids, orexin, vasoactive intestinal peptide (VIP) and corticotropin-releasing hormone (CRH) (Steidl et al., 2004). Norepinephrine signalling was demonstrated to cause G-CSF-mediated mobilisation of murine HSCs, suppression of osteoblasts and downregulation of CXCL12 (Katayama et al., 2006). Furthermore, dopamine and norepinephrine signalling were shown to regulate the motility of human HSCs as well as inducing BM repopulation through the activation of Wnt signalling (Spiegel et al., 2007). A recent report also suggested that norepinephrine induces the production of HSCs from their haemogenic endothelium precursors in AGM; thus, providing compelling evidence on the role of neurotransmitters in the formation of HSCs (Fitch et al., 2012). Additionally, the nonmyelinating Schwann cells, which are a subset of the neural cells in BM niche, were shown to be required for the activation of the latent form of TGF-β and therefore, indirectly maintain the quiescence of HSCs (Yamazaki et al., 2011).

1.6.5 Transcription factors in haemopoietic development

Transcription factors are key intrinsic players in cell fate determination during HSC development from AGM and during haemopoietic lineage differentiation (Orkin, 2000). Mutations or chromosomal translocations of haemopoietic transcription factors are the main cause of developing haematological malignancies; therefore, it is required to understand their role during normal haemopoiesis. Conventional and conditional gene knockouts in mice, zebrafish, chicken, *Drosophila* and *Xenopus* have provided thorough insights into the function and role of key haemopoietic transcription factors (Iwasaki and Akashi, 2007, Orkin and Zon, 2008, Rothenberg, 2007).

The transcription factors that are crucially linked to the formation of HSCs during embryogenesis are MLL, runt-related transcription factor 1 (RUNX1), T cell acute lymphocytic leukaemia 1 (TAL1/SCL), LIM domain only 2 (LMO2), and ETS variant 6 (ETV6/TEL); the absence of which results in the abolishment of haemopoiesis (Orkin, 2000, Orkin and Zon, 2008). Chromosomal translocations of these genes are found in the majority of leukaemic events; for example TAL1/SCL and LMO2 translocations in T-cell acute lymphoblastic leukaemia (T-ALL) and translocations

of MLL, RUNX1, and ETV6/TEL that result in the formation of fusion oncogenes in the cases of various myeloid and lymphoid leukaemias (Orkin and Zon, 2008). On the other hand, the major transcription factors that are associated with lineage differentiation are GATA binding protein 1 (GATA1), CCAAT/enhancer binding protein-alpha (CEBPA), SPI1/PU.1, growth factor independent 1 transcription repressor (GFI1), and IFN consensus sequence binding protein (ICSBP) (Figure 1.8) (Orkin and Zon, 2008, Rosenbauer and Tenen, 2007). Nevertheless, it is suggested that all lineage-specific transcription factors are expressed at low levels in HSCs and MPPs at a single-cell level, which is referred to as the lineage priming model (Akashi, 2005, Orkin, 2003). The lineage priming model proposes a gradual fate determination route that is open to switch between lineages at all times. This phenomenon is associated with open chromatin configurations at the promoters of haemopoietic regulators to enable accessible transcription when needed, which offers plasticity to the multipotent precursors. The open chromatin regions can be transiently repressed upon early fate determinations and subsequently permanently silenced in mature blood cells (Orkin and Zon, 2008).
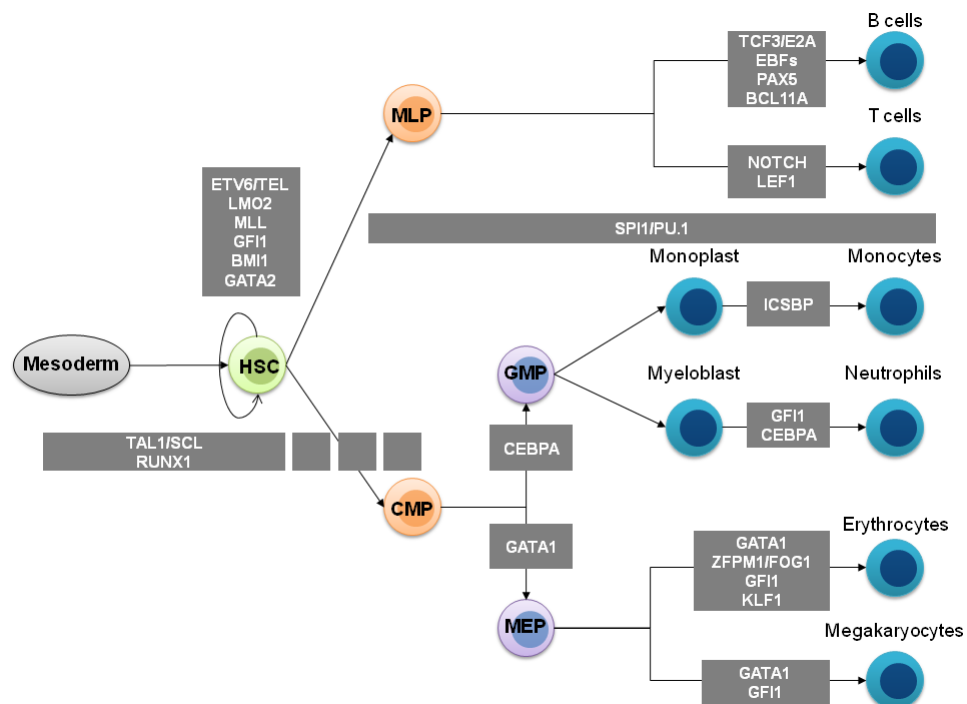
Figure 1.8: Essential transcription factors in haemopoiesis.
Schematic illustration of the transcription factors (grey boxes) that are required for the formation of HSCs from mesoderm and haemopoietic lineage differentiation. TAL1/SCL and RUNX1 are essential for the generation of foetal liver HSCs and subsequently adult HSCs, whose levels are gradually diminished in the course of lineage commitment. SPI1/PU.1 is required for the formation of GMPs as well as lymphoid progenitors. CEBPA is essential for granulocytosis whereas ICSBP is required in the monocytes development. On the other hand, the development of megakaryoctes and erythrocytes require the expression of GATA1. The development of T cells require the activity

of Notch signalling, whereas B cell development is mediated by the function of early B-cell factors (EBFs), PAX5 and TCF3/E2A transcription factors. Some of the transcription factors required for the HSC maintenance, such as GFI1, are also essential in the formation of terminally differentiated cells, such as neutrophils and megakaryocytes (Orkin and Zon, 2008, Rosenbauer and Tenen, 2007).

A global analysis of gene expression signatures in human HSCs, intermediate progenitors and terminally differentiated haemopoietic cells has discovered a complex interconnected network of transcription factors which significantly overlap with the above findings (Novershtern et al., 2011). This report has identified modules of transcription factors which are coregulated at different stages of haemopoiesis. Furthermore, this study has provided supporting evidence for the lineage priming model by showing the binding of lineage-specific transcription factors to their target genes in HSCs, which has been referred to as anticipatory regulation. The global gene expression analysis of HSCs and haemopoietic progenitors are reviewed in Chapter 3.

1.6.6 Epigenetics of haemopoietic development

The importance of epigenetic in transcriptional regulation was discussed in Section 1.5. Therefore, it is important to review the genome-wide studies of chromatin changes during haemopoietic development which provide a better understanding of the lineage priming model.

1.6.6.1 Role of DNA methylation in lineage commitment

The genome-wide studies of DNA methylation has been recently performed on several mouse haemopoietic progenitors, which include MPPs, CMPs, GMPs and CLPs, as well terminally differentiated granulocytes, monocytes and T cells (Ji et al., 2010). This study showed that a number of myeloid- and lymphoid-lineage-specific genes were subjected to differential DNA methylation in the course of haemopoietic development from MPPs. The lymphoid- and myeloid-specific genes showed a significant DNA demethylation during lymphopoiesis and myelopoiesis, respectively, which was associated with transcriptional upregulation. DNA methylation patterns during haemopoiesis demonstrated significant epigenetic plasticity. Lymphopoiesis was significantly linked to the gain of methylation, whereas myelopoiesis showed a higher tendency towards the loss of DNA methylation. Thus, the observation indicated that the lymphopoiesis is more dependent on DNA methylation than myelopoiesis (Ji et al., 2010). In addition to

the identification of many novel loci that were differentially methylated in the course of myelopoiesis and lymphopoiesis, a subset of genes was also identified which showed a progressive hypermethylation in both myeloid and lymphoid lineages which was subsequently linked to the maintenance of multipotent state, such as MEIS1, hepatic leukaemia factor (HLF), HOXA9, and PR domain containing 16 (PRDM16) which have been previously reported to be important during haemopoiesis (Ji et al., 2010). Thus, overall this study showed a significant difference in DNA methylation programming between myelopoiesis and lymphopoiesis, which suggested a DNA-methylation-dependent mechanism of differentiation for the latter.

DNA methylation has also been directly linked to the maintenance of HSC self-renewal and fate determination upon the loss of multipotency (Broske et al., 2009, Trowbridge et al., 2009). These studies demonstrated that by reducing the expression of DNMT1 in HSCs, their self-renewal capacities were impaired and their differentiation was skewed towards the myeloid lineage. The latter could be explained by the above observation which showed a less DNA-methylation-dependent mechanism of differentiation during myelopoiesis. On the contrary, the loss of DNMT3A was associated with failed differentiation, but did not affect the HSCs self-renewal capacities (Challen et al., 2012). Thus, *de novo* DNA methylation by DNMT3A is essential to block the HSC-specific transcription factors which in turn induces lineage commitment and differentiation.

1.6.6.2 Role of histone modifications in lineage priming

The analysis of genome-wide histone modification patterns in HSCs have demonstrated large numbers of bivalent promoters which are noticeably reduced during lineage differentiation in both murine and human haemopoiesis (Cui et al., 2009, Weishaupt et al., 2010). The tracking of CD133[+] human haemopoietic stem and progenitor cells during differentiation to CD36[+] erythroblasts identified a subset of CD133[+] bivalent genes that lost H3K27me3 and became activated in CD36[+] cells. These genes were also associated with paused RNAPII, H3K4me1 and H3K9me1 at their promoters and basal level transcription in CD133[+] cells. On the contrary, the bivalent genes that lost H3K4me3 during differentiation were not associated with RNAPII, H3K4me1 or H3K9me1 at their promoters in CD133[+] cells (Cui et al., 2009). Therefore, the results indicated that genes were primed for

activation or permanent repression in HSCs and progenitor cells and, thus, providing further evidence for the lineage priming model that was discussed above. This study also demonstrated that the majority of bivalent genes in CD133[+] cells (53%) lost their H3K4me3 in CD36[+] cells and only 24% of the bivalent promoters remained bivalent after differentiation, although they exhibited higher levels of H3K27me3 than H3K4me3. Therefore, the differentiation was associated predominantly with the silencing of the developmentally regulated and other lineages-specific loci (Cui et al., 2009). Similar observations was made in the genome-wide analysis of murine haemopoiesis (Weishaupt et al., 2010). Nevertheless, the bivalent promoters in murine HSCs were shown to be significantly associated with the repressive H3K9me3 marks more frequently than the bivalent promoters in T cells, which may propose a unique epigenetic suppression mechanism in HSCs that may be required for blocking immature lineage commitment. Furthermore, the results suggested that T cell maturation is also associated with a global silencing mechanism, as observed in the case of human CD36[+] cells (Weishaupt et al., 2010). The above observations were further supported by a recent study in murine LSK cell population which showed a positive correlation between the levels of H3K4me3 in bivalent promoters in LSK cells and the related gene expression levels in differentiated cell types (Adli et al., 2010). Overall, the genome-wide histone modification analyses provided compelling evidence for the lineage priming model and the role of epigenetics in anticipatory regulation.

In addition to the changes of histone modification signatures during haemopoiesis development, several studies have also investigated the role of PcG proteins in this process. These reports demonstrated negative correlation between the PRC2 expression and long term HSC repopulating capacities (Majewski et al., 2008, Majewski et al., 2010). In other words, these studies showed that EZH2, SUZ12 and EED subunits restrict the function of HSCs, whereas the PRC1 complex and in particular BMI1 are essential for the maintenance of HSC niche (Lessard et al., 1999).

Overall, the above observations indicate an important role for the epigenetic mechanisms of gene regulation in the maintenance of HSC self-renewal as well as lineage differentiation through inducing the lineage priming model.

1.7 Chronic myeloid leukaemia (CML)

CML is a rare haematological disorder affecting of 1 or 2 cases per 100,000 individuals per year with the median age onset of around 65 years. Women are less frequently affected and show higher survival rates compared to men (Berger et al., 2005, Hehlmann et al., 2007). CML is a bi- or triphasic disorder which initiates in the chronic phase (CP) that is a more stable stage of disease and can typically last 3-8 years. CP CML patients can progress to the more advanced phases of disease, known as accelerated phase and blast crisis (BC) (Mughal and Goldman, 1999).

This disease is identified by the presence of Philadelphia (Ph) chromosome which is the outcome of a reciprocal chromosomal translocation t(9;22)(q34;q11) in haemopoietic cells (Rowley, 1973). This translocation brings a proto-oncogene, c-abl oncogene 1 (ABL1), from chromosome 9 to the juxtaposition of breakpoint cluster region (BCR) gene on chromosome 22 (Bartram et al., 1983, Groffen et al., 1984). ABL1 is a non-receptor tyrosine kinase which becomes constitutively active under the control of the BCR promoter (Laneuville, 1995); thus, the BCR-ABL1 fusion oncogene functions as a constitutive tyrosine kinase (Lugo et al., 1990). The underlying mechanisms leading to this translocation are currently unknown.

CML can be diagnosed by excessive granulocytosis in the differential blood count analysis. A white buffy coat is noticeable upon sedimentation that consists of white blood cells. The full diagnosis is confirmed by identifying the presence of the BCR-ABL1 fusion oncogene and/or transcripts in peripheral blood (PB) and/or BM cells by cytogenetic analysis, fluorescence *in situ* hybridisation (FISH) or real time-polymerase chain reaction (RT-PCR), respectively (Hehlmann et al., 2007, Hughes et al., 2006). The entry to BC is diagnosed by the presence of the blast cells with 70% of patients show a shift towards the myeloid lineage and the rest towards the lymphoid lineage or show a mixed leukaemia (Mughal and Goldman, 1999).

1.7.1 The anatomy of BCR-ABL1 oncogene

The relationship between the BCR-ABL1 protein and CML was initially elucidated in BCR-ABL1-transduced mouse cells *in vivo* (Daley et al., 1990, Heisterkamp et al., 1990). Furthermore, BCR-ABL1-expressing murine HSC cells bearing a

mutation in the ATP-binding pocket, which has the kinase activity, of ABL1 did not develop CML-like disease (Zhang and Ren, 1998). Additionally, through ectopically expressing BCR-ABL1 in human CD34$^+$ cells *in vitro* CML-like characteristics could be induced (Ramaraj et al., 2004, Zhao et al., 2001). Therefore, these observations suggest that high levels of BCR-ABL1 expression is sufficient to induce CP CML disease and additional mutations are not required for developing CP characteristics in haemopoietic subclones.

The ABL1 breakpoints are typically located upstream of exon 1b, or downstream of exon 1a, or between exon 1b and 1a (Melo, 1996). On the other hand, the breakpoints in the BCR gene typically occur in a 5.8 kilobase (kb) window spanning exons 12-16 (b1-b5), known as the major breakpoint cluster region (M-bcr). Fusion transcripts of two main forms can be found due to alternative splicing, which present either b2a2 or b3a2 junctions, that translate into a 210-kDa protein (p210$^{BCR-ABL1}$) (Faderl et al., 1999).

1.7.2 Signalling pathways downstream of BCR-ABL1

The constitutive tyrosine activity of BCR-ABL1 results in the activation of several downstream pathways. The active BCR-ABL1 is phosphorylated at BCR tyrosine 177 (Zhang et al., 2001). The growth factor receptor-bound protein 2 (GRB2) binds to this phosphorylation site and recruits son of sevenless (SOS) that in turn activates RAS and the scaffold protein GRB2-associated binding protein 2 (GAB2) (Ren, 2005). BCR-ABL1-mediated phosphorylation of GAB2 is responsible for the constitutive activation of the phosphatidylinositol 3-kinase (PI3K)/AKT and extracellular signal-regulated kinase (ERK) pathways (Sattler et al., 2002, Skorski et al., 1995). Activation of the PI3K/AKT pathway is thought to block the activity of ICSBP, a negative regulator of granulocytosis (Section 1.6.5), by inhibiting its DNA binding capacity that results is subsequent activation of its target genes, such as the antiapoptotic B-cell CLL/lymphoma 2 (BCL2) gene (Middleton et al., 2006).

Furthermore, three SRC family kinases, haemopoietic cell kinase (HCK), LYN and FGR, are phosphorylated by BCR-ABL1. The transcription factor STAT5 is recruited by phosphorylated HCK (Ilaria and Van Etten, 1996, Ren, 2005), which subsequently upregulates cyclin D1 (CCND1) that causes G_1 to S progression and increased proliferation (Nosaka et al., 1999). STAT5 knockdown in primary CML

cells blocked the formation of Ph[+] myeloid lineage and *Stat5a*[-/-] *Stat5b*[-/-] mice foetal liver cells failed to develop leukaemia upon retroviral transduction of BCR-ABL1 (Hoelbl et al., 2006). Furthermore, STAT5 activates the antiapoptotic protein BCL2L1/BCLX that leads to the inhibition of apoptosis (Gesbert and Griffin, 2000).

Another intracellular signalling pathway mediated by the RAC family of guanosine triphosphatases (GTPases), which include RAC1, RAC2 and RAC3, is also demonstrated to be active in CML cells (Thomas et al., 2007). The knockdown of RAC1 and RAC2 genes in murine cells expressing p210[BCR-ABL1] caused a delay in developing a myeloproliferative disorder which was associated with reduced phosphorylation of BCR-ABL1 downstream effectors CRK-like (CRKL), ERK, c-Jun N-terminal kinase (JNK) and p38; proposing a RAC-dependent mechanism of function for the BCR-ABL1 oncoprotein (Thomas et al., 2007).

Thus, the overall net effect of BCR-ABL kinase activity is increased proliferation and survival as well as reduced apoptosis (Figure 1.9).
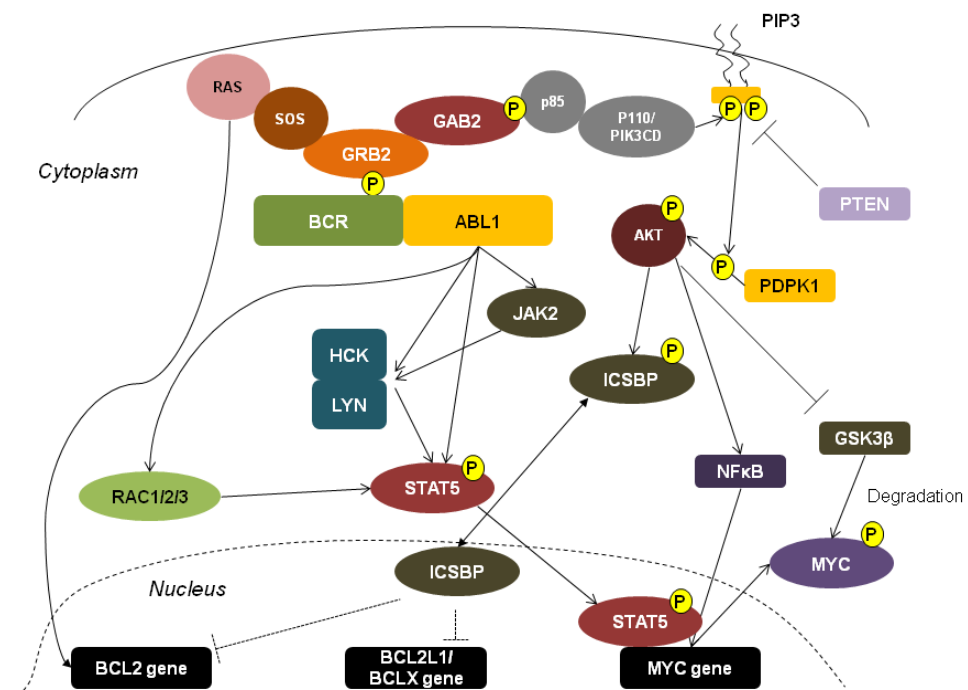
Figure 1.9: The downstream signalling mediators of BCR-ABL1 tyrosine kinase.
GRB2 binds to the phosphorylated tyrosine 177 residue of BCR and recruits SOS, an activator of RAS-MAPK pathway, and GAB2, which activates p85 and p110 subunits of PI3K pathway. Activation of PI3K pathway leads to downstream activation of AKT that induces the expression of MYC oncogene, through NFκB signalling, and its stabilisation by inhibiting GSK3β. Activated AKT also induces the expression of the antiapoptotic factors BCL2 and BCL2L1/BCLX by inhibiting ICSBP function. On the contrary, the activation of phosphatase and tensin homolog (PTEN) blocks the activity of AKT by inhibiting its phosphorylation that in turn block the activation of AKT downstream targets. Activation of Janus kinase 2 (JAK2) by ABL1 and subsequent activation of

downstream HCK and LYN SRC kinases result in the phosphorylation and activation of STAT5, which can also be activated by the RAC family proteins. STAT5 activity can also cause MYC expression. The overall outcome of BCR-ABL1 oncogenic function is increased proliferation and survival. P indicates phosphorylation. PIK3CD, PI3K catalytic delta polypeptide.

1.7.3 CML stem cells (LSCs)

The early evidence regarding the clonality and HSC origins of CML came from the observations in which cells isolated from the BM of CML patients were associated with increased proliferation and the capacity to differentiate into different myeloid lineages (Dameshek, 1951). The identification of Ph chromosome and BCR-ABL1 oncogene in multiple myeloid lineages (Tough et al., 1963, Whang et al., 1963) and similar X-chromosome inactivation patterns in the granulomonocytic lineages, platelets and B cells (Fialkow et al., 1967, Fialkow et al., 1977) further elucidated the clonality of CML and self-renewing, multipotent HSCs as the origin of the disease. The initial evidence for the presence of repopulating LSCs came from the observations that the white blood cells obtained from CML patients were capable of repopulating in severely neutropenic recipients and generate Ph^+ progeny (Levin et al., 1963), which was later explained to be due to the presence of high numbers of mobilised LSCs into PB in CML patients (Dazzi et al., 2000, Petzer et al., 1996). Furthermore, the CMP or GMP origin of CML was ruled out as the ectopic expression of BCR-ABL1 in murine haemopoietic progenitor cells failed to induce self-renewal capacities; therefore, the results provided further evidence to support the HSC origin of CML (Huntly et al., 2004).

The isolation of LSCs from their normal counterparts is one of the major challenges in studying LSCs. LSCs display similar cell surface markers that define HSCs, which include CD34, CD90, and aldehyde dehydrogenase in the absence of Lin and CD38 markers (Florian et al., 2006, Jiang et al., 2008, Petzer et al., 1996). Despite the similarities between HSCs and LSCs in terms of cell surface markers, LSCs exhibit a higher proportion of cells in cycle which can be used for their isolation (Holyoake et al., 1999).   Nonetheless, recent studies have demonstrated novel cell surface markers, including IL-1 receptor accessory protein (IL1RAP) and dipeptidyl peptidase 4 (DPP4/CD26), to distinguish between LSCs and HSCs (Herrmann et al., 2012, Jaras et al., 2010). Ultimately, the genotyping approaches are applied to characterise the sorted cell populations.

LSCs are characterised with the highest levels of BCR-ABL1 transcripts (Jamieson et al., 2004, Jiang et al., 1999, Jiang et al., 2007). The expression of BCR-ABL1 in LSCs has been linked to the secretion of autocrine IL-3 and G-CSF that results in stem cell differentiation and inhibition of self-renewal (Holyoake et al., 2001, Jiang et al., 1999). In line with these findings, LSCs were shown to have reduced self-renewal capacities than their normal counterparts as observed with their limited potential to compete with HSCs in *in vitro* and *in vivo* assays (Petzer et al., 1997, Udomsakdi et al., 1992). Furthermore, LSCs are associated with increased proliferative activities in comparison to HSCs (Holyoake et al., 1999) which could be explained by the autocrine loop of IL-3 and G-CSF that induces the proliferation and survival of CML progenitors. The increased sensitivity of LSCs to the stromal-secreted chemokines such as CXCL12, chemokine (C-C motif) ligand 2 (CCL2), and CCL3/MIP-1α may decrease the quiescence of LSCs relative to HSCs and can increase their mobilisation (Cashman et al., 2002, Eaves et al., 1993). However, several reports suggested a more profound role for TGF-β in the maintenance of quiescence of LSCs similar to their role in HSCs (Section 1.7.6) (Cashman et al., 1992, Naka et al., 2010). LSCs, unlike HSCs, are not dependent on the external growth factors to inhibit apoptosis and therefore, are associated with significantly reduced apoptosis which could be also linked to the autocrine secretion of IL-3 and G-CSF (Jiang et al., 1999, Maguer-Satta et al., 1998). Overall, LSCs demonstrate increased proliferation potential and reduced quiescence, self-renewal and apoptotic capacities in comparison to HSCs.

The identification of LSCs hypothesised two models which could explain the progression of CP CML to BC. First, LSCs in CP may acquire further mutations that would in turn skew their differentiation towards myeloid lineage or the second model which suggests the acquisition of further mutations by the GMPs or other lineage-restricted progenitors which cause them to abnormally self-renew and block differentiation. Recent studies have provided compelling evidence in favour of the second model which show abnormal activation of β-catenin in GMPs that in turn results in the formation of undifferentiated blasts (Jamieson et al., 2004, Neering et al., 2007).

Based on the above observations, a hypothetical model of LSC formation during CP and BC is proposed (Figure 1.10) (Quintas-Cardama and Cortes, 2009).
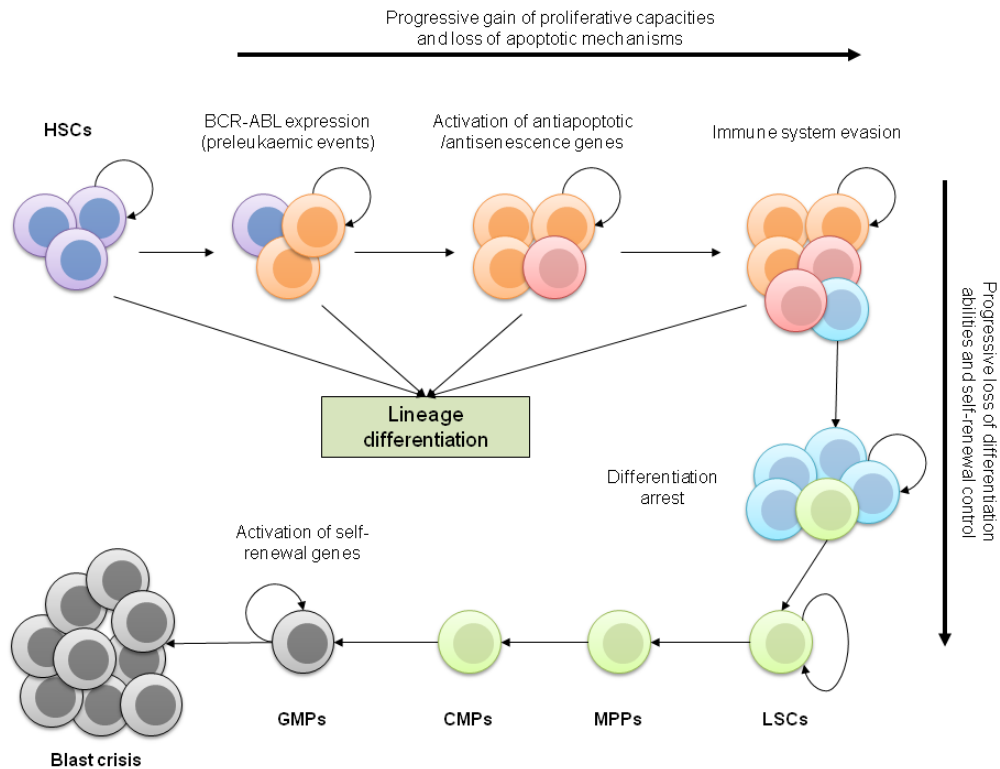
Figure 1.10: The hypothetical model of LSC formation during CP and BC.
Several lines of evidence indicate that LSCs are originated from multipotent self-renewing HSCs. The reciprocal translocation and BCR-ABL1 expression are the first preleukaemic events that occur in an HSC. Subsequently by activation of antiapoptotic genes such as BCL2 and the activity of proliferative genes such as MYC, a progressive acquisition of proliferative potential and loss of apoptotic abilities occur. Further loss of control on self-renewal due to the inactivity of tumour suppressor genes such as p53 and CDKN2A/p16 results in the formation of LSCs. The reduction of differentiation capacities as a result of Ikaros family zinc finger 1 (IKZF1) or CEBPA suppression in LSCs and re-activation of self-renewal genes such as β-catenin in the leukaemic GMPs in addition to accumulation of further genetic and/or epigenetic abnormalities can give rise BC (Quintas-Cardama and Cortes, 2009).

1.7.4 Tyrosine kinase inhibitors (TKIs) in CML therapy

The small molecule TKIs are the central line of treatment against CML. Over the past decade the development of TKIs that directly target the constitutive tyrosine kinase activity of BCR-ABL1 has resulted in significantly improved survival rates and disease management in CP CML patients. Nevertheless, it has to be noted that allogeneic transplantation remains the most effective long-term therapy for CML, particularly in the more aggressive stages of the disease.

The first TKI to be discovered was imatinib which could specifically discriminate CML cells from their normal counterparts by directly targeting BCR-ABL1 (Druker et al., 1996). Imatinib treatment of CML in CP is associated with an overall survival rate of 89% over a five-year clinical evaluation and a progression-free survival rate of 93%, which was much higher than earlier treatment strategies involving

interferon-alpha (IFN-α) (Druker et al., 2006). However, resistance to imatinib can arise as a result of mutations in the kinase domain that either make direct interactions with imatinib or are important in formation of the inactive BCR-ABL1 conformation that is required for drug interaction (Gorre et al., 2001).

Therefore, two second generation TKIs were designed to overcome the observed imatinib resistance, which included nilotinib (Weisberg et al., 2005), a derivative of imatinib with ~30-fold higher strength, and dasatinib (Shah et al., 2004), with ~300-fold higher potency than imatinib. The administration of second generation TKIs resulted in achieving a quicker response, significantly reduced progression to the later stages of the disease and increased overall survival (Jabbour et al., 2011). Nevertheless, BCR-ABL1^{T315I} is the most TKI-insensitive mutation which is referred to as the gatekeeper mutation that could not be targeted by second generation TKIs (Azam et al., 2008). Therefore, there have been increased efforts to develop third generation TKIs that can target BCR-ABL1^{T315I} mutant, such as ponatinib (O'Hare et al., 2009)and DCC-2036 (Chan et al., 2011).

CML patients who have responded to therapy and are in a state of remission harbour a very suppressed clone of CML cells which is referred to as minimal residual disease (MRD) or LSCs (Chomel et al., 2011). LSCs are responsible for the relapse of CP CML after the withdrawal of TKIs and therefore, patients are kept on lifelong TKI treatment after achieving remission. The survival of LSCs in the presence of TKIs is the current challenge in CML management and treatment strategies. Thus, understanding of the LSC biology is essential for developing effective chemotherapies that can completely eradicate CML.

1.7.5 Targeting CML stem cells

The TKI-insensitivity of LSCs raised the question of whether resistance is conferred in a BCR-ABL1-dependent or independent manner. TKIs have been shown to block the kinase activity of BCR-ABL1 in LSCs (Corbin et al., 2011, Hamilton et al., 2012); however, the rate of apoptosis in LSCs upon TKI treatments is significantly lower than in progenitors. Thus, LSCs are believed not to be BCR-ABL1-kinase addicted (Hamilton et al., 2012).

It was mentioned above that TKI treatment can inhibit BCR-ABL1 function in LSCs, but cannot eliminate them. Therefore, an LSC with suppressed BCR-ABL1 activity upon TKI treatment could be associated with a loss-of-function, gain-of-function, or neutral status relative to an untreated LSC harbouring fully functional BCR-ABL1 proteins. The loss-of-function LSCs are therefore, in a dominant-negative state which are cytokine-dependent and their myeloid cells express only very low levels of BCR-ABL1 (Corbin et al., 2011, Kumari et al., 2012). The neutral variants, on the other hand, are indistinguishable from HSCs and are suggested to be more dependent on HIF1α and promyelocytic leukaemia (PML) proteins (Ito et al., 2008, Zhang et al., 2012). In the gain-of-function variants the pathways downstream of BCR-ABL1 exhibit high activity despite the inhibition of the kinase domain, which subsequently enhance migration and minimise adhesion capacities (Ramaraj et al., 2004). It is also suggested that the gain-of-function variants can represent epigenetic alterations as a result of initial BCR-ABL1 activity which still persist after TKI-induced suppression. Thus, blocking kinase-independent or epigenetically active pathways may be required in order to eliminate LSCs. These reports indicate that there are multiple subclones of LSCs, which could exist simultaneously, and potentially a combinatorial treatment strategy is required to eliminate LSCs (Kaelin, 2005). Therefore, it is required to dissect the role of key biological pathways in the maintenance of LSCs, as discussed below.

1.7.6 Major biological pathways involved in the survival of LSCs

1.7.6.1 Wnt signalling pathway

It was demonstrated in a murine CML model that the absence of β-catenin reduced the severity of the disease as a result of impaired self-renewal capacities of LSCs (Zhao et al., 2007a). However, contradictory observations suggested that the inhibition of GSK3β, a negative regulator of β-catenin, in combination with imatinib caused apoptosis in LSCs (Reddiconto et al., 2012). Thus, it is plausible that β-catenin levels should be tightly regulated in LSCs to act as a pro-survival factor.

1.7.6.2 TGF-β signalling pathway and the role of FOXO3A and BCL-6

BCR-ABL1 induces the cytoplasmic translocation and subsequent degradation of FOXO3A, which is phosphorylated by the activated AKT, in CML progenitor cells.

Imatinib treatment, therefore, results in the nuclear stabilisation of FOXO3A which causes cell cycle arrest and apoptosis through upregulation of CDKN1B/p27 and BCL2L11/BIM genes (Essafi et al., 2005). A recent study on the role of TGF-β in the maintenance of LSCs, demonstrated inhibition of AKT phosphorylation in the presence of TGF-β ligand that resulted in nuclear localisation of FOXO3A in LSCs, which was independent of TKI treatment (Naka et al., 2010). The murine FOXO3A-deficient CML model developed leukaemia in the first engraftment, but failed to promote leukaemia in later engraftments; thus, implying an important function for TGF-β-FOXO3A signalling in the maintenance of LSCs. Through inhibition of TGF-β signalling by LY364947, a TGFBR1 inhibitor, and in combination with imatinib, the CML mice demonstrated better prognosis than the imatinib-treated controls (Naka et al., 2010). This was mediated by the activation of AKT which caused cytoplasmic translocation and degradation of FOXO3A. This study provided substantial evidence on the role of TGF-β signalling in maintaining LSC quiescence despite the expression of BCR-ABL1. However it remains unclear whether TGF-β signalling is regulated by autocrine or paracrine means in CML.

FOXO3A also targets BCL6 in CML (Hurtz et al., 2011). The nuclear localisation and activity of FOXO3A in the presence of TKIs or TGF-β stimulation may result in upregulation of BCL6 which in turn protects LSCs by suppressing p53 and CDKN2A/p16 tumour suppressors. Therefore, the administration of retro-inverso BCL6 peptide inhibitor (RI-BPI) resulted in increased survival of mice transplanted with CML cells (Hurtz et al., 2011). Thus, a specific LSC survival mechanism can be proposed in which the TGF-β-FOXO3A signalling suppresses LSC apoptosis by activating BCL6.

1.7.6.3 Hedgehog signalling pathway

The hedgehog pathway is implied in conferring self-renewal capacities to LSCs through the function of SMO (Dierks et al., 2008). TKIs are unable to inhibit the hedgehog signalling, but hedgehog-blocking antibodies can successfully impair the pathway (Zhao et al., 2009); therefore, indicating a BCR-ABL1-independent mechanism for hedgehog activation. Cyclopamine, a potent antagonist of hedgehog signalling by inducing an inactive conformation in SMO, was shown to target LSCs specifically over HSCs; therefore, it can be used as a potent chemotherapeutic in combination with TKIs (Zhao et al., 2009).

1.8 Epigenetic origins of cancer

There has been a growing interest over the past decade to understand the role of epigenetic abnormalities as an equally important factor as genetic abnormalities in cancer development (Herman and Baylin, 2003; Jones and Baylin, 2007). In fact, the first evidence of epigenetic involvement in cancer came from the observations of abnormal changes in DNA methylation patterns in cancer genome (Riggs and Jones, 1983). Therefore, the understanding of underlying epigenetic mechanisms and abnormalities that give rise to the gene expression signatures of cancer cells is crucial for a better understanding of the cancer biology. Tumours often form as a result of the stress responses such as exposure to DNA damaging agents. Therefore, under these circumstances predisposition to epigenetic abnormalities can result in the induction of cell survival mechanisms instead of apoptosis or senescence (Ohm and Baylin, 2007). Nevertheless, it can be argued that the genetic mutations in key developmental and cell signalling pathways, such as Wnt, hedgehog and Myc, are sufficient for tumourigenesis (Taipale and Beachy, 2001; Hanahan and Weinberg, 2000), but it has been demonstrated that the induction of such genetic mutations in mature cells results in apoptosis and senescence (Lowe et al, 2004). Thus, other cellular changes are required for the cells to survive and initiate cancer in the presence of these mutations. These key players can be the epigenetic modifications such as DNA methylation or histone modifications whose roles during early stages of cancer formation are discussed below.

1.8.1 Role of DNA methylation in the formation of cancer stem cells

Abnormal promoter DNA hypermethylation which results in abnormal gene silencing has been reported as one of the key epigenetic changes that confers advantage to cancer cell survival (Jones and Baylin, 2007). The promoter hypermethylation is identified in association with the genes that regulate stem and progenitor cells maintenance, such as CDKN2A/p16, GATA4, GATA5 and several regulators of Wnt signalling pathway. These epigenetic abnormalities are believed to occur during early neoplastic transformation (Jones and Baylin, 2007). This model has been investigated in colon cancers with respect to the Wnt signalling regulators. An inhibitor of Wnt signalling, SOX17, which promotes the degradation of β-catenin, was found to be abnormally hypermethylated and silenced in benign colon neoplasams (Zhang et al., 2008). This epigenetic abnormality can, therefore,

result in the formation of cancer stem cells by maintaining the constitutive activity of Wnt signalling pathway which is important in stem cell self-renewal and maintenance.

1.8.2 Role of histone modifications in the formation of cancer stem cells

The genome-wide studies of histone modifications and DNA methylation in colon cancer showed that 50% of abnormally hypermethylated genes in cancer cells were associated with PRC2 complex and bivalent domains at their promoters in ES cells or embryonic mesenchymal cells (Schlesinger et al., 2007, Widschwendter et al., 2007). DNA demethylation of these genes by 5-aza-2′-deoxycytidine treatment or as a result of DNMT knockout in colon cancer cells did not result in the resolution of bivalency, but was associated with increased levels of H3K27me3 at these promoters (McGarvey et al., 2006). The bivalent genes in adult stem cells are poised as their fate will be determined upon lineage differentiation. However, as discussed earlier, the lineage priming model indicates very low expression levels of the bivalent genes whose activities are required subsequently during differentiation. Therefore, the observation of hypermethylated bivalent genes in cancer cells suggests a permanent blockage on transcriptional activation of lineage-specific transcription factors that consequently confers a less mature phenotype to cancer cells. The maintenance of stem cell-like epigenetic signature at early stages of tumourigenesis would in turn promote the expansion of immature pre-malignant cancer cells. These pre-malignant clones can provide foundations for further genetic or epigenetic abnormalities that lead to a malignancy. GATA4, which is a transcription factor required for development of mature epithelial cells in colon, is found to be bivalent and expressed at very low levels in ES cells. Furthermore, upon *in vitro* induction of differentiation, GATA4 promoter predominantly retains H3K4me3 marks that is associated with transcriptional activation (Tiwari et al., 2008). GATA4 promoter retains its bivalency but becomes significantly hypermethylated in colon cancer cells which results in the permanent silencing of this loci (Tiwari et al., 2008). Thus, hypermethylation completely abolishes the poised state and eliminates the accessibility of promoter for activation during differentiation.

1.9 Aims of this thesis

CML is a stem-cell driven and clonal haematological malignancy that has played a significant role in the understanding of cancer stem cell biology. A systems biology approach was implemented to investigate the epigenetic programming that defines stem cell identity and entry into the proliferative compartment in CML. The aims of the work presented in this thesis were:

1. To elucidate the gene expression signatures of LSCs and CML progenitor cells (LPCs) to identify biological pathways important in maintenance of stem cell features as well as exit from the stem cell programme.

2. To develop a ChIP method that can reliably be coupled to high throughput DNA sequencing for small primary stem cell populations in order to study and compare histone modification patterns of HSCs and LSCs.

3. To establish global histone modification maps of HSCs, LSCs and LPCs to identify the relationship between the epigenetic programming and the observed gene expression signatures.

4. To functionally validate the novel biological pathways that are identified by global gene expression and epigenetics analyses to be important in the maintenance of LSCs.

2. Materials and Methods

2.1 Materials

2.1.1 Composition of molecular biology solutions

Sterile high-performance liquid chromatography (HPLC)-grade water (VWR International, UK) was used to prepare all solutions.

Cell lysis buffer (CLB):

- 10 mM Tris-hydrochloric acid (HCl) (pH 8.0)

- 10 mM sodium chloride (NaCl) (VWR International)

- 0.2% Igepal (Sigma-Aldrich, Dorset, UK)

- 10 mM sodium butyrate (NaBu) (Sigma-Aldrich)

- 50 µg/ml phenylmethanesulfonyl fluoride (PMSF) (Sigma-Aldrich)

- 1 µg/ml Leupeptin (MP Biomedicals, California, USA)

Nuclear lysis buffer (NLB):

- 50 mM Tris-HCl (pH 8.1)

- 10 mM ethylenediaminetetraacetic acid (EDTA) (Sigma-Aldrich)

- 1% sodium dodecyl sulphate (SDS) (Sigma-Aldrich)

- 10 mM NaBu

- 50 µg/ml PMSF

- 1 µg/ml Leupeptin

Immunoprecipitation (IP) dilution buffer (IPDB):

- 20 mM Tris-HCl (pH 8.1)

- 150 mM NaCl

- 2 mM EDTA

- 1% Triton X-100 (VWR International)

- 0.01% SDS

- 10 mM NaBu

- 50 µg/ml PMSF

- 1 µg/ml Leupeptin

IP wash buffer 1 (IPWB1):

- 20 mM Tris-HCl (pH 8.1)

- 50 mM NaCl

- 2 mM EDTA

- 1% Triton X-100

- 0.1% SDS

IP wash buffer 2 (IPWB2):

- 10 mM Tris-HCl (pH 8.1)

- 250 mM lithium chloride (LiCl) (Sigma-Aldrich)

- 1 mM EDTA

- 1% Igepal

- 1% Deoxycholic acid (Sigma-Aldrich)

IP elution buffer (IPEB):

- 100 mM sodium bicarbonate ($NaHCO_3$) (VWR International)

- 1% SDS

1X phosphate buffered saline (PBS):

1X PBS used for washing the cells in ChIP assay was prepared by dissolving one PBS tablet (Sigma-Aldrich) in 500 ml HPLC water.

10X deoxyribonucleotide triphoshphate (dNTP) mix for labelling ChIP samples:

The following mix was prepared in sterile labelling water (Invitrogen, Paisley, UK) and used for labelling ChIP samples prior to microarray hybridisations:

- 1 mM deoxycitidine triphosphate (dCTP) (Scientific Laboratory Supplies, Hull, UK)

- 2 mM each of deoxyadenosine triphosphate (dATP), deoxyguanosine triphosphate (dGTP), and deoxythymidine triphosphate (dTTP) (Scientific Laboratory Supplies)

20X saline-sodium citrate (SSC) buffer:

The following were dissolved in 800 ml HPLC water:

- 175.3 g NaCl

- 88.2 g sodium citrate (VWR International)

The volume was adjusted to 1000 ml and the pH adjusted to 7.0.

Tecan hybridisation station (HS) 4800 hybridisation buffer:

- 50% formamide (Sigma-Aldrich)

- 5% dextran sulphate (AppliChem)

- 0.1% Tween 20 (Sigma-Aldrich)

- 2X SSC

- 10 mM Tris-HCl pH 7.4

PBS/0.05% Tween 20 (Hybridisation wash solution 1):

The following salts were dissolved in 1 litre of HPLC water:

- 7.33 g NaCl

- 2.36 g sodium phosphate dibasic (Na_2HPO_4) (VWR International)

- 1.52 g sodium phosphate monobasic monohydrate ($NaH_2PO_4.H_2O$) (VWR International)

- 500 μl Tween 20

10X polymerase chain reaction (PCR) buffer:

- 670 mM Tris-HCl (pH 8.8)

- 166 mM ammonium sulphate [$(NH_4)_2SO_4$] (Sigma-Aldrich)

- 67 mM magnesium chloride ($MgCl_2$) (Sigma-Aldrich)

10X Tris-borate-EDTA (TBE) buffer:

The following salts were dissolved in 1 litre of HPLC water:

- 108 g Tris base (Invitrogen)

- 55 g boric acid (VWR International)

- 9.3 g EDTA (VWR International)

1X TBE buffer was prepared by mixing 100 ml 10X TBE buffer with 900 ml HPLC water.

50X Tris-acetate-EDTA (TAE) buffer:

The following salts were dissolved in 1 litre of HPLC water:

- 242 g Tris base

- 57.1 ml acetic acid (BDH)

- 18.6 g EDTA

1X TAE buffer was prepared by mixing 20 ml 10X TAE buffer with 980 ml HPLC water.

2.1.2 Composition of tissue culture solutions

"DAMP" solution for thawing cryopreserved CD34$^+$ or mononuclear cells (MNCs) aliquots from liquid nitrogen:

- 2 ml DNase I solution (1mg/ ml) (StemCell Technologies, Vancouver, Canada)

- 1.25 ml MgCl$_2$ (400X, 1M solution) (Sigma-Aldrich)

- 53 ml trisodium citrate (0.155M) (Sigma-Aldrich)

- 25 ml human serum albumin (20%) (Scottish National Blood Transfusion Service)

Volume was adjusted to 500 ml with Dulbecco's PBS (Mg^{2+}/Cl$^-$ free) (Invitrogen).

Serum Free Medium (SFM):

The following were made up and filter sterilised using vacuum bottle top filters:

- 25 ml bovine serum albumin (BSA)/insulin/transferrin (BIT) (StemCell Technologies)

- 1.25 ml L-glutamine (200 mM) (Invitrogen)

- 1.25 ml penicillin/streptomycin (5000U/ ml) (Invitrogen)

- 250 µl 2-mercaptoethanol (50 mM) (Invitrogen)

- 500 µl low density lipoprotein (LDL) (10mg/ ml) (Sigma-Aldrich)

- 97.25 ml Isocove's modified Dulbecco's medium (IMDM) (Sigma-Aldrich)

SFM supplemented with physiological growth factors cocktail:

The following growth factors were added at indicated volumes per 10 ml SFM and filter sterilised through 0.22µm filter:

- 5 µl G-CSF (2 µg/ ml) (StemCell Technologies)

- 20 µl GM-CSF (0.1 µg/ ml) (StemCell Technologies)

- 2 µl IL-6 (5 µg/ ml) (StemCell Technologies)

- 5 µl LIF (0.1 µg/ ml) (Peprotech, USA)

- 20 µl macrophage inflammatory protein (MIP)-1α (0.1 µg/ ml) (StemCell Technologies)

- 4 µl SCF (0.5 µg/ ml) (StemCell Technologies)

2.1.3 Antibodies

Name	Supplier	Catalogue no.
H3K4me1	Abcam, Cambridge,UK	ab8895
H3K4me3	Abcam	ab8580
	Active Motif, London, UK	39159
H3K27me1	Millipore, Watford, UK	07-448
H3K27me3	Millipore	07-449
H3K36me1	Abcam	ab9048
H3K36me3	Abcam	ab9050
H3K79me1	Abcam	ab2886
H3K79me3	Abcam	ab2621
H3K9/K14ac	Millipore	06-599
Normal rabbit immunoglobulin G (IgG)	Millipore	12-370
H3K4me3 (LowCell ChIP Kit)	Diagenode, Liege, Belgium	pAb-003-050
H3K27me3 (LowCell ChIP Kit)	Diagenode	CS-069-50
Human Anti-CD34-APC	Becton Dickinson (BD), Oxford, UK	555824
Human Anti-CD38-FITC	BD	555459
Human Anti-CD38-PerCP	BD	561106
Annexin V-PE	BD	556422
Mouse Anti-SSEA1-PE	BD	560142

2.2 Methods

2.2.1 Cell culture

2.2.1.1 Culture of cell lines

The mES cell line E14 (Hooper et al., 1987) was obtained from Bill Skarnes (Sanger Institute, Cambridge, UK). E14 cell line was established from the inbred mouse strain 129/Ola. E14 cells were cultured in a feeder-independent manner on gelatinised 6-well plates or 25 cm^2 tissue culture flasks (Corning, Amsterdam, The Netherlands) from frozen in Dulbecco's modified Eagle medium (DMEM) (Invitrogen) supplemented with 15% foetal bovine serum (FBS) (Invitrogen), 2% LIF-conditioned medium (CM), and 1.5X 10^{-4} M monothioglycerol (MTG) (Sigma-Aldrich). Cells were maintained at 37°C, 5% CO_2 and were passaged with warm fresh medium after reaching 70-80% confluency (approximately every 2 days). Cells were reseeded at 150,000 cells per well of a 6-well plate, or 1x10^6 cells per 75 cm^2 flask, or 3x10^6 per 225 cm^2 flask. Since 10^8 cells were required for the conventional ChIP assays, five 225 cm^2 flasks were normally cultured to obtain sufficient number of cells.

GFP-tagged murine stromal cell line OP9 (Kodama et al., 1994) was received as a gift from Dr. Alison Michie (Institute of Cancer Sciences, University of Glasgow). OP9 cells were derived from newborn B6C3F2 *op/op* mouse calvaria with a mutation in the gene encoding for macrophage-colony stimulating factor (M-CSF), thus resulting in the lack of secretion of functional M-CSF (Yoshida et al., 1990). The OP9 cells were demonstrated to augment haemopoiesis and facilitate the differentiation of ES cells to CD34$^+$ cells *in vitro* (Vodyanik et al., 2005). OP9-GFP cells were cultured in minimum essential medium alpha (MEM α) (Invitrogen) containing 20% FBS (Invitrogen), 1% penicillin/streptomycin, 1% L-glutamine, 1% non-essential amino acids (NEAA) (Invitrogen), 50 µM 2-mercaptoethanol (Invitrogen). Cells were maintained at 37°C, 5% CO_2 and were passaged with warm fresh medium after reaching confluency (approximately every 3 days). The OP9-GFP co-culture systems were set-up by seeding 50,000 cells/well of collagen-coated 6-well plates and fresh medium was replenished after 3 days until reaching confluency prior to plating the primary haemopoietic cells.

2.2.1.2 Collection of human primary cell samples

All samples were collected with approval from the Local Research and Ethics Committee and with written informed patient consent from patients at diagnosis of CP CML. Cells were collected by leukapheresis prior to any drug treatment. Each sample was determined to be Ph$^+$ by FISH and BCR-ABL$^+$ by PCR (Dr. Elaine Allan, Institute of Cancer Sciences, University of Glasgow). Further samples were also obtained from patients, with normal BM undergoing autologous stem cell collection for either non-Hodgkin's lymphoma or normal bone marrow donors, who had been mobilised with G-CSF and had excess CD34$^+$ cells remaining after those required for clinical use had been processed.

2.2.1.3 Purification of the MNC fraction from whole blood cell samples

Either 6 ml of ficoll/Histopaque solution was added to a 15 ml falcon tube, or 20 ml of ficoll/Histopaque solution was added to a 50 ml falcon tube (depending on the volume of blood sample) and brought to room temperature (RTm). The whole blood sample was first diluted (1:2) with PBS, carefully layered drop-wise onto the ficoll/Histopaque solution, until it reached the top of the centrifuge tube and centrifuged at 1,500 rounds per minute (rpm) for 20 minutes (mins) at RTm. Following centrifugation, the opaque interface containing the MNCs was carefully aspirated with a pastette. The interface was then transferred into a sterile centrifuge tube with a pastette and washed twice with sterile PBS (centrifuged at 1,000 rpm for 5 mins) (Dr. Alan Hair, Institute of Cancer Sciences, University of Glasgow). The resultant MNCs were either used fresh or cryopreserved according to Section 2.2.1.5 until required.

2.2.1.4 Selection of CD34$^+$ cells from MNC samples

Enrichment for CD34+ cells was achieved using the sterile CliniMACS system (Miltenyi Biotec, Bisley, UK), which positively selects for CD34$^+$ cells according to the manufacturers' instructions (Dr. Alan Hair, Institute of Cancer Sciences, University of Glasgow). Briefly, total MNCs were incubated with a specific anti-CD34 monoclonal antibody (Miltenyi Biotec) to which super-paramagnetic MACS beads (~50 nm in diameter) had been conjugated. The cell sample was then passed through a high-gradient magnetic separation column, where the target CD34$^+$ cells were retained in the column and the unlabelled CD34$^-$ cells flushed

through and discarded. The bound CD34$^+$ cells were then eluted after removal from the magnetic field, collected and an aliquot was removed for flow cytometry assessment of CD34 purity, which confirmed that all samples were >95% CD34$^+$ post-selection. All samples were stored at the indicated concentrations (Section 2.2.1.5) in liquid nitrogen until required.

2.2.1.5 Cryopreservation of cells

E14 cells were frozen down in cryotubes at a density of 3-5x10^5 cells per ml in 10% dimethyl sulfoxide (DMSO), 40% FBS, and 50% E14 culture medium. OP9-GFP cells were frozen down at 1x10^6 cells per cryotube in 10% DMSO and 90% OP9 culture medium.

For CML and non-CML cells between 4x10^6-2x10^7 CD34$^+$-enriched cells or 10^8 MNCs were cryopreserved by adding an equal volume of 20% DMSO in 4.5% ALBA (Human Albumin Solution, Scottish National Blood Transfusion Service) to each cell suspension to obtain a final concentration of 10% DMSO (Dr. Alan Hair, Institute of Cancer Sciences, University of Glasgow). The cryotubes were transferred to a cryofreezer container and first cooled at -80°C to provide a controlled temperature reduction and then transferred to liquid nitrogen for long-term storage.

2.2.1.6 Recovering frozen primary cells

CML/non-CML CD34$^+$-enriched cells or MNCs were removed from liquid nitrogen and immediately thawed at 37°C in a water bath until the ice crystals had melted. Using a pastette, the cells were added to a 15 ml sterile falcon tube and recovered by drop-wise addition of 10 ml of DAMP over a 20-minute period. This step was performed at RTm to enhance the activity of the DNase I, with constant agitation to prevent clumping of the cells. The cells were centrifuged at 1,000 rpm for 10 mins and after discarding the supernatant, the pellet was washed twice in DAMP (1,000 rpm for 10 mins) and resuspended in SFM supplemented with physiological growth factors cocktail for counting and cell viability using the Trypan blue dye. The cells were then plated and cultured in SFM supplemented with physiological growth factors cocktail overnight in 75 cm^2 hydrophobic tissue culture flasks at ~2x10^6 cells/ml.

2.2.1.7 OP9-CD34$^+$ CD38$^-$ cells co-culture

The sorted CML CD34$^+$ CD38$^-$ cells were seeded at a density of 5x10^3-5x10^4 per well on top of confluent OP9-GFP cultures (Section 2.2.1.1). The OP9 media was replaced with 2 ml Myelocult H5100 media (StemCell Technologies) supplemented with freshly prepared and filter-sterilised hydrocortisone sodium succinate (StemCell Technologies) to give a final concentration of 10^{-6} M. The co-culture was maintained for 3 days at 37°C and 5% CO$_2$. A stromal-free condition was also set up in a well of a 96-well plate in the presence of 200 µl SFM supplemented with physiological growth factors to monitor the effect of OP9 stromal cells on the primitive cells. On day 3, the media in each well of the co-culture experiment was replaced by 1 ml collagenase IV/Myelocult (1 mg/ml) solution and incubated for 20 min at 37°C. Subsequently, 1 ml trypsin (Invitrogen) was added and the cells were incubated for an extra 15 mins at 37°C. The cells were harvested by using cell scrapers and 1 ml FBS was added to each well to neutralise trypsin. Cells were centrifuged at 300 g for 5 mins and washed once with PBS/2% FBS. After passing through 0.7 µm filters, cells were then used for gene expression analysis (Section 2.2.4.3), monitoring cell division progress (Section 2.2.2.2), or colony-forming cell (CFC) assays (Section 2.2.1.9).

The concentration of neurotransmitters used for the treatment conditions in OP9 co-culture assays are summarised in Table 2.1. The neurotransmitters were added to each well upon plating the CML CD34$^+$ CD38$^-$ cells.

Neurotransmitter	Supplier	Solvent	Stock concentration	Final concentration	Dilution 1	Dilution 2
DL-Norepinephrine HCl	Sigma-Aldrich	Water	200 mM	1 µM	1:200	1:1,000
Serotonin HCl	Sigma-Aldrich	Water	50 mM	5 µM	1:10	1:1,000
Histamine di-HCl	Sigma-Aldrich	Water	100 mM	100 µM	none	1:1,000
Acetylcholine chloride	Sigma-Aldrich	Water	100 mM	100 nM	1:1,000	1:1,000
L-glutamic acid	Sigma-Aldrich	1 M HCl	100 mM	10 µM	1:10	1:1,000

Table 2.1: The concentration of the neurotransmitters used in OP9 co-culture treatment assays.
Each neurotransmitter was prepared in stock solutions in the appropriate solvent. To achieve the final concentration levels two dilutions were performed; one with the Myelocult medium prior to adding to the well and the second one at the time of mixing with the medium in each well.

2.2.1.8 CFC and replating assays

After harvesting the OP9-CD34$^+$ CD38$^-$ co-culture cells on day 3, cells were stained for CD34-APC for 20 min in the dark followed by washing twice with PBS/2% FBS. The viable GFP$^-$ CD34$^+$ cells were isolated using the FACSAria cell sorting system and equal number of cells (3-5x10^3) in each condition was used to set up duplicate CFC assays for each treatment condition. The appropriate volume of cell suspension for duplicate wells was added to 2 ml of MethoCult H4034 Optimum (StemCell Technologies) methylcellulose-based medium and thoroughly mixed. 1 ml of this mixture was added to a 35 mm tissue culture dish and was gently swirled, so that the bottom of the dish was completely coated, and then incubated for 10 days at 37°C and 5% CO_2 under humidified conditions. At the end of this time, the number of viable erythroid (CFU-E and BFU-E), granulocyte-macrophage (CFU-GM, CFU-G, and CFU-M), and multipotential granulocyte, erythroid, macrophage, megakaryocyte (CFU-GEMM) colonies were counted in each dish.

On day 11, individual colonies from each plate were carefully plucked under the microscope in the hood with a p10 pipette and were inoculated into individual wells in a 96-well plate, which were covered with 100 µl of MethoCult medium and 10 µl of SFM. Even dispersion was achieved by aspirating the SFM into the colony containing pipette tip and carefully mixing with the MethoCult medium by pipetting action without introducing air. Normally, 50 colonies were plucked at random and the 96-well plate was incubated at 37°C and 5% CO_2 under humidified conditions for 10 days before measuring the replating efficiency which can be expressed as the total number and the types of colonies in each treatment condition.

2.2.2 Flow cytometry

Flow Cytometry or fluorescence-activated cell sorting (FACS) is a quantitative technique that permits the visualisation and sorting of cells by multiple parameters according to their fluorescence intensity after antibody labelling. A flow cytometer can also measure the size of a cell using forward scatter and the granularity of a cell using side scatter parameters. All the flow cytometric analyses were carried out employing a FACSCanto system (BD).

2.2.2.1 Isolation of CD34$^+$ CD38$^-$ and CD34$^+$ CD38$^+$ cells from recovered primary samples

The recovered primary cells were harvested after overnight culture and centrifuged at RTm (1,200 rpm for 5 mins). The pellet was washed once in PBS/2% FBS. After cell count, aliquots of cells (10^5 cells) were removed for appropriate isotype controls which were used to correctly set the detectors, so that the negative isotype population was placed in the first log decade for each flow cytometry channel. CD34-APC and CD38-FITC or CD38-PerCP (for co-culture experiments) single positive staining (2 µl of antibody per control) were also set up. The remaining cells were stained with both CD34-APC (5 µl/10^6 cells) and CD38-FITC (10 µl/10^{6}cells) or CD38-PerCP (5 µl/10^6 cells) antibodies and samples were incubated for 20 mins in the dark. Following incubation, samples were washed twice with PBS/2% FBS (1,200 rpm for 5 mins). The controls were resuspended in 200 µl PBS/2% FBS and the rest of the sample was resuspended in 2 ml PBS/2% FBS. All samples were passed through 0.7 µm filters to remove cell clumps before sorting. After adjusting the isotype and positive control thresholds (gates), CD34$^+$ CD38$^-$ and CD34$^+$ CD38$^+$ cells were then isolated using FACSAria (BD) cell sorting system.

2.2.2.2 Carboxyfluorescein diacetate succinimidyl ester (CFSE) analysis for monitoring cell division

2.2.2.2.1 CFSE staining

The use of CFSE intracellular fluorescence dye was initially demonstrated to be a powerful flow cytometry technique to monitor cell divisions *in vitro* and *in vivo* in lymphocytes (Lyons and Parish, 1994). CFSE molecule is highly permeable via cellular membranes and once inside the cell cannot be passively lost. CFSE covalently binds to the intracellular amino groups and is diluted between daughter cells at each division. CFSE was shown to stably label intracellular molecules with sequential halving of fluorescence intensity at every cell division for up to eight divisions before the fluorescence intensity gets reduced to background level (Lyons and Parish, 1994). CFSE staining was also successfully reported for HSCs (Nordon et al., 1997).

After isolating the CD34$^+$ CD38$^-$ population and prior to plating for OP9-GFP co-culture experiments, half of the cells were removed for CFSE analysis. Since we

were using GFP-tagged cells (FITC wavelength), the CellTrace Violet dye (Invitrogen) was used which could be monitored at the Pacific Blue wavelength. Cells were resuspended in 1 ml PBS/2% FBS and 1 µl CellTrace Violet dye (5 mM in DMSO) was added to give the final concentration of 5 µM. The staining suspension was subsequently incubated for 20 mins at 37°C and protected from light. The unbound dye was quenched by adding 10X the original staining volume of ice-cold PBS/20% FBS and washed at 300 g for 5 mins. After discarding the supernatant, the pellet was washed once in PBS/2% FBS. Cells were then resuspended in the co-culture or stromal-free media (SFM supplemented with physiological growth factors) and a 20 µl aliquot was removed, diluted 1:10 with PBS/2% FBS and used for cell counting by the CountBright Absolute Counting Beads (Invitrogen) using flow cytometry according to the manufacturer's instructions. The efficiency of CFSE staining was also established at this stage by flow cytometry. Additionally, a colcemid control was set up using CFSE-stained cells in stromal-free conditions to determine the position of the undivided population ($CFSE_{max}$ peak) after co-culture. The colcemid solution (100 µg/ml) was directly added to the cultured cells to give a final concentration of 100 ng/ml. The cells unstained for CFSE were used to adjust the voltage settings and optimally compensate for spectral overlap.

2.2.2.2.2 FACS analysis of CFSE experiments

Following harvesting the co-culture on day 3, a 20 µl aliquot was removed from each different condition for cell counting by the CountBright Absolute Counting Beads. CFSE-stained cells were washed once with PBS/2% FBS and used for flow cytometry. Isotype controls were used to correctly set the detectors, so that the negative isotype population (CFSE$^-$ cells) was placed in the first log decade for each flow cytometry channel. The colcemid-treated assay was used as the CFSE$^+$ control and also to identify the undivided population, which was termed as the $CFSE_{max}$ peak. The co-cultured cells were stained for CD34-APC for 20 mins in the dark followed by washing with PBS/2% FBS. The CFSE fluorescence was measured in the GFP$^-$ (FITC$^-$) and CFSE$^+$ (Pacific Blue$^+$) population.

2.2.2.2.3 Calculation of the undivided ($CFSE_{max}$) cell population

To determine the anti-proliferative effect of different treatment conditions and assess the size of the non-proliferating primitive progenitor population, the

percentage recovery of viable CD34$^+$ population in the undivided population remaining after co-culture was assessed. The number of viable cells harvested from each culture condition was recorded, as was the percentage of CD34$^+$ cells found in the undivided fraction (CFSE$_{max}$). Percentage recovery of input cells in the undivided peak could then be calculated by dividing the absolute number of CFSE$_{max}$ CD34$^+$ cells by the total number of input CD34$^+$ cells (day 0) and multiplying by 100%. This allowed direct comparison of different treatment conditions on the non-proliferating primitive progenitor population.

2.2.3 Total RNA extraction for Affymetrix and TaqMan-based RT-PCR gene expression analysis

Total RNA was extracted by RNeasy Micro Kit (QIAGEN, Crawley, UK) for sample sizes smaller than 5x10^5 cells such as the sorted CD34$^+$ CD38$^-$ cell populations according to the manufacturer's instructions. RNeasy Mini Kit (QIAGEN) was used for sample sizes between 5x10^5-1x10^7 cells such as the sorted CD34$^+$ CD38$^+$ populations according to the manufacturer's instructions. QIAshredder columns (QIAGEN) were used to homogenise the cell lysates according to the manufacturer's instructions. All the RNA extraction samples were DNase I-treated on-column using the ribonuclease (RNase)-Free DNase Set (QIAGEN) according to the manufacturer's instructions. The purity of RNA sample (A260/280) and its concentration were measured by NanoDrop ND-1000 spectrophotometer. The Agilent Bioanalyzer and Agilent RNA 6000 Nano Kit (Agilent Technologies, California, USA) were used to measure the integrity of the RNA samples according to the manufacturer's instructions.

2.2.4 PCR-based gene expression analyses

2.2.4.1 384-well TaqMan Micro Fluidic cards

180 ng RNA was converted to cDNA for each sample using High Capacity RNA-to-cDNA Kit (Applied Biosystems, Warrington, UK) according to the manufacturer's instructions. A pool of RNA was used for normal CD34$^+$ CD38$^-$ cells (60 ng of RNA from each individual sample), but individual RNA samples were used for CML CD34$^+$ CD38$^-$ cells.

Custom 384-well TaqMan Micro Fluidic cards (Applied Biosystems) were designed for Affymetrix gene expression validation analysis. The 96a format was used which consists 96 TaqMan gene expression assays pre-implanted in four replicates. Each card uses eight sample-loading ports, each connected to 48 reaction-wells. Each cDNA sample was divided into triplicates; therefore, for each port 30ng equivalent cDNA (60 ng per replicate) was added to 100 µl of TaqMan Gene Expression Master Mix (Applied Biosystems), and samples run on the Applied Biosystems 7900HT Real-Time PCR System according to the manufacturer's instructions. Data was quantified using RQ Manager Analysis and DataAssist softwares (Applied Biosystems). Relative gene expression was calculated using the $2^{-\Delta\Delta CT}$ method. The RQ ratio between target gene and endogenous controls were plotted as a measure of mRNA gene expression. The full list of TaqMan probes used to design the custom arrays can be found in the Appendix 1.

2.2.4.2 Fluidigm Dynamic Array integrated fluidic circuits (IFCs)

The 48.48 Dynamic Array IFCs (Fluidigm, Amsterdam, The Netherlands) were used to analyse the gene expression levels of primary haemopoietic cells in the OP9-GFP co-culture system. Cells were stained with CD34-APC and CD38-PerCP after harvesting on day 3 for 20 mins in the dark followed by washing twice in PBS/2% FBS. 300 GFP$^-$ CD34$^+$ CD38$^-$ viable cells from each treatment condition were sorted directly into the pre-amplification master mix. The pre-amplification master mix could be made up of up to 100 TaqMan probes at 0.2X final concentration. Each master mix was prepared as follows: 5.0 µl CellsDirect 2X reaction mix (Invitrogen), 2.5 µl 0.2X probe mix, 0.2 µl SuperScript III RT/Platinum Taq Mix (Invitrogen), 1.3 µl Tris-EDTA (TE) buffer (pH 8.0), and 0.1 µl RNase inhibitor SUPERase-In (Ambion). Cells could also be sorted to the pre-amplification master mix lacking the probe mix and the SuperScript III enzyme to be kept at -20°C until required. The negative control sample was also set up at this stage by sorting the cells into a pre-amplification master mix in the absence of SuperScript III enzyme. The pre-amplification mix was then placed in a thermal cycler using the following programme: 50°C for 15 mins (reverse transcription step), 95°C for 2 mins (Taq activation), and 18 cycles of 95°C for 15 secs and 60°C for 4 mins. At the end of pre-amplification the samples were diluted 1:5 by the TE buffer (pH 8.0).

The 48.48 Dynamic Array IFCs format would allow analysis of up to 48 gene expression assays (45 tests + 3 housekeeping genes) for 16 samples in triplicates (14 tests + negative control + water). Dynamic Array IFCs were primed according to the manufacturer's instructions in the IFC Controller MX system (Fluidigm). 48 10X assay mixes were prepared by mixing 2.5 µl each 20X TaqMan Gene Expression Assay (Applied Biosystems) and 2.5 µl 2X Assay Loading Reagent (Fluidigm) and loaded to the probe inlets of the arrays. Each cDNA sample (individual replicates) was added to the samples inlet by mixing 2.5 µl 2X TaqMan Universal PCR Master Mix (Applied Biosystems), 0.25 µl 20X GE Sample Loading Reagent (Fluidigm), and 2.25 µl pre-amplified cDNA. The chip was then inserted in the IFC Controller MX system to complete the loading process according to the manufacturer's instructions. Finally, the chip was transferred to the BioMark HD system (Fluidigm) to perform the RT-PCR reactions followed by scanning and data collection by the BioMark Data Collection software (Fluidigm). Data analysis was performed in the Fluidigm Real-Time PCR Analysis software and the raw data were exported to Microsoft Office Excel to calculate the differential gene expression using the $2^{-\Delta\Delta CT}$ method. The full list of TaqMan probes used for the Fluidigm analysis can be found in the Appendix 2.

2.2.4.3 Primer design for gene expression analyses

The TaqMan assays used for the Micro Fluidic cards and Fluidigm Dynamic Array IFCs were selected from a catalogue of TaqMan gene expression assays available from Applied Biosystems website (http://bioinfo.appliedbiosystems.com/genome-database/gene-expression.html). The selected primers could amplify all alternatively spliced transcripts of the genes of interest. All TaqMan probes were designed as FAM-conjugated. For Fluidigm gene expression analyses, in particular, probes were selected that amplify <100bp amplicon sizes and from exon-intron junctions exclusively as the pre-amplification master mix cannot be DNase I-treated.

2.2.5 Gene expression analysis using Affymetrix GeneChip Whole-Transcript (WT) Gene Arrays

2.2.5.1 Preparation of labelled target DNA and array hybridisation

After verifying the purity, integrity, and concentration of total RNA by the Bioanalyzer (Section 2.2.3.2), 50 ng of total RNA was used to prepare sense DNA targets using random priming-based Ambion WT Expression Kit (Ambion) according to the manufacturer's instructions. The synthesised sense DNA strands were fragmented and labelled using the GeneChip WT Terminal Labelling Kit (Affymetrix, High Wycombe, UK) which contains terminal deoxynucleotidyl transferase (TdT), according to the manufacturer's instructions. The HuGene-1_0-st-v1 Affymetrix arrays were hybridised, stained, and washed using GeneChip Hybridization, Wash, and Stain Kit (Affymetrix) on the GeneTitan Instrument (Affymetrix), according to the manufacturer's instructions.

2.2.5.2 Preprocessing of raw image data

The raw Affymetrix image data (.CEL) files were imported in Bioconductor (www.bioconductor.org), which is an R programming environment (Gentleman, 2005). For data preprocessing, *affy* and *aroma.affymetrix* packages were used in Bioconductor, which perform Robust Multi-array Averages (RMA) background correction and quantile normalisation on the .CEL files. *Aroma.affymetrix package* uses RMA probe-level model (PLM) to obtain gene-level summaries for each gene (www.aroma-project.org). The array quality was assessed at this stage by the Normalised Unscaled Standard Error (NUSE) plot and the Relative Log Expression (RLE) plot. Furthermore, an extra quality control step was introduced by the *simpleaffy* package to generate the RNA degradation plots. Differential expression analysis was performed using the *limma* package which fits the microarray data into a linear model using an empirical Bayes method and coupled to the Benjamini Hochberg (BH) false discovery rate (FDR) control ((Benjamini and Hochberg, 1995) with an adjusted p-value of 0.05. The *annotate* package was used to annotate the significantly differentially expressed probe sets using the HuGene-1_0-st-v1 transcript cluster database. The *made4* and *rgl* packages were used to perform three-dimensional (3D) principal component analysis (PCA). More information about the mentioned Bioconductor packages could be found in their 'vignettes' available from the Bioconductor project website.

2.2.5.3 Biological pathways analysis for gene expression differences

Significantly differentially expressed genes in each analysis were exported to PANTHER database (www.pantherdb.org) (Thomas et al., 2003) to identify the pathways with significant numbers of changes and the pathways that show an up- or downregulation in a cell type relative to others. An input list of differentially expressed genes with their differential fold change values was provided. The statistical tool then created a distribution of values for the input dataset, which was then used as a reference for distribution. Subsequently, the input dataset was divided into functional categories, and the distribution of values in each group was then calculated and compared against the reference distribution. The probability of obtaining a random distribution from the reference distribution in each category was measured by the Mann-Whitney Rank-Sum test (*U*-test) (Clark et al., 2003). The output data was reported as p-values associated with the direction of shift for the distribution of values in each group relative to the reference distribution. The categories with non-random distribution were associated with small and significant p-values.

2.2.6 ChIP

2.2.6.1 Conventional ChIP assays

Crosslinking

1×10^8 E14 cells were collected by centrifuging at 1,200 rpm for 5 mins at RTm and resuspended in 50 ml serum-free DMEM media in a glass flask. Cells were crosslinked with 1,355 µl formaldehyde solution (38%; final concentration 1.0%) and incubated at RTm with gentle agitation for 15 mins. The crosslinking reaction was stopped by adding 3,425 µl 2 M glycine (final concentration of 0.125 M) followed by incubation for 5 mins at RTm with gentle agitation.

Cell and nuclear lysis

Cells were transferred to a 50 ml Falcon tube (kept on ice) and were centrifuged at 1,200 rpm for 6-8 mins at 4ºC. The pellet was washed with 1.5 ml ice-cold PBS (2,000 rpm for 5 mins at 4ºC). The cell pellet was gently resuspended in 1.5X pellet volume of CLB and mixed well by pipetting followed by incubation on ice for

10 mins. Nuclei were subsequently collected by centrifuging at 2,500 rpm for 5 mins at 4ºC. After removing the supernatant, the nuclei were resuspended in 1.2 ml of NLB followed by incubation on ice for 10 mins.

Sonication

0.72 ml of IPDB was added to the nuclear lysate, which was then transferred to a 5 ml glass falcon tube. Lysate was sonicated using the Sanyo/MES Soniprep sonicator to shear the crosslinked DNA into 300 – 1,000 bp fragments. The tip of the probe was maintained under the surface to avoid making bubbles which cause inefficient sonication. The samples were kept on ice at all times. Chromatin was sonicated for 11 mins (8 pulses of 30 secs 'on' and 1 min 'off') at 14 microns power amplitude. The sheared chromatin was then transferred to a 2 ml eppendorf tube and centrifuged at 13,000 rpm for 5 mins at 4ºC.

Pre-clearing

The supernatant was transferred to a 15 ml falcon tube and 4.1 ml IPDB was added to give the NLB:IPDB ratio of 1:4. Subsequently, 100 µl normal rabbit IgG was added to the chromatin solution and incubated for 60 mins at 4ºC on a rotating wheel for pre-clearing. The IgG antibodies were sequestered by adding 200 µl homogeneous protein G-agarose beads suspension (100 µl beads bed volume) followed by incubation for 3 hours (hrs) at 4ºC on a rotating wheel. The beads were centrifuged at 3,000 rpm for 2 mins at 4ºC. The pre-cleared chromatin solution could be stored at -80°C at this stage until required.

IP

The supernatant from previous step was used to set up various IP conditions in a 2 ml eppendorf tube. For each IP condition approximately 10% of the chromatin solution was used (equivalent to 10^7 cells). An NLB:IPDB buffer at the ratio of 1:4 was freshly prepared and used to ensure that the final volume of all ChIP conditions was 1,350 µl. ChIP conditions were set up as follows: 675 µl chromatin + 675 µl NLB:IPDB buffer + 10 µg of antibody raised against the histone modification of interest. Moreover, 270 µl of chromatin was aliquoted and stored

at -20°C to be used as an input sample for array hybridisations or deep sequencing.

ChIP samples were incubated overnight at 4ºC with rotation. The following day, samples were centrifuged at 13,000 rpm for 5 mins at 4ºC. The supernatants were transferred to a 1.5 ml eppendorf tubes and kept on ice whenever possible. The antibodies were sequestered by adding 50 µl homogenous protein G-agarose beads suspension (25 µl beads bed volume) followed by incubation for 3 hrs at 4ºC with rotation. Subsequently, protein G-agarose beads were centrifuged at 13,000 rpm for 20 seconds at 4ºC and the supernatant was discarded. The protein G-agarose beads were carefully washed, for which the wash buffer was added, the samples were vortexed briefly, were centrifuged at 7,500 rpm for 2 mins at 4°C and left to stand on ice for 1 min before removing the supernatant. The washes were carried out in the following sequence: (i) the beads were washed twice with 750 µl of cold IPWB1. The beads were transferred to a 1.5 ml eppendorf tube after the first wash; (ii) the beads were washed once with 750 µl of cold IPWB2; (iii) the beads were washed twice with 750 µl of cold TE buffer (pH 8.0). Finally, DNA-protein-antibody complexes were eluted from the beads by adding 225 µl of IPEB followed by vortexing and centrifugation at 7,500 rpm for 2 mins. Subsequently, the supernatant was collected in a fresh 2.0 ml eppendorf tube. The bead pellet in the original tube was resuspended in 225 µl of IPEB again, briefly vortexed and centrifuged at 7,500 rpm for 2 mins. Both of the elutions were combined in the same tube.

Crosslinks reversal and DNA purification

The reversal of the crosslinks was carried out on the input sample which was stored at -20°C previously. This was achieved by adding 0.1 µl RNase A (10 mg/ml, 50 Kunitz units/mg, ICN Biochemicals) and 16.2 µl 5 M NaCl (to the final concentration of 0.3 M) to the input DNA sample. Similarly, 0.2 µl RNase A (10 mg/ml, 50 Kunitz units/mg) and 27 µl 5 M NaCl (to a final concentration of 0.3 M) were added to each of the IP test samples. All the samples including the input DNA sample were incubated at 65ºC for 6 hrs to reverse the crosslinks. Subsequently, 9 µl proteinase K (10 mg/ml, 20 U/mg, Invitrogen) was added to each sample to digest the proteins followed by incubation at 45ºC overnight. DNA was purified by adding 250 µl phenol (Sigma-Aldrich) and 250 µl chloroform to

each sample. The samples were vortexed and centrifuged at 13,200 rpm for 5 mins at RTm. The aqueous layer (top layer) was collected in fresh 2.0 ml eppendorf tubes and 500 µl chloroform was added to each sample. The samples were vortexed and centrifuged at 13,200 rpm for 5 mins at RTm. The aqueous layer was transferred to fresh 2.0 ml eppendorf tubes. DNA was precipitated by adding 5 µg glycogen (5 mg/ml; Roche, UK), 70 µl 3 M sodium acetate (pH 5.2), and 1,375 µl 100% ethanol to each sample followed by incubation at -20°C overnight. After overnight precipitation, samples were centrifuged at 14,000 rpm for 20 mins at RTm. The DNA pellets were washed with 500 µl of ice-cold 70% ethanol and air-dried for 10-15 mins. The DNA pellets of the IP samples were resuspended in 50 µl nuclease free water and 100 µl for the input DNA samples. 5 µl of each sample was run on a 1% agarose 1X TBE gel and visualised with SYBR Safe DNA gel stain to check the DNA size. Samples were stored at -20°C.

2.2.6.2 Low-cell ChIP assays

<u>Crosslinking</u>

$1x10^5$ cells were collected by centrifuging at 1,200 rpm for 5 min at RTm and resuspended in 1 ml of plain DMEM or IMDM in a 2 ml eppendorf tube. Cells were fixed by adding 26.3 µl formaldehyde solution (38%; final concentration 1.0%) followed by incubation at RTm with gentle agitation for 15 mins. The crosslinking was stopped by adding 68.5 µl 2 M glycine (final concentration of 0.125 M) followed by incubation for 5 mins at RTm with gentle agitation to stop the crosslinking reaction.

<u>Cell and nuclear lysis</u>

Cells were centrifuged at 2,200 rpm for 6-8 mins at 4ºC and the pellet was washed with 1.5 ml ice-cold PBS (4,000 rpm for 5 mins at 4ºC). Subsequently, cells were directly resuspended in 1.2 ml of NLB and incubated on ice for 10 mins.

<u>Sonication</u>

The sonication step was performed similar to the conventional ChIP. And the sonicated DNA was directly used to set up the IP conditions without pre-clearing.

IP

The chromatin solution was used to set up various IP conditions in a 2 ml eppendorf tube. For each IP condition approximately 10% of the chromatin solution was used (equivalent to 10^4 cells). An NLB:IPDB buffer at the ratio of 1:4 was freshly prepared and used to ensure that the final volume of all ChIP conditions was 1,350 μl. ChIP conditions were set up as follows: 675 μl chromatin + 675 μl NLB:IPDB buffer + 0.5-5.0 μg of antibody raised against histone modification of interest. The antibody concentrations were optimised for each modification in the low-cell ChIP. 270 μl of chromatin was aliquoted and stored at -20°C to be used as an input sample for array hybridisations or deep sequencing.

The rest of IP steps, bead washes, elution, crosslinks reversal, and DNA purification steps were performed similar to the conventional ChIP assays. However, the gel electrophoresis of the purified ChIP DNA was not performed as the DNA concentration was lower than could be detected on the gel. The quality of low-cell ChIP was measured by ChIP-quantitative PCR (qPCR) instead (Section 2.2.7).

2.2.7 ChIP-qPCR

ChIP-qPCR reactions were performed using SYBR green-based chemistry in a Stratagene Mx3000P qPCR System (Agilent Technologies) to quantify the DNA fragments enriched for immunoprecipitated histone modifications. Each ChIP-qPCR reaction was set up by adding 5 μl of low-cell ChIP-purified DNA (diluted 1:5) or ChIP-sequencing (ChIP-seq) library (dilute 1:10); 12.5 μl 2X FastStart Universal SYBR Green Master (Rox) (Roche) and each of the forward and reverse primers at the final primer concentration of 300 nm in a final volume of 25 μl. 5 ng of input or human genomic DNA were used as reference for determining absolute enrichment levels. ChIP-qPCR reactions were performed in duplicate. In order to determine the relative fold enrichment levels negative control primers were designed amplifying previously identified genomic regions that do not interact with the histone modification of interest. The thermal cycler programme used was as follows: 95°C for 10 mins, followed by 40 cycles of 95°C for 30 secs, and 60°C or 63°C for 60 secs. Melt curve analysis was carried out between 72-95˚C at the rate of 1˚C/second with continuous measurement of fluorescence to ensure that there

was only a single product present in each reaction and to allow for detection of primer dimers. The relative fold enrichments were calculated as follows:

$$\Delta Ct_{\text{ test primer}} = Ct_{\text{ input}} - Ct_{\text{ test primer}}$$

$$\Delta Ct_{\text{ negative control}} = Ct_{\text{ input}} - Ct_{\text{ negative control}}$$

$$\text{Relative fold enrichment} = 2^{(\Delta Ct \text{ test primer} - \text{median } \Delta Ct \text{ negative controls})}$$

2.2.7.1 Primer design

Primers were designed using online primer designing tools Primer3 (http://frodo.wi.mit.edu/; Rozen and Skaletsky, 2000) and PrimerQuest (Integrated DNA Technologies, Leuven, Belgium). Primer pairs were designed to amplify genomic regions ranging from 70-150 bp, with an optimum melting temperature of 60°C, and an optimum length of 20 bp. The specificity of designed primers was checked by the BLAT and In-Silico PCR tools available from the UCSC Genome Browser website. Primers were made by standard desalting procedure in the stock solutions of 100 μM (TE pH 8.0 buffer) (Integrated DNA Technologies).

The primers designed for the validation of low-cell ChIP efficiency were chosen from the known interacting and non-interacting regions with the histone modifications of interest. On the other hand, to monitor the mouse genomic contamination in the low-cell ChIP-purified DNA and low-cell ChIP-seq libraries, primers were designed to amplify multiple mouse genomic regions (Appendix 3).

The primers designed for the validation of ChIP-seq profiles (Appendix 4) were selected from a 1 kb region spanning the TSS based on the ENSEMBL annotation for the major protein-coding transcript. The sequences of DNA fragments were obtained from the UCSC Genome Browser by entering the genomic coordinates of the promoter regions.

2.2.8 ChIP-chip

2.2.8.1 Random priming labelling of ChIP-purified DNA samples

The hybridisation of ChIP-purified DNA samples on the mouse tiling path arrays (printed at the Sanger Institute, University of Cambridge) were set up on the Tecan

HS 4800 (an automated hybridisation station). The DNA samples were fluorescently-labelled using BioPrime DNA Labelling System (Invitrogen) as follows:

300-500 ng input DNA or 20% of ChIP DNA was mixed with 60 µl 2.5X random primers (Invitrogen) and the final volume was adjusted to 130.5 µl with labelling water (Invitrogen). The samples were denatured in a 100°C heat block for 10 mins followed by snap-chill on ice. Subsequently, 15 µl dNTPs for labelling, 1.5 µl Cy3 (for ChIP samples) or Cy5 (for input samples) labelled dCTP (GE Healthcare), and 3 µl Klenow fragment (Invitrogen) were added to the DNA/random primers solution and mixed gently by pipetting. The labelling reactions were incubated overnight at 37°C and stopped by adding 15 µl of stop buffer (Invitrogen). The unlabelled nucleotides were removed by Micro-spin G50 columns (GE Healthcare) according to the manufacturer's instructions. The purified labelled DNA samples were analysed on a 1% agarose 1X TBE gel and visualised by SYBR Safe DNA gel stain.

2.2.8.2 Competitive hybridisation of labelled ChIP DNAs onto microarrays

The Cy3-labelled ChIP samples (~180 µl) and the Cy5-labelled input samples (~180 µl) were mixed and 135 µl of human or mouse Cot1 DNA, 80 µl 3 M sodium acetate (pH 5.2), and 1,600 µl 100% ice-cold ethanol were added and incubated overnight at -20°C in the dark to precipitate the labelled DNA. The precipitated DNA samples were centrifuged at 13,000 rpm for 20 mins at RTm. After discarding the supernatant, DNA pellets were washed with 500 µl 80% ice-cold ethanol (13,000 rpm for 5 mins at RTm). Residual levels of ethanol were aspirated and the pellets were resuspended in 150 µl of Tecan hybridisation buffer at 70°C with agitation for 2-3 mins. The samples were then denatured on a 100°C heat block for 10 mins followed by snap-chill on ice. The hybridisation reactions were kept at 37°C in the dark for 60 mins.

Prior to the hybridisation, the slide holders and the slide chambers of Tecan HS 4800 were carefully cleaned and glass slides were loaded on the slide holder. The hybridisation station was washed thoroughly (automated system) by HPLC water and dried with blasts of nitrogen gas. The station was also primed with all wash buffer solutions to remove the air bubbles in the liquid channels and tubing. The

glass slides were subsequently replaced with the microarrays and ~150 µl of Tecan hybridisation buffer was injected using a positive displacement pipette to initiate the pre-hybridisation step, which was performed at 37°C for 60 mins. The microarray slides were washed with PBS/0.05% Tween 20 and dried with blasts of nitrogen gas. At this stage, 150 µl of hybridisation reaction was injected slowly onto the slide using the displacement pipette. The hybridisation step was performed at 37°C for 48 hrs in the dark with medium agitation. The slides were washed using the programme shown in Table 2.2 and dried with nitrogen gas. Microarray slides were stored in a dark, low temperature, low humidity environment until ready for scanning to prevent loss of fluorescent signal.

Steps	Wash solutions	Temperature	No. of washes	Wash duration	
				Wash time	Soak time
1	PBS/0.05% Tween 20	37°C	10	60 secs	30 secs
2	0.1X SSC	52°C	5	60 secs	2 mins
3	PBS/0.05% Tween 20	RTm	10	60 secs	30 secs
4	HPLC water	RTm	2	30 secs	n/a

Table 2.2: Wash steps for the hybridisations performed on Tecan HS 4800.
The wash buffers were prepared in advance using HPLC water (Section 2.1).

2.2.8.3 Scanning and processing of ChIP-chip data

Cy3 and Cy5 images at 5 µm resolution were acquired using ScanArray 4000 confocal laser-based scanner with a laser power of 100% a photo multiplier tube (PMT) value of between 70-85%. ScanArray Express (PerkinElmer) was used to quantitate the fluorescent intensities of the spots using the adaptive circle quantitation and the TOTAL normalisation methods. This software can automatically locate the spot position on the scanned image of the array to obtain the signal intensity values. Mean intensity ratios (intensity-background) were reported for each spot representing an array element (.csv file formats). Those spots identified as 'not found' were removed from the data set. Further analysis of the ChIP-chip data was carried out in a Microsoft Office Excel spreadsheet in which each array element was associated with its genomic sequence position information. The mouse tiling path array data were visualised in the UCSC Genome Browser by uploading a 'wiggle' file which contained chromosome start and end coordinates and fold enrichment ratios.

2.2.8.4 ChIPOTle and BLOCs algorithms in ChIP-chip data analysis

ChIPOTle is a Microsoft Office Excel-based application that performs ChIP-chip data analysis using a window-sliding approach (Buck et al., 2005) This method will identify the binding sites as peaks and a confidence value is calculated based on the number of array elements used to define the peak. This method is particularly useful to detect target sites of the proteins whose binding frequency is approximately less than three times the window size and therefore, was used to analyse histone modifications that form peaks <2 kb such as H3K4me3. The window size in ChIPOTle is adjusted to approximately the average sheared chromatin size. Larger window sizes tend to include noise from adjacent loci whereas smaller windows will not be able to detect data from neighbouring array spots. Window size is independent of the array resolution. Step size is another parameter which depends on both window size and the array resolution. It should be adjusted to less than the half of the length of array elements and less than or equal to a quarter of the window size. Finally, the p-value cutoff should be set to account for FDR. Therefore, the ChIPOTle parameters were set as 500 bp window size with 125 bp step size and 0.0001 p-value cutoff.

The BLOCs algorithm was developed in Perl language to identify the broad local enrichments larger than 5 kb, such as H3K27me3, which are significantly above the background (Pauler et al., 2009). Each BLOC initiates with 10-13 consecutive array elements where 10 of these elements show positive $\log_2$ fold enrichments and ends with 6-8 successive elements where six of them show negative $\log_2$ fold enrichments. BLOCs larger than 5kb with a median $\log_2$ value above the 0.25X standard deviation of all the signals on the array were reported. A browser extensible data (BED) file which contained the genomic coordination of BLOCs enrichment domains was generated for UCSC Genome Browser visualisation.

2.2.9 ChIP-seq

2.2.9.1 Preparation of ChIP-seq libraries for Illumina Genome Analyzer

The preparation of ChIP-seq DNA libraries for Illumina Genome Analyzer II consists of five major steps, which include end-repair of ChIP DNA fragments to generate blunt-ended fragments, dA-tailing, ligation of adapters, PCR amplification of adapter-ligated fragments, and gel size-selection. The ChIP-seq libraries were

prepared similarly for both low-cell and conventional ChIP assays. The ChIP-seq DNA Sample Prep Kit (Illumina) and the NEBNext DNA Sample Prep Reagent Set 1 (New England BioLabs, Hitchin, UK) were used to prepare the ChIP-seq libraries as follows:

End-repair

The end repair reaction was set up as follows: 30 μl of low-cell ChIP DNA or 20 μl of conventional ChIP DNA, 5 μl 10X Phosphorylation Reaction Buffer, 2 μl dNTP mix (10 mM), 1 μl T4 DNA Polymerase, 1 μl Klenow DNA Polymerase (diluted 1:5), 1 μl T4 Polynucleotide Kinase (PNK), and the final volume was adjusted to 50 μl with ultra-pure nuclease-free water. The reaction mix was incubated in a thermal cycler for 30 mins at 20°C. The blunt-ended fragments were purified using QIAquick PCR Purification Kit (QIAGEN) in a final elution volume of 34 μl, according to the manufacturer's instructions.

Deoxyadenosine (dA)-tailing

The 3′-dA overhang was added to the blunt-ended fragments by mixing the purified DNA from the previous step with 10 μl dATP (1.0 mM), 5 μl 10X Klenow Fragment (3′→5′ exo⁻) Buffer, and 1 μl Klenow Fragment (3′→5′ exo⁻) followed by 30 mins incubation at 37°C. The dA-tailed fragments were then purified using MinElute PCR Purification Kit (QIAGEN) in a final elution volume of 10 μl, according to the manufacturer's instructions.

Ligation of adapters

The sequencing adapters were ligated to the DNA purified in the previous step in the following reaction: 1 μl Illumina Adapter Oligo Mix (diluted 1:10), 15 μl 2X Quick Ligation Reaction Buffer, and 4 μl Quick T4 DNA Ligase. The reaction was incubated for 15 mins at RTm. The unligated adapters were removed by using the QIAquick PCR Purification Kit in a final elution volume of 36 μl, according to the manufacturer's instructions.

PCR amplification of adapter-ligated ChIP DNA

The adapter-ligated DNA fragments were amplified by preparing the following PCR reaction mix: 10 µl 5X Phusion High-Fidelity (HF) Buffer, 1.5 µl dNTP mix (10 mM), 1 µl Illumina PCR Primer 1.1, 1 µl Illumina PCR Primer 2.1, and 0.5 µl Phusion DNA Polymerase. The thermal cycler conditions were set as follows: 98ºC for 30 secs; 18 cycles of 98ºC for 10 secs, 65ºC for 60 secs, and 72ºC for 30 secs; and finally 72°C for 5 mins. The PCR products were purified by MinElute PCR Purification Kit in a final elution volume of 10 µl, according to the manufacturer's instructions.

Agarose gel size-selection of amplified libraries

The purified PCR-amplified DNA fragments were loaded onto a 2% agarose 1X TAE gel in alternate wells to avoid cross-contamination of the samples. 1 kb ladder (New England BioLabs) was also loaded as a marker for the size-selection procedure. The electrophoresis was performed for 2 hrs with 60-70 voltage power. Subsequently, the gel was stained with 50 µl ethidium bromide for 30 mins in sterile water or 1X TAE buffer. After discarding the staining buffer, DNA fragments were visualised on the UV transilluminator and the bands corresponding to 200-300 bp size range were excised from each sample lane using a scalpel blade. The excised DNA fragments were purified using QIAquick Gel Extraction Kit (QIAGEN) in a final elution volume of 30 µl, according to the manufacturer's instructions.

At the end of this step the ChIP-seq DNA libraries were analysed by the Agilent Bioanalyzer to determine the average length of library fragments for flowcell cluster generation purposes. The quality of ChIP-seq libraries was assessed by qPCR as described in Section 2.2.7.

2.2.9.2 Processing of the ChIP-seq results

2.2.9.2.1 Sequence alignments

Sequence reads of 76 bp were obtained from Illumina Genome Analyzer II platform. Using the Solexa analysis pipeline (Dr. Peter Saffrey, Institute of Cancer Sciences, University of Glasgow), the sequence reads were clipped back to 36 bp. All reads were aligned to the reference genome - human (UCSC hg18) and mouse (UCSC mm8) – using the Bowtie algorithm (Langmead et al., 2009) and only uniquely aligned reads were retained. In all analyses, reads of multiple identical

copies (duplicate reads) were excluded. The uniquely aligned reads were extended to 150 bp, which was the average length of ChIP-seq libraries (without the length of sequencing primers), and converted to BED files detailing the genomic coordinates of each tag. In order to visualise the ChIP-seq data sets on the UCSC Genome Browser the BED files were converted to bigWig files which are in an indexed binary format and therefore, faster to upload and process for the larger data sets. Normalised bigWig files were also generated by normalising the number of sequencing tags in each genomic coordinate to the total number of uniquely aligned reads in the relevant ChIP-seq dataset (library size) (Dr. Andrew Crossan, Institute of Cancer Sciences, University of Glasgow).

Previously published ChIP-seq data sets for haemopoietic progenitor cells (HPCs) (Cui et al., 2009) were obtained from publicly available GEO database under the accession number of GSE12646. The ChIP-seq data sets for H3K4me3 and H3K27me3 were analysed similar to above.

2.2.9.2.2 SICER and BayesPeak algorithms

SICER is a spatial clustering approach which uses the enrichment context of a local window to determine the significance of enrichment domains (Zang et al., 2009). This approach employs a control ChIP-seq library, normally generated from the input DNA to identify potential islands. This involves identifying clusters of windows occupied by reads that are unlikely to appear by chance. All eligible summary windows were first identified, with an eligible window being defined as those windows satisfying a required window tag-count threshold determined from a preset p-value based on a Poisson background model. Islands were then identified by grouping consecutive eligible windows allowing gaps of at most two ineligible windows. The resulting islands identified for each library have an estimated FDR of less than 1%. SICER was used for the H3K27me3 and H3K36me3 ChIP-seq data sets with a gap size of 1kb.

BayesPeak algorithm (Spyrou et al., 2009) was used to identify the H3K4me3 ChIP-seq enrichment regions using a Bayesian hidden Markov model. This method uses a binomial distribution as opposed to the Poisson distribution used in SICER and therefore, overdispersion can be taken into account in measuring the abundance of sequencing reads at various genomic locations. Default parameters were used in this study.

2.2.9.2.3 seqMINER clustering tool

The extended BED files were imported into the seqMINER tool, a *k*-means clustering approach (Ye et al., 2011), for classifying all annotated transcripts in human hg18 build based on the histone modification enrichments at their promoters. The TSSs of all annotated transcripts were used as the reference coordinates or the middle values. Each putative promoter was defined as 2,500 bp upstream and downstream of the reference coordinate, also considering the orientation of the transcripts. Tag densities were defined as the number of tags present or overlapping in that 5 kb window. Data sets were linearly normalised to reduce the bias generated as a result of inherent ChIP-seq variations between samples. The number of clusters is defined by the user and the output data can be visualised as clusters of heatmaps based on the similarities observed between groups of transcripts across all analysed samples. Initially, all the annotated transcripts were divided into 20 clusters, from which four major classes of promoters could be detected including H3K4me3, bivalent, H3K27me3^{hi}, and neither modifications. The genes identified in each promoter class were then reclustered with twice as many original clusters that were associated with each promoter class to identify the falsely categorised transcripts, which were subsequently transferred to the appropriate promoter class.

2.2.9.3 Biological pathways analysis for epigenetic classes

The list of genes in each cluster was exported to PANTHER database to identify the significantly affected pathways in each cluster. For the analysis a reference list was provided which contained all the annotated gene promoters in the hg18 build. The reference list was divided into groups based on functional classifications. An input list of genes associated with each promoter class was also provided, which was divided into similar categories as the reference. A binomial test was then applied to statistically determine how significantly a group of genes is over- or under-represented in the input list relative to the reference by taking into account the number of genes expected in the input list for a particular category, which was calculated by the number of genes in that category in the reference list and the size of input list, and subsequently measuring the p-value by comparing the number of observed genes in each category to the expected values (Mi et al., 2010).

2.2.9.3.1 Composite plots for ChIP-seq data sets

The composite plots for different promoter classes identified by the seqMINER were generated by dividing the defined promoters into 100 equally-sized bins and the maximal number of overlapping tags for each histone modification was averaged in each bin for each promoter class. The library-size normalised average tag densities of each histone modification (y-axis) were plotted over an average promoter length (x-axis).

2.2.10   Statistical analysis

Most of the statistical analysis was performed using Microsoft Office Excel. This included the calculation of standard deviations and confidence intervals, which were calculated using the STDEV and CONFIDENCE (α=0.05) functions, respectively. The standard error of mean (SEM) was calculated by dividing the standard deviation of dataset by the square root of dataset size in Excel.

For the Spearman's rank correlation coefficients the output of scanned and processed array images in .csv formats (Section 2.2.8.3) were used for low-cell ChIP-chip and conventional ChIP-chip assays. The statistical correlation between the two datasets were analysed and the Spearman's rank correlation values between -1 and +1 were assigned, where -1 showed the highest negative correlation and +1 showed the highest positive correlation between two datasets. A p-value was also generated as a measurement for the significance of observed correlation (Dr. Peter Saffrey, Institute of Cancer Sciences, University of Glasgow).

The box plots were created by GraphPad Prism 5.0. The whiskers indicate 5-95 percentile. P-values were calculated by two-tailed unpaired t tests and 95% confidence intervals.

For z-score analysis of histone modification distributions at the promoters of annotated genes, the total number of sequencing tags at a 2 kb window centred at the TSS of each promoter was computed (x). For H3K4me3 analysis the promoters of H3K4me3 and bivalent/H3K27me3^{hi} genes were used for which a normal distribution of H3K4me3 levels could be established. Furthermore, for H3K27me3 analysis the promoters of bivalent/H3K27me3^{hi} genes were used for which a normal distribution of H3K27me3 levels could be established. The mean

and standard deviation of total number of sequencing tags for all the promoters that were included in the normal distribution was computed. The z-score for each promoter was obtained using the following formula in Microsoft Office Excel: z-score$_{promoter}$ = (x – mean) ∕ standard deviation

3. CML stem cells share the characteristics of both stem cells and proliferative cells at the gene expression level

3.1 Abstract

In order to understand the phenotype of CML stem cells (LSCs), studying their gene expression profile is essential to identify the genes and pathways that are involved in malignant transformation as well as the genes and pathways that maintain their stem cell-like phenotype. However, this aim could only be achieved by first understanding normal haemopoiesis and the pathways regulating HSC maintenance as well as the mechanisms involved in the differentiation of early haemopoietic progenitors. Therefore, in this Chapter, the genome-wide gene expression profiles of HSCs, LSCs, normal haemopoietic progenitors (HPCs,) and CML progenitor cells (LPCs) were determined using Affymetrix technology. 13 biological pathways were identified to be associated with the HSC identity, which included TGF-β signalling, Wnt signalling, chemokines and cytokines signalling, and several novel neurotransmitter signalling pathways. Furthermore, it was demonstrated that the HSC identity pathways retained their expression levels in LSCs, whereas several pathways associated with proliferative capacity were upregulated in LSCs. Therefore, LSCs shared similarities with both HSCs and HPCs at the level of biological pathways. Furthermore, the exit from stem cell compartment in CML was associated with the loss of HSC identity pathways and acquisition of further proliferative capacities. It was also revealed that a global programming regulate gene expression changes at the genome-wide level in both normal and CML cells. Consequently, this global programming controls gene expression changes at many biological pathways, as was also observed in the HSC identity pathways. However, LPCs were found to be more effective in downregulating the genes in the HSC identity pathways than their normal counterparts. Thus, it can be argued that the CML cells implement this global programming in a different manner from their normal counterparts.

3.2 Introduction

As described in Chapter 1, transcriptomics is the science of defining and quantifying all the various transcripts present in a cell, or population of cells,

including mRNAs, non-coding RNAs (ncRNAs) and small RNAs. Studying the complete set of transcripts in a cell and their levels is a crucial step in understanding molecular phenotype and functions of cells and tissues. The knowledge of transcriptomics is required for identifying the mechanisms involved in development and disease. The study of transcriptome has been made possible through development of various technologies which were described in Table 1.1. The work presented in this Chapter is generated by the Affymetrix microarrays which use a hybridisation-based approach. A brief summary of Affymetrix microarray platform and the downstream bioinformatics analyses that were used in this work is presented below.

Several studies have investigated the gene expression signatures of HSCs in human systems. These studies have helped identifying the mechanisms involved in HSC self-renewal and quiescence (Sections 1.6.4 and 1.6.5). In one study, the genome-wide gene expression analysis of human foetal HSCs (Lin⁻ CD34⁺ CD38⁻), murine foetal HSCs (Lin⁻ AA4.1⁺ c-Kit⁺ Sca-1⁺) and murine adult HSCs (Lin⁻ c-Kit⁺ Sca-1⁺) were analysed using Affymetrix arrays (Ivanova et al., 2002). The comparison of gene expression signatures between these three cell types identified evolutionarily conserved genes that may have a crucial role in HSC maintenance. Intriguingly, these HSC maintenance genes were shown to be predominantly associated with, signalling ligands and cell surface receptors, intracellular signalling, adhesion molecules and transcription factors. This signature included several important genes associated with TGF-β, Wnt and Notch signalling pathways, such as BMP8, frizzled 4 (FZD4), TCF3, and jagged 2 (JAG2), which could be important in the maintenance of HSCs (Section 1.6.4). Furthermore, the homeobox family of transcription factors, particularly HOXA5, HOXA10, HOXA2 and HOXB4, were also identified in this signature (Ivanova et al., 2002). HOXA10 and HOXB4 were reported to have an important role in HSC self-renewal (Magnusson et al., 2007, Sauvageau et al., 1995). On the contrary, the gene expression signature of early progenitors (Lin⁻ CD34⁺ CD38⁺ in human or Lin⁻ c-Kit⁺ Sca-1⁻ in mouse) was significantly enriched for cell cycle-associated genes, DNA repair mechanisms and protein synthesis and metabolism. The observation was in agreement with the proliferative capacities of these populations (Ivanova et al., 2002). The gene expression signatures of adult human HSCs were first described by isolating the Lin⁻ CD34⁺ CD38⁻ population from BM, PB, and cord blood using Affymetrix arrays (Georgantas et al., 2004). Genes that were

significantly upregulated in HSC populations relative to the differentiated haemopoietic cells (Lin$^+$ CD34$^+$ CD38$^+$) in all three tissues were described as HSC signature. The HSC signature was significantly enriched for signalling proteins as well as transcription factors, such as CEBPB, GATA3, HLF, HOXA3, HOXB6, KLF2, KLF4, RBPMS, and MECOM (MDS1 and EVI1 complex) (Georgantas et al., 2004). These two studies reported gene expression signatures of human HSCs for the first time and both highlighted the importance of signalling pathways and the HOX family of transcription factors in these signatures. The importance of signalling pathways in the maintenance of adult HSCs was also shown in mouse (Park et al., 2002, Ramalho-Santos et al., 2002).

A global analysis of gene expression signatures have been performed recently in human HSCs (CD34$^+$ CD38$^-$), lineage progenitors and terminally differentiated haemopoietic cells using Affymetrix arrays (Novershtern et al., 2011). This report has identified a subset of genes that are predominantly expressed in HSCs and early progenitors. This gene cluster comprises several transcription factors, such as GATA2, HOXA9, HOXA10 and MEIS1, as well as cell surface markers, such as CD34 and KIT. These genes were also expressed in early myeloid progenitors. However, their expression was suppressed after the branching of GMPs and MEPs and also in lymphoid branch (Novershtern et al., 2011). This HSC gene expression signature was used as a validation tool for the work presented in this Chapter.

There have been a number of studies investigating the gene expression of CP CML in recent years. Several reports have conducted their gene expression analyses in order to identify biomarkers for disease progression, response to TKIs and prognosis (McWeeney et al., 2010, Oehler et al., 2009, Radich et al., 2006). However, they did not address the differences between CML cells and their normal counterparts and more importantly the underlying gene expression signature that defines the LSC compartment. Three studies have attempted to answer these questions which are reviewed in this Section.

The first study investigated the gene expression signatures of CD34$^+$ cells isolated from the BM of CML patients and normal individuals using customised cDNA arrays covering 1,185 genes (Kronenwett et al., 2005). This study identified 158 genes that were significantly misregulated in CML relative to normal. This gene

signature showed an upregulation in the genes associated with proliferation and cell cycle progression, such as CDK4 and cell division cycle 25C (CDC25C), as well as downstream effectors of BCR-ABL1, such as RAS family and JAK2. On the contrary, the gene expression levels of two CDKIs, CDKN1A/p21 and CDKN2D/p19, were significantly downregulated in CD34$^+$ CML cells. Additionally, several chemokine- and cytokine-related genes were also significantly downregulated in CML cells, such as IL10, chemokine (C-X-C motif) receptor 4 (CXCR4), platelet-derived growth factor-α (PDGFA) and G-CSF receptor (CSF3R). This study also investigated the gene expression differences between the CML CD34$^+$ cells isolated from BM and PB, which identified no significant differential expression between the two populations. However, the CML cells isolated from either BM or PB were closely related to the normal CD34$^+$ cells that were isolated from BM at the gene expression level. Intriguingly, this report showed a significant upregulation of neurotransmitter receptors on CD34$^+$ CML cells, which include adenosine A1 receptor (ADORA1), opioid μ1 receptor (OPRM1) and chloride channel 1A (CLNS1A). This finding was combined with immunofluorescence analysis of various neurotransmitter receptors on the surface of CD34$^+$ CML cells. The results demonstrated an upregulation of GABA-B, serotonin 1F, corticotropin-releasing hormone (CRH) receptor, adenosine A1 and A2B receptors, opioid μ1 and κ1 receptors, and orexin receptor 1 and 2 (Kronenwett et al., 2005). The observation was in agreement with an earlier report demonstrating the expression of the above neurotransmitter receptors on the surface of HSCs (CD34$^+$ CD38$^-$) which were diminished in the CD34$^+$ CD38$^+$ progenitor cells and were completely absent in terminally differentiated cells (Steidl et al., 2004).

In another study, the gene expression signature of non-dividing (G_0) CD34$^+$ CML cells were reported using Affymetrix arrays (Graham et al., 2007). This study showed a small differential expression between the CML G_0 and dividing populations, whereas the normal CD34$^+$ cells in G_0 were significantly segregated from their dividing counterparts. This analysis also identified a large subset of significantly misregulated genes in CML G_0 relative to normal G_0. The significantly upregulated genes in CML G_0 relative to their normal counterparts were mainly associated with cell cycle progression and CDK activities, DNA replication and mitosis, whereas the significantly downregulated genes were mainly associated with chemokine activity and antigen presentation. These observations were in agreement with the above study. However, the differential expression analysis

between CML G_0 and dividing populations revealed a significant upregulation in chemokine activity in the quiescent compartment. This study identified several chemokine ligands, such as CCL19, CXCL3, CXCL12, CXCL13, which were significantly upregulated in the quiescent population of CML and suggested a role for the chemokines in the maintenance of LSCs.

Gene expression signatures of BM LSCs (CD34$^+$ CD38$^-$) cells were also established in comparison with three myeloid progenitors in CML, which include CMPs, GMPs and MEPs (Bruns et al., 2009). This study also indicated a significant degree of similarity between the gene expression signatures of LSCs and CMPs. Furthermore, the differential expression analysis between LSCs and HSCs demonstrated a significant downregulation of several cell adhesion proteins and chemokine receptors, such as CXCR4 and cadherin 2 (CDH2) and it was suggested that this downregulation could explain the inherent defective adhesion and migratory capacities of LSCs. Additionally, several stem cell self-renewal promoting transcription factors were also significantly downregulated in LSCs, such as PTEN and FLT3. Therefore, the report suggested a possible explanation for the reduced self-renewal and quiescence capacities of LSCs (Section 1.7.3).

Overall, the above studies indicated a proliferative-like gene expression signature for LSCs which was associated with elevated levels of cell cycle regulators and reduced levels of adhesion proteins and stem cell regulators relative to their normal counterparts. Additionally, an important role for the chemokine ligands and receptors was suggested in LSCs, although their levels were significantly reduced relative to their normal counterparts. Intriguingly, one study suggested a role for neurotransmitter regulation of LSCs similar to HSCs (reviewed in Section 1.6.4.7). However, all these studies only reported a small subset of genes and their association with biological processes which tend to be broad descriptions. They did not investigate the interaction between the misregulated genes in the context of biological pathways which provide a better understanding of CML biology. Moreover, the gene expression signatures were reported without considering the role epigenetic regulation that was shown to be responsible for lineage priming in haemopoietic development (Section 1.6.7). Therefore, the work presented in this Chapter is going to address the impact of misregulated genes in LSCs in the context of biological pathways. Subsequently, the role of chromatin modifications in transcriptional regulation of LSCs will be investigated in Chapter 5.

3.3 Methods

3.3.1 The Affymetrix GeneChip expression arrays

Typically, fluorescently-labelled complementary DNAs (cDNAs) are hybridised onto customised or commercial high-density oligonucleotide microarrays. Genomic tiling microarrays, described in Chapter 4, with whole genome coverage at high density were initially used for high resolution mapping (ranging from several base pairs to ~100 bp) of transcribed regions (Bertone et al., 2004). However, the emergence of commercial platforms such as Affymetrix GeneChips revolutionised this technology by producing high resolution arrays with a genome-wide coverage. Affymetrix GeneChip microarrays were first constructed by adapting semiconductor technology using quartz wafer and the photolithographic process (McGall and Christians, 2002). The Affymetrix GeneChip expression array system is a one-colour microarray system. In a one-colour array, control and experimental samples are hybridised onto different arrays, labelled with the same fluorescent dye. The one-colour design provides flexibility when comparing multiple different samples from the same platform (Russell et al., 2009). The GeneChip arrays are mainly classified into Whole-Genome 3′ *in vitro* transcription (IVT) arrays and Whole-Transcript (WT) Exon/Gene arrays (Figure 3.1).
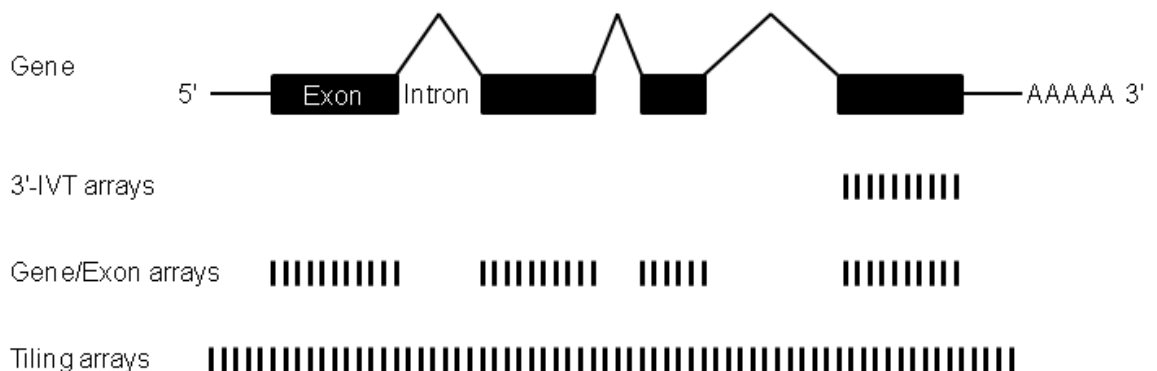
Figure 3.1: Comparison between the design of Affymetrix arrays and genomic tiling arrays. The spread of the oligonucleotide probes (black bars) across the gene body is demonstrated for different types of arrays used for gene expression analysis.

The 3′ IVT arrays were the first generation of commercial arrays using probes designed against the 3′-end of each gene to estimate the expression of the entire gene (Dalma-Weiszhausz et al., 2006). This technology required 3′-oligo(dT) cDNA synthesis and labelling procedure. However, transcripts with unknown 3′-

ends, which were non-polyadenylated, had alternative polyadenylation sites, or were truncated, could not be analysed by this approach. Therefore, due to the above limitations and the presence of a large number of splice variants, the next generation of Affymetrix arrays - the WT Gene/Exon arrays - were designed which allowed the analysis of alternative splicing as well as transcripts with undefined 3′-ends, truncations, or degradation products (Dalma-Weiszhausz et al., 2006). WT arrays have replaced oligo(dT) cDNA synthesis used for 3′ IVT arrays with a random priming approach. To perform hybridisation, total RNA or mRNA extracted from the cell or tissues of interest is first reverse-transcribed using a T7-promoter random primer to generate double-stranded cDNA. The cDNA then undergoes an *in vitro* transcription reaction in the presence of T7 RNA polymerase and biotinylated ribonucleotides to generate biotin-labelled complementary RNAs (cRNAs). The biotinylated cRNAs are fragmented (to optimise target-probe hybridisation kinetics) and hybridised onto the probe array. The hybridised probe array is stained with a streptavidin phycoerythrin (PE) conjugate and scanned. The PE conjugate is excited by laser and emits fluorescence for detection (Dalma-Weiszhausz et al., 2006) (Figure 3.2).

Human WT Gene arrays were specifically used for the work presented in this Chapter. Human WT Gene arrays cover 28,869 annotated genes (University of California Santa Cruz (UCSC) hg18 and National Centre for Biotechnology Information (NCBI) build 36 annotations). The oligonucleotide probes used in WT Gene array design are perfect match (PM) probes with a complete complementarity to the target sequence and the background noise is analysed using a set of 20,000 generic background probes. Each gene is represented by an average of 26 25-mer probes positioned across the exons which provides a more accurate and complete profiling of gene expression than the 3′ IVT arrays (Russell et al., 2009).
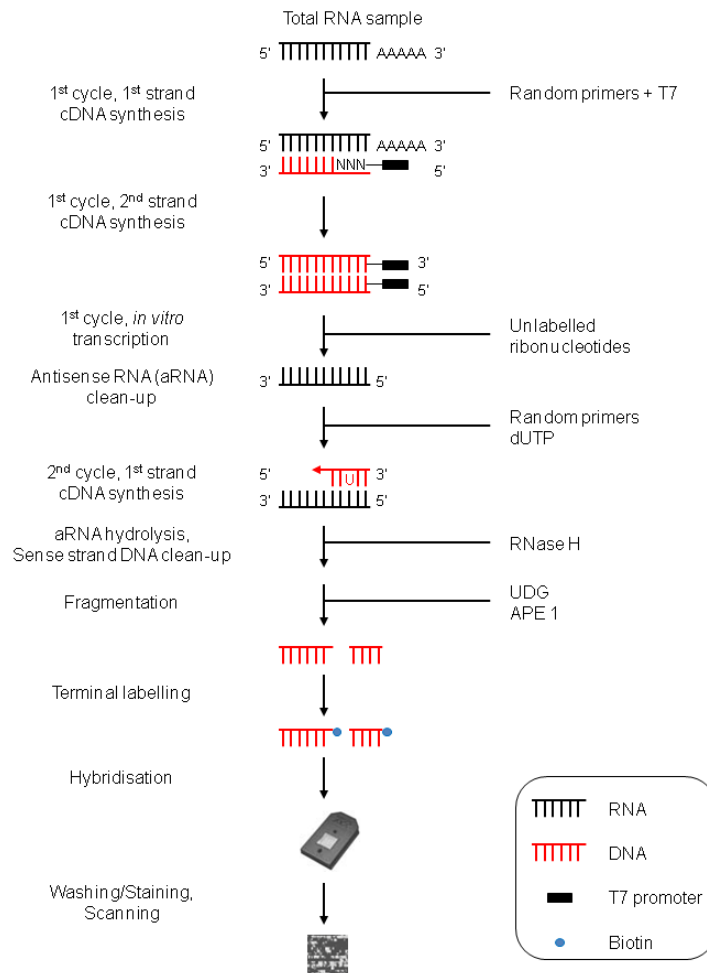
Figure 3.2: Schematic overview of cDNA synthesis for Affymetrix WT Gene/Exon arrays.
The overall work-flow of RNA-to-cDNA preparation for the Affymetrix WT arrays is illustrated. cDNA synthesis is performed in two cycles initiated by random priming. This process results in generation of biotinylated sense DNA targets that are required for hybridisation on Affymetrix arrays. T7: T7 RNA polymerase; UDG: uracil-DNA glycosylase; APE1: apurinic/apyrimidinic endonuclease.

Despite the high-throughput advantage of the hybridisation-based approaches, they suffer from several limitations. These include dependency on the existing knowledge of genomic sequence and annotation, cross-hybridisation which could result in high background, and limited dynamic range of detection as dictated by the signal saturation and background (Wang et al., 2009). Furthermore, comparison between experiments requires complicated normalisation approaches and experiments performed on different array platforms could not be directly compared owing to differences in the array or probe design (Wang et al., 2009). These limitations resulted in the development of sequence based approaches (Table 1.1), such as RNA-sequencing (RNA-seq), which offer a high-throughput and unbiased analysis of all types of transcripts. Nevertheless, these approaches require large amounts of RNA which cannot be obtained from small adult stem cell

populations. Therefore, these approaches could not be employed for the analysis of the gene expression signatures in HSCs and LSCs in this work.

3.3.2 Affymetrix data analysis

3.3.2.1 Data pre-processing

Data pre-processing is a crucial step prior to downstream statistical analysis or data clustering. Pre-processing largely consists of various data normalisation steps to adjust for the systematic errors as a result of experimental variation as opposed to real biological differences (Russell et al., 2009). The major steps of data normalisation are: (i) probe summarisation which merges intensities from multiple probes that correspond to a gene to generate a single value for each gene; (ii) background correction which is a 'within-array' normalisation procedure to account for background levels; (iii) normalisation which is a 'between-array' normalisation that accounts for intensity variations between multiple arrays.

A. Probe summarisation – As mentioned earlier, the WT Gene arrays are designed based on PM probes and a probe set is defined as a set of probes for a given gene. Robust Multi-array Averages (RMA) and GeneChip RMA (GCRMA) are the two methods that have been introduced for the probe summarisation of PM probe sets (Irizarry et al., 2003). In RMA approach, the observed signal is regarded as a measure of true expression in addition to a probe-specific affinity factor. The output is comprised of log transformed intensities which is fitted to Tukey's median polish model (Irizarry et al., 2003).

B. Background correction – Background noise is a common artefact of any microarray technology and Affymetrix arrays are not an exception. The background level should be determined and corrected prior to any further evaluation of the signals. The noise in Affymetrix arrays could be caused by the scanning device, non-specific binding of labelled DNA to the surface of arrays, artefacts of labelling, or other imperfections in array design. The background correction should be accurate as otherwise true signals might be mistakenly removed in the process (Russell et al., 2009). As mentioned earlier, a set of 20,000 generic background probes are designed on WT Gene arrays. These generic probes are used in a global background correction method performed in RMA algorithm (Russell et al., 2009).

C. Normalisation – 'Between-array' normalisation is also important to maintain the ability to compare results generated on different arrays. The RMA method uses a quantile normalisation approach which assumes the same amount of RNA is hybridised to identical arrays and therefore, expects an identical overall intensity distribution. In other words, RMA normalisation works by making the intensity distribution of multiple arrays identical (Irizarry et al., 2003).

3.3.2.2 Differential expression analysis

A powerful differential expression analysis requires a powerful experimental design. In other words, microarray experiments should be designed to contain an appropriate number of biological replicates in a cost efficient manner. Previous studies using microarrays for gene expression analysis suggested an experimental design which contains a minimum, but not an optimum, of five biological replicates per condition group providing that only two groups of conditions are to be compared (Allison et al., 2006, Pavlidis et al., 2003). Biological replicates would allow evaluation of biological differences between two groups as well as measuring the variability within each group. However, technical replicates can be substituted if the number of samples is limited or the cost of obtaining another sample is more than the cost of an array (Allison et al., 2006).

Another important step in the analysis of differential expression is the hypothesis test required for making statistical decisions on whether a gene is differentially expressed. Four types of possible outcomes are predicted for a pair of hypotheses, two of which are regarded as correct decisions: when a differential gene expression value is not significant (p-value above the threshold) and there are no gene expression differences, and when the difference is statistically significant (p-value below the threshold) and there is a change of expression. The other two outcomes are known as 'inferential errors': type 1 errors or false positives, when p-value is significant but there is no change of gene expression, and type 2 errors or false negatives, when the p-value is not significant but there is a change of expression. Power of the test is indicated by the probability of not making false negative errors, whereas the confidence is defined by the probability of not obtaining false positive errors (Russell et al., 2009, Tsai et al., 2003).

Fold change can be used as a sole measure of gene expression differences; however, it is deemed statistically inadequate as the variance and confidence levels cannot be measured. Therefore, using fold change cutoffs regardless of the variance or sample size, results in increased false positive rates (Allison et al., 2006). In order to tackle this issue statistical tests are performed on many transcripts to verify the candidates that show expression differences based on the p-value (Allison et al, 2005; Russell et al, 2009). False discovery rate (FDR) control methods, such as Benjamini-Hochberg (BH) correction (Benjamini and Hochberg, 1995), are common statistical tests in microarray analysis.

3.3.2.3 Principal component analysis (PCA)

PCA is a class of unsupervised methods which do not require prior knowledge of potential gene expression differences. PCA algorithm converts a set of potentially correlated variables, such as gene expression, into a set of values of linearly uncorrelated types known as principal components (PCs). In the gene expression model, PCs are employed for visualisation purposes by reducing the dimensionality of expression data and filtering noise. PCA analysis results in the identification of gene or sample clusters with similar expression patterns (Russell et al., 2009).

3.3.2.4 Data mining

In order to convert the differences in gene expression into meaningful biological clusters, different tools have been developed to divide differentially expressed genes based on their biological function, molecular characteristics, or cellular location amongst others. These variants are referred to as the GO terms. PANTHER database (www.pantherdb.org) (Thomas et al., 2003) provides integrated GO analysis tools, which are used for biological pathways analyses in this Chapter.

3.3.2.5 Validation of microarray data

FDR control is not very stringent in detecting false positive errors to allow the discovery of gene expression differences, particularly in the absence of large numbers of biological replicates. These false positive signals could be due to the cross-hybridisation or hybridisation artefacts. Furthermore, the limited intensity

range of the spots on the array does not reflect true fold change differences, but as a factor of normalised intensity relative to the other spots and arrays in the experiment. Therefore, it has been recommended to use RT-PCR for validation of the observed gene expression differences. RT-PCR provides a better measure of quantitative differences in gene expression levels between two experimental conditions or samples. Furthermore, the false positive signals can be identified by designing gene-specific RT-PCR primers (Rajeevan et al., 2001, Rockett and Hellmann, 2004).

The microarray data can also be compared with information available in literature. An agreement between the microarray analysis and data from other sources provides confidence in the identified biological processes. Taking the analysis in this Chapter as an example, gene expression analysis of HSCs and CML cells have been previously reported (reviewed below). If there is an overlap between the observed gene expression signatures in this Chapter and the previous reports, it could be suggested that the observed results are likely to be meaningful.

3.4 Aims of the Chapter

The gene expression patterns of CML cells versus their normal counterparts will be investigated.

1. To purify normal and CML cell populations based on the CD34 and CD38 cell surface markers by FACS.

2. To derive gene expression profiles of HSCs (CD34$^+$ CD38$^-$), LSCs (CD34$^+$ CD38$^-$), HPCs (CD34$^+$ CD38$^+$) and LPCs (CD34$^+$ CD38$^+$).

3. To identify the genes and pathways involved in the maintenance of HSCs.

4. To identify the genes and pathways involved in the proliferation of HPCs.

5. To identify the genes and pathways, which are uniquely affected in CML.

6. To determine whether the pathways identified in 4, 5 and 6 change in their levels of expression as LSCs become LPCs.

Results

3.5 Overall strategy

HSCs (CD34$^+$ CD38$^-$), LSCs (CD34$^+$ CD38$^-$), HPCs (CD34$^+$ CD38$^+$) and LPCs (CD34$^+$ CD38$^+$) were isolated by FACS from non-CML (Ph$^-$) (n=3) and chronic phase CML (Ph$^+$) (n=3) samples. Total RNA was extracted followed by cDNA synthesis, labelling and hybridisation on Affymetrix WT Gene arrays. After pre-processing the raw image data, significant gene expression differences between each cell type were deduced. Gene expression profiles were used in PCA analysis to cluster the datasets according to the similarities and differences between different cell types. Differential expression was used to identify the significant biological pathways associated with each transition using the PANTHER database. Putative pathways involved in maintaining HSC quiescent and self-renewal (HSC identity) or those involved in proliferation (HPC identity) were identified by comparing the gene expression profiles of HSCs to those of HPCs. Furthermore, CML identity pathways were identified as the pathways which showed expression changes uniquely in in the CML cells. The expression levels of HSC identity and HPC identity pathways were analysed in LSCs and LPCs to establish the links between the stem cell phenotype of LSCs and the pathways which control it, as well as the changes in these pathways that are required for LSCs to become proliferative progenitors.

3.6 Isolation of HSCs, LSCs, HPCs, and LPCs

HSCs and HPCs were purified from G-CSF-mobilised PB samples of a healthy donor and two non-Hodgkin lymphoma patients. Non-Hodgkin lymphoma is a haematological malignancy that affects lymphoid branch of haemopoiesis and abnormalities have not been reported for their HSC compartment. Therefore, these samples were referred to as 'non-CML' in this work. LSCs and LPCs were purified from the PB samples of three CP CML patients. The CD34$^+$ populations from these patients were assessed for the presence of Ph chromosome by FISH technique at the time of sample collection (Table 3.1; Dr. Mhairi Copland, unpublished data). All CML samples were >96% Ph$^+$.

HSCs and LSCs were defined as cell populations expressing CD34 cell surface antigen in the absence of CD38 antigen and therefore, referred to as CD34$^+$CD38$^-$ population, whereas HPCs and LPCs were identified by expression of both CD34 and CD38 molecules on their cell surface and therefore, referred to as CD34$^+$ CD38$^+$ population. The CD34$^+$ CD38$^-$ population was previously reported to be ~100 folds more enriched for engraftable HSCs than the CD34$^+$ CD38$^+$ population (Bhatia et al., 1997, Larochelle et al., 1996). The Ph$^+$ LTC-IC CML cells were also significantly associated with the CD34$^+$ CD38$^-$ population (Petzer et al., 1996). However, as discussed in Section 1.7.3, cytogenetic analysis was required for further characterisation of sorted LSCs and LPCs. (Table 3.1; Dr. Mhairi Copland, unpublished data).

The isolation of CD34$^+$ CD38$^-$ and CD34$^+$ CD38$^+$ subpopulations was achieved by FACS using antibodies against CD34 and CD38 antigens (Figure 3.3). The proportion of stem cell compartment varied between samples, although non-CML individuals showed significantly larger stem compartment than the CML samples. This could be associated with the G-CSF mobilisation that released HSCs from their BM niche. FISH analysis for the Ph chromosome in the sorted LSCs showed that >89% of the cells were Ph$^+$ (Table 3.1, Dr. Mhairi Copland, unpublished data).

Label		Cell surface markers	Sample name	FISH Positive (%)	
				CD34$^+$	CD34$^+$CD38$^-$
CML1	LSC1	CD34$^+$CD38$^-$	CP CML 264	99	98
	LPC1	CD34$^+$CD38$^+$			
CML2	LSC2	CD34$^+$CD38$^-$	CP CML107	96.2	Data not available*
	LPC2	CD34$^+$CD38$^+$			
CML3	LSC3	CD34$^+$CD38$^-$	CP CML110	99.7	89
	LPC3	CD34$^+$CD38$^+$			
Non-CML1	HSC1	CD34$^+$CD38$^-$	Normal 400	Not tested.	
	HPC1	CD34$^+$CD38$^+$			
Non-CML2	HSC2	CD34$^+$CD38$^-$	Non-CML 015		
	HPC2	CD34$^+$CD38$^+$			
Non-CML3	HSC3	CD34$^+$CD38$^-$	Non-CML 012		
	HPC3	CD34$^+$CD38$^+$			

Table 3.1: Molecular characteristics of the HSCs, LSCs, HPCs, and LPCs used in this study. The phenotype and the percentage of Ph+ cells measured by FISH are reported for the samples used in this study. The percentage of Ph$^+$ cells reported for the bulk CD34$^+$ compartment was obtained at the time of collection from newly diagnosed patient. The percentage of Ph$^+$ reported for CD34$^+$ CD38$^-$ population was obtained after cell sorting. Data was not available for the post-sort LSC2 sample, although the CML2 CD34$^+$ population was significantly Ph$^+$ (Dr. Mhairi Copland, unpublished data).
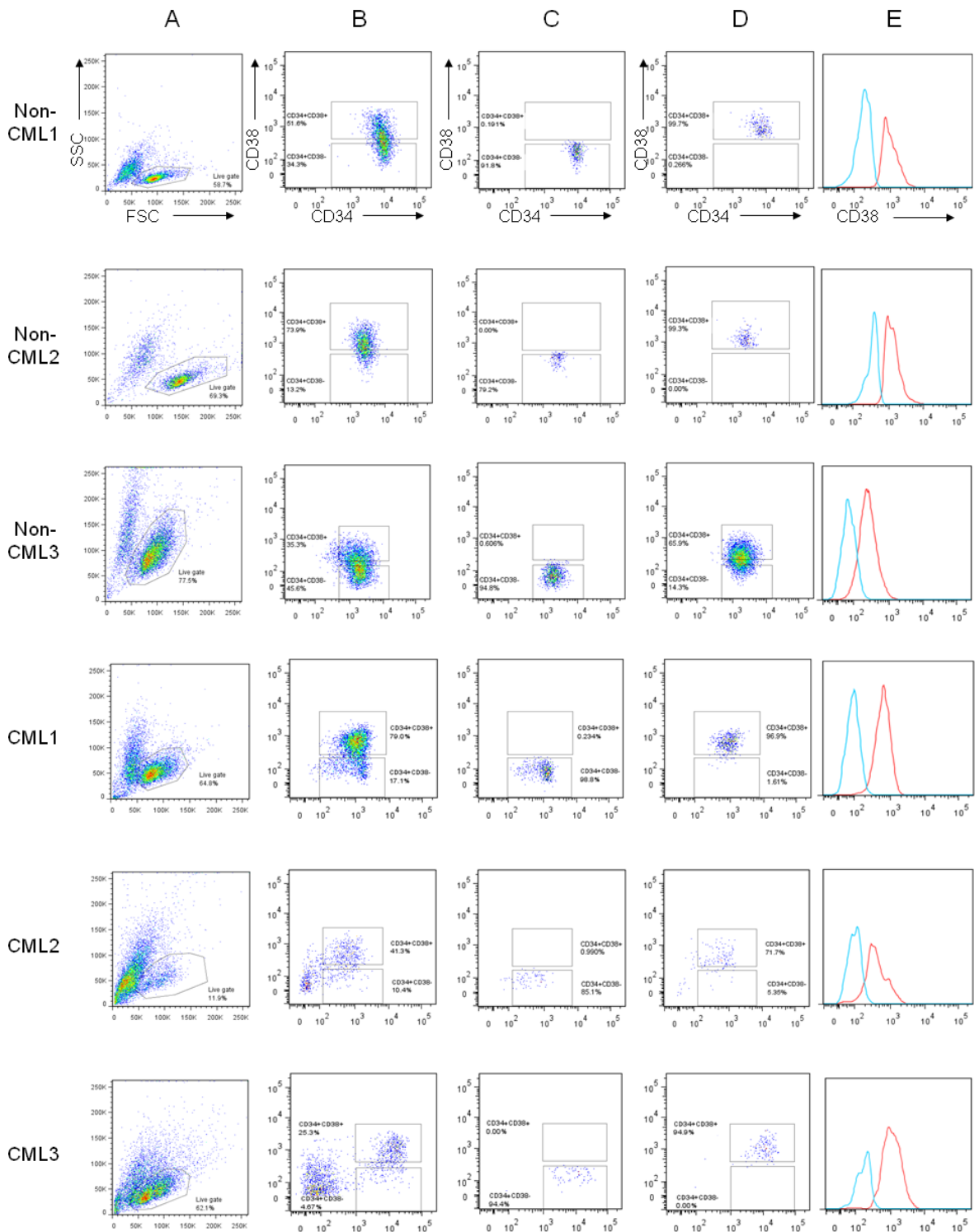
Figure 3.3: Isolation of HSCs, LSCs, HPCs, and LPCs by FACS-sorting.
Dot plots represent FACS profiles of non-CML (n=3) and CML (n=3) samples. (A, B) Cell sorting was performed on the CD34[+]-enriched in the three non-CML and CML1 samples, whereas unenriched mononuclear cells were used for CML2 and CML2 samples. (A) The live cell population was identified by using the side scatter (SSC) and forward scatter (FSC) parameters. The percentage of live cells is shown. (B) The CD34 and CD38 gates were adjusted based on the isotype controls. The percentage of live cells in CD34[+] CD38[-] and CD34[+] CD38[+] populations are shown. (C) The HSCs and LSCs were isolated by the presence of CD34 (x-axis) and the absence of CD38 (y-axis) cell surface antigens. (D) The HPCs and LPCs were isolated by the presence of both CD34 and CD38 markers. The post-sort plots demonstrate the purity of samples in each case as represented by the percentage of cells in each gate. (E) The distinction between cell surface CD38 expression in isolated CD34[+] CD38[-] (blue line) and CD34[+] CD38[+] (red line) cells is shown as histograms representing the CD38 intensity on the x-axis with the cell count on the y-axis.

3.7 RNA extraction for Affymetrix hybridisations

Total RNA was extracted from the isolated HSCs, LSCs, HPCs, and LPCs, as described in Chapter 2. For HSCs and LSCs, two technical replicates per sample were obtained by FACS purification of two separate aliquots of each sample at independent time points. In total, six RNA samples were generated for each of HSC and LSC cell types. For HPCs and LPCs, three RNA samples were purified for each individual and were subsequently divided into two technical replicates. The quality and integrity of purified DNase-treated RNA samples were investigated using the Agilent Bioanalyzer prior to cDNA synthesis for Affymetrix hybridisation (Figure 3.4). RNA integrity was calculated by the RNA integrity number (RIN) which assigns a value between 1 and 10. RIN of 1 indicates complete degradation and 10 complete integrity of RNA samples. The RIN values of 8 or above are recommended for the Affymetrix hybridisations.
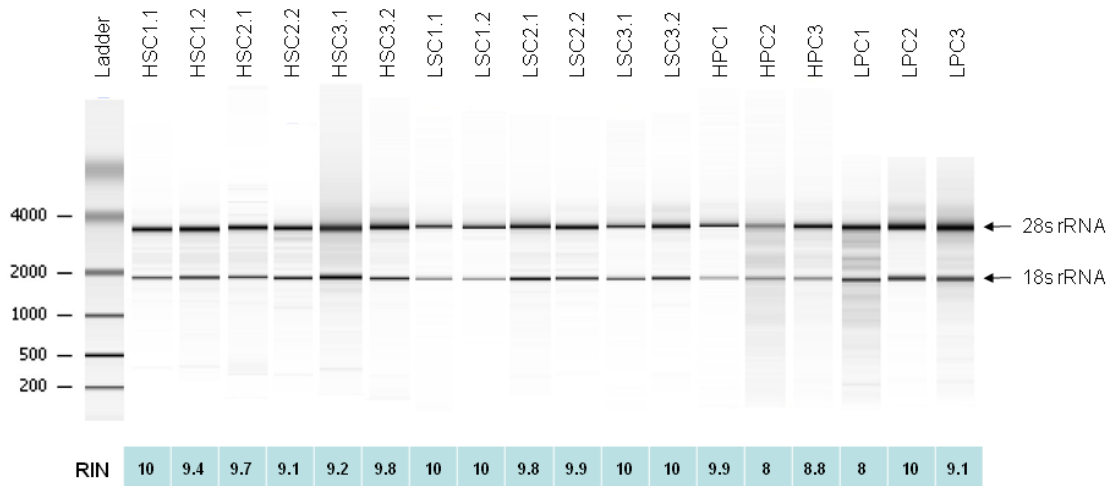
Figure 3.4: Bioanalyzer quality control of purified total RNA.
The Bioanalyzer gel results for the total RNA samples purified from HSC, LSC, HPC, and LPC samples. Two major bands are detectable in each lane which belong to 28s (top) and 18s rRNAs (bottom). Moreover, no significant genomic DNA or protein contamination was detected in any of the samples. Ladder and the bp ruler can be seen on the left side. The integrity of RNA samples were measured by the RIN tool. RIN values of above 8.0 are considered as acceptable for Affymetrix hybridisation.

The Bioanalyzer gel analysis showed no major contamination of genomic DNA or proteins and only two bands could be significantly detected for all samples which belonged to 28s and 18s rRNAs (Figure 3.4). Furthermore, all RNA samples showed an RIN of 8 or above and therefore, passed the required integrity threshold for Affymetrix hybridisation.

3.8 Affymetrix hybridisation and pre-processing of data

cDNA synthesis and labelling, using the random priming approach described earlier, was performed and the samples were hybridised onto Affymetrix human GeneChip WT Gene arrays (HuGene-1_0-st-v1). The raw image data (.CEL files) were obtained and the pre-processing of data was achieved by RMA probe summarisation and background correction, and quantile normalisation in the *aroma.affymetrix* package of Bioconductor, as described in Chapter 2. RNA degradation, Normalised Unscaled Standard Error (NUSE), and Relative Log Expression (RLE plots) were generated to check the quality of the hybridisation and to identify potential outlier datasets (Figure 3.5).
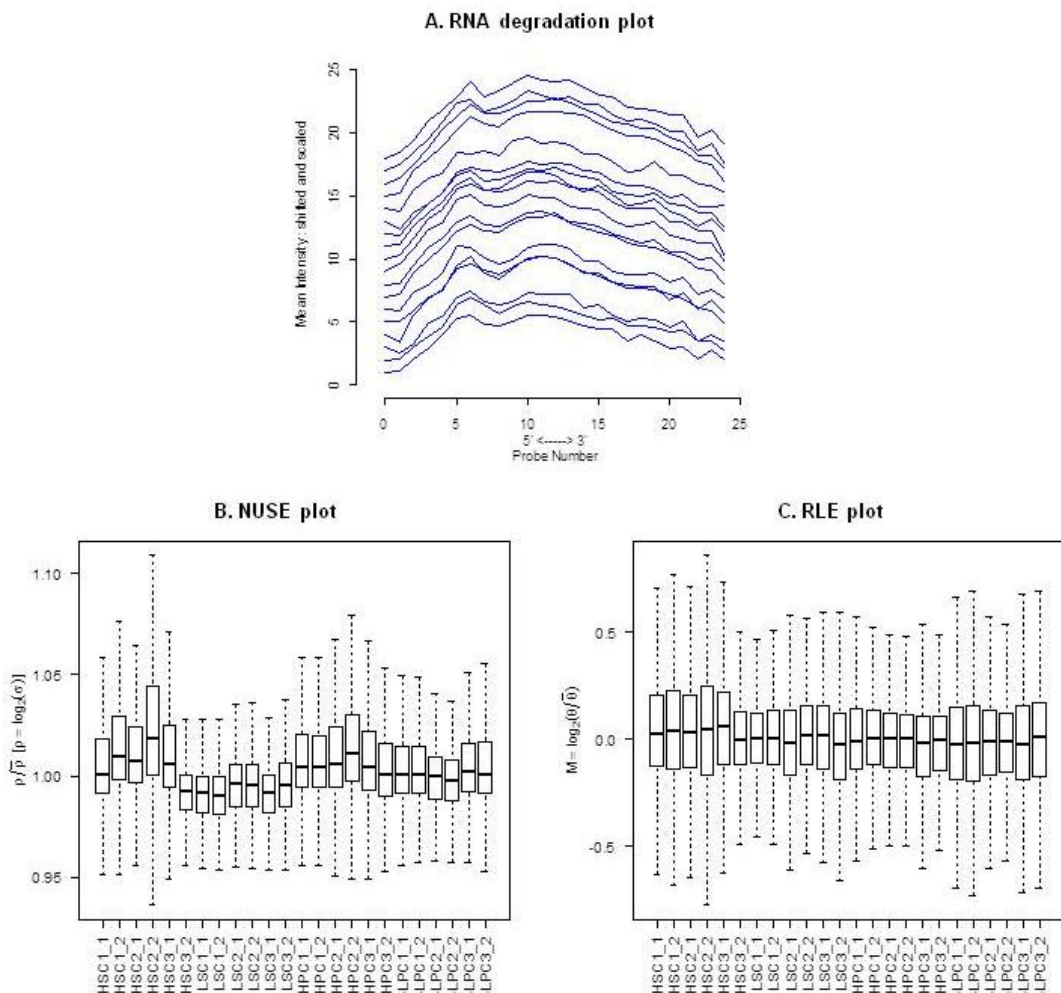
Figure 3.5: Affymetrix pre-processing quality control (QC) plots.
(A) RNA degradation plots which show the mean intensity of probes (y-axis) as laid down from 5′ to 3′ in an average transcript (x-axis). The RNA degradation plots demonstrate higher intensity of probes at the 5′ end with a gradual decline towards the 3′ end. Furthermore, the slope of the plots are in agreement. (B) NUSE plot in which the medians of box plots are expected to align close to 1.00 (y-axis). (C) RLE plot in which the medians of box plots are expected to align at M=0 (y-axis).

RNA degradation plot is a visualisation tool to inspect the quality of hybridisation across the probes in an average gene as laid down from 5′ to 3′. The RNA is degraded from the 5′-end and therefore, intact RNAs at the time of hybridisation show higher probe intensity at the 5′-end with a gradual decline towards the 3′-end. Furthermore, the arrays that show a different slope relative to others are normally regarded as outlier. In this case, all the arrays showed a similar slope in RNA degradation plots with higher intensities at 5′-end and lower intensities towards the 3′-end (Figure 3.5A). Therefore, the expression differences between the samples would not be due to significant quality variations of the hybridised material. NUSE and RLE plots were used as the next line of QC after pre-processing. NUSE is calculated by standardising the standard error of the estimated expression value for each probe set in order to obtain a median standard error for each array which is close to 1.00. The median of standard error for each array in this study was observed to be close to 1.00 (Figure 3.5B). RLE plots are generated by calculating log-scale estimates for the expression of each probe set on each array. Furthermore, it is presumed that the number of differentially expressed genes between datasets are relatively few and therefore, the box plot representing each array is expected to centre close to 0. All the hybridised arrays in this work were shown to be aligned at 0 (Figure 3.5C). Therefore, the hybridised arrays passed the QC tests and can be used for the downstream PCA and differential expression analyses.

3.9 PCA clustering of the gene expression datasets

In order to demonstrate the relationship between the gene expression profiles of CML cells and their normal counterparts, three-dimensional (3D) PCA plots were generated from the preprocessed microarray datasets (Figure 3.6). PCA method is typically employed for visualisation purposes by reducing the dimensionality of expression data and filtering noise. Unlike hierarchical clustering, the clustering decisions are not made locally in the PCA approach and the relationship between different data sets can be re-evaluated at a later stage in a 3D environment (Russell et al., 2009).
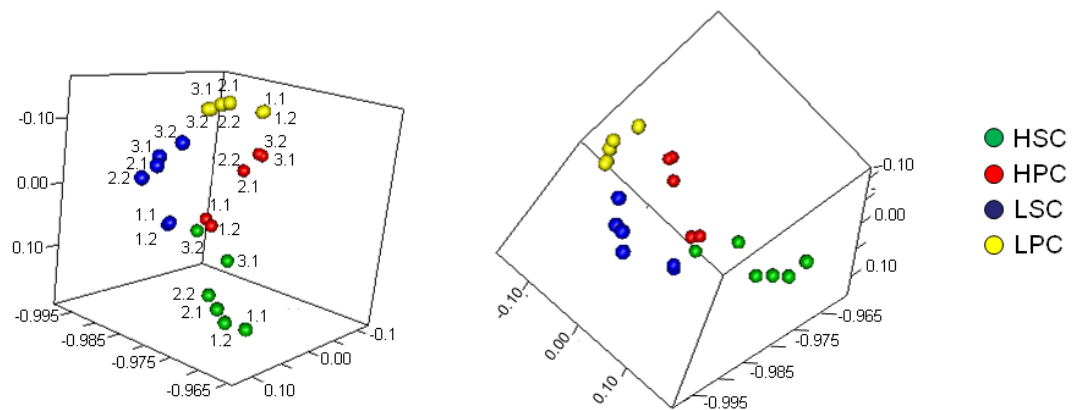
Figure 3.6: PCA clustering of the gene expression datasets.
3D PCA clustering of the preprocessed gene expression datasets is shown from two different angles. The relationship between the gene expression signature of the four cell can be observed. HSCs (green circles), HPCs (red circles), LSCs (blue circles), and LPCs (yellow circles) are labelled with regards to their sample number (Table 3.1).

The 3D PCA clustering demonstrated that there is a considerable heterogeneity in the LSC population (Figure 3.6, blue circles). The LSC1 sample was found in close proximity to their normal counterparts (Figure 3.6, green circles), whereas LSC2 and LSC3 samples were in the close proximity of LPCs (Figure 3.6, yellow circles). Intriguingly, the distance between LSCs and HSCs in the 3D environment was similar to the distance between HSCs and HPCs (Figure 3.6, red circles), although LSCs and HPCs did not overlap. These observations indicated that the gene expression signature of LSCs was located between the gene expression profiles of HSCs and LPCs. Therefore, LSCs could be expected to demonstrate the intermediate characteristics of both cell type, or in other words, less primitive but more proliferative than their normal counterparts. The observations were also in agreement with the reported similarities between the gene expression signature of LSCs and early myeloid progenitors (Bruns et al, 2009; reviewed in Section 3.1.4).

The LPCs, on the other hand, were more homogenous and discrete than LSCs. Furthermore, LPCs were observed in the proximity of their normal counterparts, HPCs. This observation was in agreement with the proliferative capacities of both cell types. However, LPCs and HPCs did not overlap which indicated distinct gene expression profiles between the two progenitor cell types. Furthermore, LPCs and HPCs were the two most completely segregated cell populations in the 3D PCA plot. Therefore, it could be expected that LPCs and HSCs demonstrate the largest number of gene expression differences amongst all the possible comparisons.

These global observations at the level of gene expression profiles showed interesting characteristics for LSCs and LPCs relative to their normal counterparts. However, the relationship between the gene expression signatures of CML cells and their normal counterparts needs further investigation by analysing differential expression and the biological pathways that were affected by the changes of gene expression between cell types.

3.10 Analysis of differential gene expression

Four types of differential gene expression analysis were performed. First comparison was performed between HPCs and HSCs in order to identify the genes that are associated with the HSC identity and the proliferative HPC identity. Second comparison was performed between LSCs and HSCs to distinguish the similarities that LSCs share with the HSC and HPC identities. This comparison was also required in order to identify the unique gene expression changes that occur during LSC formation which define the CML identity. Third comparison was performed between LPCs and LSCs in order to investigate gene expression differences that promoted exit from the stem cell compartment which define LPCs characteristics. Fourth comparison was performed between LPCs and HSCs in order to establish the overall gene expression differences that occurred in the course of LPC development from HSCs. The fourth comparison can also be used to establish the similarities and differences that the formation of LPCs from HSCs share with normal differentiation to HPCs.

Statistically significant gene expression differences in each comparison were deduced using the *limma* package in Bioconductor with BH FDR control adjusted at 0.05 confidence threshold, as described in Chapter 2. The hg18 RefSeq annotation was used to annotate the significantly differentially expressed Affymetrix probe sets. The total number of significantly upregulated or downregulated genes in each of the above comparisons is shown in Table 3.2.

	HPCs vs HSCs	LSCs vs HSCs	LPCs vs LSCs	LPCs vs HSCs
↑	2,612	1,996	2,366	4,948
↓	1,877	1,632	855	4,107
Total	4,489	3,628	3,221	9,055

Table 3.2: Total number of significant gene expression differences in the four comparisons. The total number of significantly (p<0.05) upregulated (red) and downregulated (green) genes is indicated for each of the studied comparisons.

The first comparison which was performed between HPCs and HSCs identified 4,489 significantly differentially expressed genes (p<0.05). 60% of the gene expression changes (n=2,612) in this comparison were in favour of upregulation which could be potentially associated with the proliferative characteristics of HPCs. Furthermore, the significantly downregulated genes (n=1,877) could be potential regulators of HSC identity or the genes whose expression were no longer required in HPCs. The second comparison which was performed between LSCs and HSCs identified 3,628 significantly differentially expressed genes, 55% of which (n=1,996) were associated with significant upregulation and 45% of them were significantly downregulated in LSCs (n=1,632). The third comparison was performed between LPCs and LSCs which identified 3,221 significantly differentially expressed genes. 73% of the gene expression differences (n=2,336) in this comparison was linked to upregulation, whereas only 27% of changes were in a downward direction indicating a large tendency for gene upregulation to promote exit from the stem cell compartment. The fourth comparison which was performed between LPCs and HSCs identified 9,055 significantly differentially expressed genes. 55% of the changes (n=4,948) were linked to upregulation and 45% of the changes (n=4,107) were associated with downregulation. This observation indicated that although the gene expression changes during LPC formation from LSCs are largely skewed towards upregulation, the gene expression differences between LPCs and HSCs were not skewed and resembled that pattern that was observed between HPCs and HSCs. The comparison between LPCs and HSCs identified the largest subset of gene expression differences amongst the four comparisons which was in agreement with earlier observations of 3D PCA clustering (Figure 3.6). However, the sum of total number

of changes between LSCs and HSCs as well as LPCs and LSCs was 6,849 which only accounted for 75% of the overall changes between HSCs and LPCs. The results could propose that a subset of genes accumulated non-significant gene expression changes in the HSC-to-LSC and LSC-to-LPC transitions which became significant gene expression differences in the overall comparison between HSCs and LPCs.

The above results demonstrated significant number of changes in both upward and downward directions in each of the studied comparisons. However, the relationship between the significant expression changes in one comparison and the gene expression changes in other comparisons was not revealed. In order to address that the expression changes (both significant and non-significant) of all the annotated genes in all four comparisons were plotted as heatmaps (Figure 3.7Ai). Furthermore, heaptmaps were also generated for the genes that were only significantly differentially expressed in each of the comparisons (Figure 3.7Aii).

The heatmap of all the expression changes which included both significant and non-significant changes in the four comparisons revealed that the direction of gene expression changes in different comparisons was concordant (Figure 3.7Ai). However, the magnitude of gene expression change and their significance level in different comparisons were not the same. In other words the genes that were downregulated in HPCs relative to HSCs were also predominantly associated with downregulation in other comparisons. Similar observation was also made for the upregulation events. Furthermore, the heatmap that was generated for the significant gene expression changes, which were a subset of the genes from the previous heatmap, showed a concordance in the direction of significant gene expression changes in other comparisons if the expression change was also significant (Figure 3.7Aii). Both heatmaps provided evidence for the earlier suggestion that the accumulation of non-significant gene expression changes during HSC-to-LSC and LSC-to-LPC transitions could result in significant gene expression changes in the overall comparison between HSCs and LPCs. Furthermore, the heatmap for significant gene expression changes (Figure 3.7Aii) also demonstrated a concordance in the direction of gene expression changes between HSC-to-HPC and HSC-to-LPC transitions. This could propose similar changes in the gene expression signature of HSCs in the course of differentiation to HPCs and leukaemic transformation to LPCs. This was further investigated by

analysing the overlap between the significant gene expression differences in HSC-to-HPC transition and the other three comparisons involving LSCs and LPCs (Figure 3.7B).
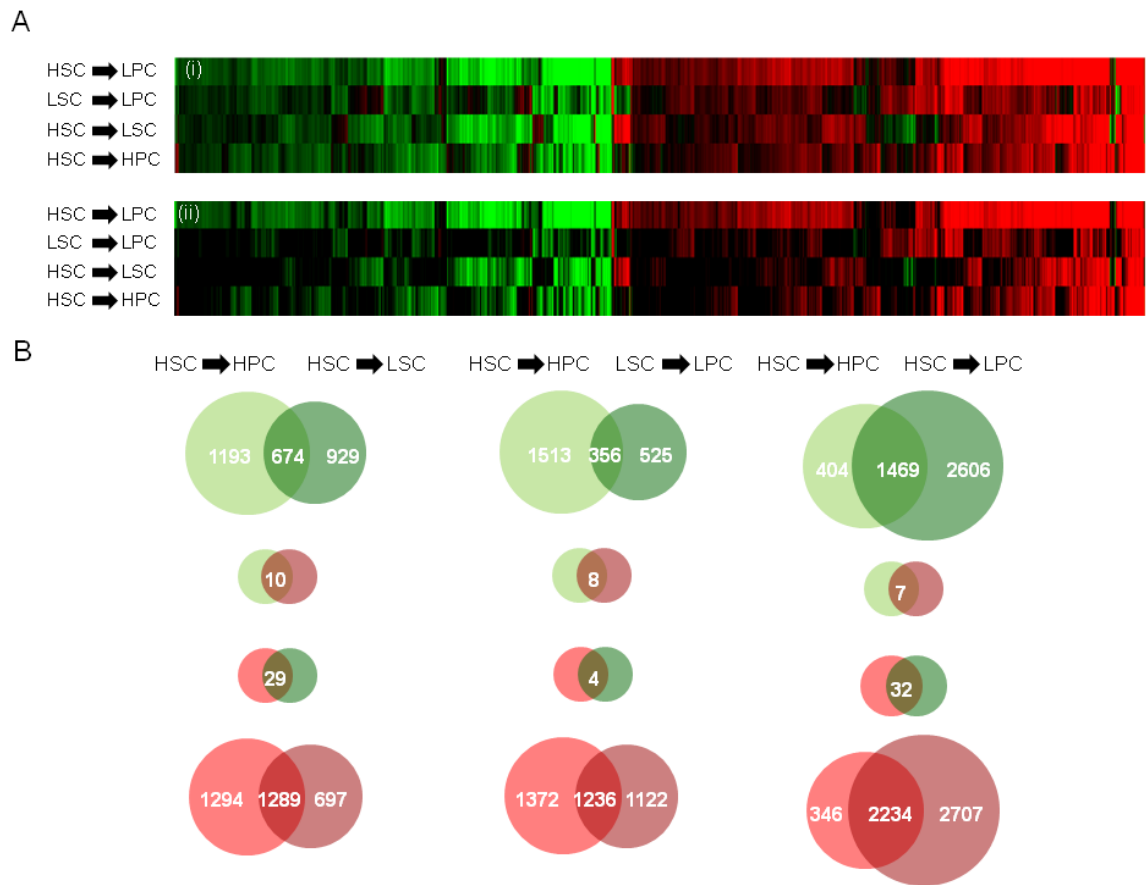
Figure 3.7: The direction of gene expression differences in the four studied comparisons is concordant.
(A) The heatmaps represent the genome-wide expression differences observed in each of the studied comparisons. (i) All expression changes that include both significant and non-significant and (ii) the significant (p<0.05) expression changes are depicted. The green bars indicate downregulation events and the red bars indicate upregulation events. (B) Venn diagrams represent the overlap between the significantly up- and downregulated genes (p<0.05) in the HSC-to-HPC transition and the HSC-to-LSC, LSC-to-LPC, and HSC-to-LPC comparisons. The number of significantly downregulated genes are shown in green circles, whereas the number of upregulated genes is represented in the red circles. The small Venn diagrams in between represent the discordant changes between the comparisons. Both heatmaps and Venn diagrams demonstrate a concordance in the direction of gene expression changes despite differences in their magnitude and significance.

The Venn diagrams highlighted earlier observations in heatmaps that the direction of gene expression changes were concordant in the four studied comparisons (Figure 3.7B). 78% of the genes that were significantly downregulated in HPCs relative to HSCs were also significantly downregulated in LPCs relative to HSCs. Furthermore, 87% of the significantly upregulated genes in HPCs relative to HSCs were also significantly upregulated in LPCs relative to HSCs. This observation demonstrated that the majority of the genes that were significantly differentially

expressed in the course of normal differentiation were also differentially expressed during LPC formation which can be associated with the proliferative status of both cell types. However, 65% of significantly downregulated and 55% of significantly upregulated genes between HSCs and LPCs were observed uniquely in this comparison. These gene expression differences, therefore, could be linked to the CML identity of LPCs.

The gene expression differences in the HSC-to-LSC and LSC-to-LPC transitions did not demonstrate similar degree of overlap with the gene expression differences between HSCs and HPCs (Figure 3.7B). 42% of the genes that were significantly downregulated in LSCs were also significantly downregulated in HPCs relative to HSCs. Moreover, 64% of the genes that were significantly upregulated in LSCs were also upregulated in HPCs relative to HSCs. These observations indicated that the gene expression signature of LSCs share similarities with HPCs, in particular with respect to the upregulation events.

Furthermore, 40% of the significantly downregulated and 52% of significantly upregulated genes between LPCs and LSCs were also significantly downregulated and upregulated in HPCs relative to HSCs, respectively (Figure 3.7B). This observation indicated that exit from stem cell compartment in CML shared similarities with exit from stem cell compartment in their normal counterparts. Nevertheless, as the gene expression signatures of LSCs shared a lot in common with HPCs, as shown above, and also closely-related to LPCs, as indicated by 3D PCA clustering, fewer gene expression changes are required to promote the LPC formation.

Overall, the key findings of the results presented in this section are summarised here. The direction of gene expression changes, both significant and non-significant, in the four studied comparisons was concordant despite differences in their significance level and magnitude of change. In other words, a subset of genes in HSCs are only programmed to be downregulated or repressed upon leaving the stem cell compartment, whereas another subset of genes in HSCs are only programmed to be upregulated or activated upon leaving the stem cell compartment. This finding is in agreement with the lineage priming model which was reviewed in Section 1.6.5. Furthermore, the leukaemic transformation was not

able to abnormally upregulate or downregulate genes that were destined for downregulation or upregulation, respectively, during normal differentiation.

Another finding of the above results indicated that the gene expression differences during LSC formation shared similarities with the expression changes during HPC differentiation. The transition from LSC to LPC also shared similarities with the HPC differentiation but to a lower level. This could provide further evidence that LSCs share similarities with the progenitor cells.

3.11 Validation of the identified gene expression differences by real-time PCR (RT-PCR)

The validation of gene expression differences observed by the Affymetrix hybridisation is a crucial step due to the presence of false positive discoveries as a result of cross-hybridisation or other labelling and array artefacts, as described earlier. Furthermore, due to the limited variance in the spot intensities as well as the normalisation processes, the fold expression differences reported by the arrays could be under- or over-estimates in some cases (Rajeevan et al., 2001).

A total of 77 genes were selected from significantly ($p<0.05$) differentially expressed genes between LSCs and HSCs that were identified by the Affymetrix analysis – 44 upregulated and 33 downregulated in LSCs. The genes were selected from three pathways that were of special interest to this research group, which include TGF-β signalling pathway, PI3K signalling pathway and the intracellular signalling downstream of BCR-ABL1 oncoprotein. TaqMan 384-well micro fluidic cards were designed using 18S rRNA, GAPDH, and RPLP0 as housekeeping genes. Each RT-PCR assay was performed in triplicate for each LSC sample and an RNA pool of HSC samples that were hybridised on the Affymetrix arrays, as described in Chapter 2. The absolute gene expression values were quantified for each probe relative to the average Ct values of the housekeeping genes and subsequently the $\log_2$ fold change differences were calculated using the HSC expression levels as the calibrator (Figure 3.8). 17 genes failed to demonstrate significant differential expression between LSCs and HSCs reported by Affymetrix. Therefore, 78% of the reported differences could be successfully validated by RT-PCR, which was greater than the previously reported success rate of 71% (Rajeevan et al., 2001).
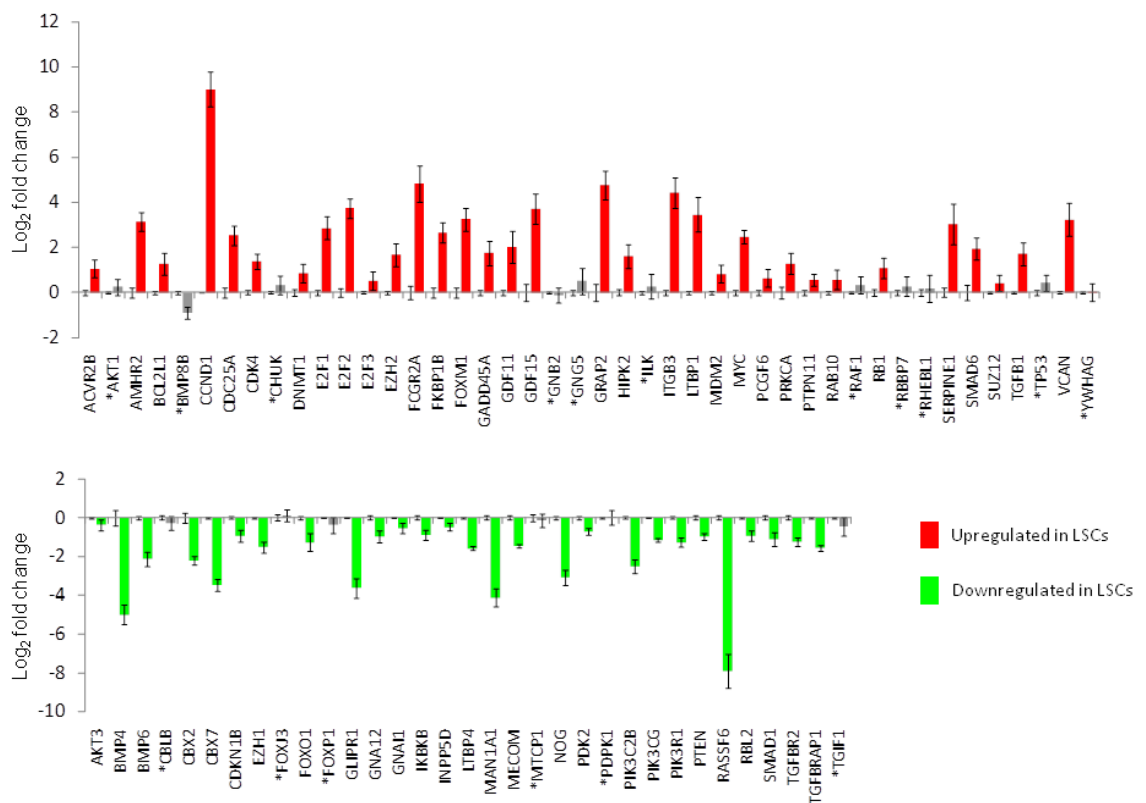
Figure 3.8: RT-PCR validation of differentially expressed genes identified by Affymetrix.
77 differentially expressed genes as identified by Affymetrix analysis between LSCs and HSCs were chosen from which 44 were upregulated (red bars) and 33 were downregulated (green bars) in LSCs. The log$_2$ fold enrichments (y-axis) were measured using the HSC expression levels as the calibrator. Each gene is represented with two sets of data; HSC log$_2$ fold expression which is set to 0 for all the samples (baseline) and LSC log$_2$ fold expression which is represented at the height of the bars. The error bars show the confidence intervals at 95% and in the absence of overlap between the error bars of HSC and LSC data points the differential expression was called significant and therefore, passed the validation step. The grey bars and the genes marked with asterisks failed the validation (n=17).

3.12 Identification of genes and pathways important for the maintenance of HSCs and their progression to the proliferative state

The implication of the identified gene expression differences in the context of biological pathways was investigated for each of the studied comparisons. The comparison between HPCs and HSCs identified 1,877 genes that were significantly downregulated in HPCs, which could be associated with the maintenance of HSCs. However, before analysing the biological pathways that were significantly associated with this subset of genes, the expression levels of several known HSC maintenance genes were investigated as an extra validation level. A recently published dataset identified 830 genes that were significantly upregulated in human HSCs relative to all the known haemopoietic progenitors and terminally differentiated haemopoietic lineage cells (Novershtern et al., 2011;

Section 3.1.3). A comparison was made between this published dataset and the HSC upregulated gene set from this study which identified 135 genes in common. The common gene set included several known HSC maintenance transcription factors and cell surface receptors (Section 3.1.3) including FLT3, GATA3, HLF, HOXA2, HOXA4, HOXA5, HOXA9, HOXB2, HOXB5, MECOM, MEIS1, NANOG, RBPMS, and SOCS2 (Figure 3.9).
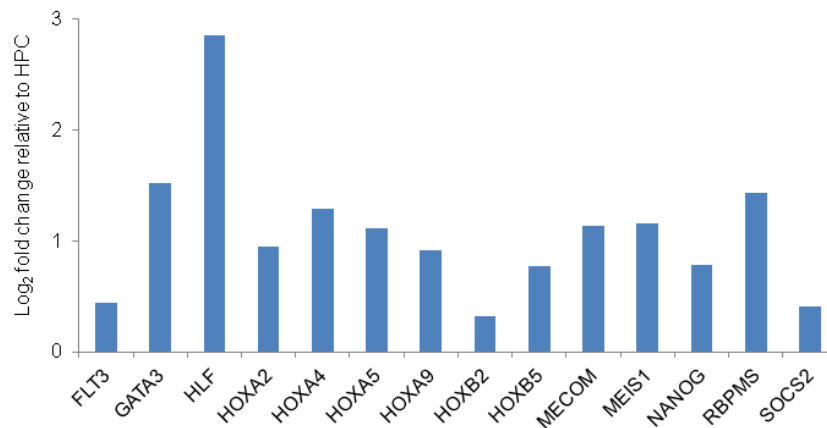
Figure 3.9: Several known HSC maintenance transcription factors are upregulated in HSCs relative to HPCs.
The log$_2$ fold change (y-axis) of gene expression in HSCs relative to HPCs was extracted from the significantly differentially expressed genes identified by Affymetrix analysis ($p<0.05$).

The significant gene expression differences between HSCs and HPCs were used to identify the biological pathways that were significantly upregulated or downregulated in this comparison (Table 3.3). Pathway analysis was performed using PANTHER database tool (Section 3.1.2.4). The list of significantly differentially expressed genes and their fold change differential expression values was used as an input. The statistical tool then created a distribution of values for the input data set, which is then used as a reference for distribution. Subsequently, the input data set was divided into functional categories that comprise biological pathways, and the distribution of values in each group was calculated and compared against the reference distribution. The probability of obtaining a random distribution from the reference distribution in each category was measured by the Mann-Whitney Rank-Sum test (*U*-test) (Clark et al, 2003; Thomas et al, 2006). The output data was represented as p-values associated with the direction of shift for the distribution of values in each group relative to the reference distribution. The categories with non-random distribution were associated with small and significant p-values.

Pathways	No. of genes in differential expression list	p-value	
Wnt signalling pathway	80	3.99E-04	
Histamine H1 receptor mediated signalling pathway	15	4.98E-03	
5HT2 type receptor mediated signalling pathway	16	7.10E-03	
Muscarinic acetylcholine receptor 1 and 3 signalling pathway	18	9.95E-03	
Oxytocin receptor mediated signalling pathway	17	1.05E-02	
PDGF signalling pathway	39	1.17E-02	
Endothelin signalling pathway	26	1.18E-02	↑ HSCs
TGF-beta signalling pathway	40	1.46E-02	
Inflammation mediated by chemokine and cytokine signalling pathway	72	1.79E-02	
Thyrotropin-releasing hormone receptor signalling pathway	17	1.95E-02	
Interleukin signalling pathway	40	2.72E-02	
Alpha adrenergic receptor signalling pathway	7	2.90E-02	
Angiotensin II-stimulated signalling through G proteins and beta-arrestin	10	3.91E-02	
De novo purine biosynthesis	21	4.65E-06	
De novo pyrimidine deoxyribonucleotide biosynthesis	12	2.67E-04	
DNA replication	12	4.85E-04	
Ubiquitin proteasome pathway	27	1.08E-03	
Serine glycine biosynthesis	6	5.65E-03	
Parkinson disease	26	5.84E-03	
Haem biosynthesis	7	5.96E-03	
Transcription regulation by bZIP transcription factor	15	5.98E-03	
TCA cycle	8	8.76E-03	
Folate biosynthesis	4	9.07E-03	↑ HPCs
Methylmalonyl pathway	3	9.78E-03	
Succinate to proprionate conversion	3	9.78E-03	
p53 pathway	40	1.14E-02	
Cell cycle	10	1.59E-02	
Formyltetrahydroformate biosynthesis	5	1.63E-02	
p53 pathway feedback loops 2	26	2.08E-02	
Adenine and hypoxanthine salvage pathway	4	3.17E-02	
General transcription regulation	11	3.56E-02	
Salvage pyrimidine ribonucleotides	6	3.95E-02	

Table 3.3: Pathways significantly upregulated in HSCs or HPCs.
The table summarises a list of pathways showing a significant (p<0.05) shift of distribution in the HSC or HPC direction relative to the overall distribution. The number of genes affected in each pathway and the calculated p-values (*U*-test) are also shown. The pathways are ranked based on their significance in terms of up-regulation or down-regulation in HSCs or HPCs respectively.

Thirteen biological pathways showed significantly higher expression in HSCs relative to HPCs. Conversely, nineteen pathways were significantly upregulated in

HPCs relative to HSCs. Intriguingly, all the 13 pathways that were downregulated upon commitment to HPCs were involved in cell signalling, including Wnt, platelet-derived growth factor (PDGF), endothelin, TGF-β, interleukins, and inflammation mediated by chemokines and cytokines signalling pathways. Furthermore, several neurotransmitter signalling pathways were also downregulated in HPCs, which included histamine via H1 receptor, serotonin via 5HT2 receptors, acetylcholine via muscarinic receptors 1 and 3, oxytocin, thyrotropin-releasing hormone (TRH), catecholamines via α-adrenergic receptors, and angiotensin II signalling pathways. The importance of Wnt, TGF-β, chemokines and cytokines signalling pathways in the maintenance of self-renewal and quiescence of HSCs have been previously reported (reviewed in section 1.6.4) (Reya et al., 2003, Sitnicka et al., 1996, Steidl et al., 2004). Furthermore, neurotransmitter receptors have also been reported on the cell surface of HSCs which were downregulated in the course of differentiation (Steidl et al., 2004; Section 1.6.4.7). These 13 pathways were, therefore, referred to as the HSC identity pathways.

Conversely the pathways upregulated in HPCs were associated with cellular metabolism, DNA replication, ubiquitin proteasome pathway, transcriptional regulation, p53 induction and regulation, and cell cycle. The upregulation of these pathways was in agreement with the proliferative nature of HPCs and therefore, were referred to as the HPC identity pathways. These pathways were also shown to be upregulated in the $CD34^+$ $CD38^+$ population relative to the $CD34^+$ $CD38^-$ population in previous studies (Georgantas et al., 2004, Ivanova et al., 2002).

The identification of previously reported biological pathways that regulate HSC and HPC identities could confirm that appropriate cell populations were purified and the differential expression analyses were performed accurately to reflect true and significant biological differences.

3.13 LSCs share similarities with both HSCs and HPCs at the pathway level.

The significant gene expression differences between HSCs and LSCs were also subjected to pathway analysis in order to identify the similarities and differences between HSCs and LSCs at the pathway level (Table 3.4).

Pathways	No. of genes in differential expression list	p-value	
Toll receptor signalling pathway	11	5.37E-03	
Endothelin signalling pathway	20	3.08E-02	↓ LSCs
T cell activation	22	3.99E-02	
Parkinson disease	25	3.15E-05	
Ubiquitin proteasome pathway	28	1.12E-04	
Cell cycle	12	9.77E-04	
Cytoskeletal regulation by Rho GTPase	26	2.81E-03	
p53 pathway	40	3.85E-03	
Cholesterol biosynthesis	6	5.37E-03	
De novo purine biosynthesis	14	6.25E-03	
Integrin signalling pathway	41	9.66E-03	
De novo pyrimidine deoxyribonucleotide biosynthesis	7	1.08E-02	
Blood coagulation	11	1.09E-02	↑ LSCs
Formyltetrahydroformate biosynthesis	7	1.13E-02	
DNA replication	10	1.35E-02	
Apoptosis signalling pathway	33	1.47E-02	
Methylmalonyl pathway	3	2.76E-02	
Succinate to proprionate conversion	3	2.76E-02	
Haem biosynthesis	5	3.90E-02	
Xanthine and guanine salvage pathway	2	4.21E-02	
Transcription regulation by bZIP transcription factor	8	4.35E-02	
General transcription by RNA polymerase I	4	4.75E-02	
Ornithine degradation	2	4.81E-02	

Table 3.4: Pathways significantly upregulated or downregulated in LSCs relative to HSCs.
The table summarises a list of pathways showing a significant (p<0.05) shift of distribution in a positive (LSC) or negative (HSC) direction relative to the overall distribution. The number of genes affected in each pathway and the calculated p-values (*U*-test) are also shown. The pathways are ranked based on their significance in each direction.

Endothelin signalling was the only HSC identity pathway that was significantly downregulated in LSCs, whereas the remaining 12 HSC identity pathways were not significantly down- or upregulated in LSCs. These pathways include TGF-β, Wnt, PDGF, chemokine and cytokine, interleukins and several neurotransmitter signalling pathways (Section 3.10). Therefore, the HSC identity pathways were expressed at similar levels in both HSCs and LSCs, which may propose their function in LSC maintenance. The role of TGF-β and Wnt signalling pathways in

the maintenance of quiescence and self-renewal capacities of LSCs were reported previously (Section 1.7.6) (Naka et al., 2010, Zhao et al., 2007). Conversely, two reports showed significant downregulation of several chemokine ligands and receptors in LSCs relative to their normal counterparts (Bruns et al., 2009, Graham et al., 2007). However, this was not in agreement with the above findings that showed no significant change of expression in the chemokine and cytokine signalling pathway between HSCs and LSCs. Furthermore, several neurotransmitter receptors were also reported to be expressed on the surface of LSCs (Kronenwett et al., 2005), although the reported receptors were not associated with the neurotransmitter pathways that were identified above.

Twenty pathways were significantly upregulated in LSCs relative to HSCs, 12 of which were involved with the HPC identity (Section 3.10). These included ubiquitin proteasome pathway, cell cycle, DNA replication, p53 induction and several metabolic pathways. This observation was in agreement with the previous reports that showed the upregulation of cell cycle and DNA replication genes in LSCs relative to their normal counterparts (Graham et al., 2007, Kronenwett et al., 2005).

In addition to the pathways that were associated with the HSC and HPC identities, several other biological pathways were also identified to be significantly misregulated between HSCs and LSCs. Toll-like receptor signalling and T-cell activation pathways showed a significant downregulation in LSCs, whereas integrin signalling, blood coagulation, cytoskeletal regulation by Rho GTPase, and apoptosis signalling were upregulated in LSCs. Therefore, these pathways could be associated with the CML identity.

LSCs, therefore, share similarities with both HSCs and HPCs at the pathway level. The shared similarities with HPCs was expected from the overlap between the gene expression differences that observed in the HSC-to-HPC and HSC-to-LSC transitions (Figure 3.7B). The persistent activity of the HSC identity pathways in LSCs could provide further evidence for the stem cell features of LSCs such as self-renewal and quiescence despite the upregulation of many proliferative pathways.

3.14 LPCs show a significant loss of HSC identity.

The significantly differentially expressed genes between LSCs and LPCs were analysed at the pathway level to identify the pathways that were significantly down- or upregulated in the course of LSC commitment to LPCs (Table 3.5).

Pathways	No. of genes in differential expression list	p-value	
Angiogenesis	29	1.75E-04	
Integrin signalling pathway	21	7.81E-04	
5HT2 type receptor mediated signalling pathway	15	1.05E-03	
PDGF signalling pathway	28	1.68E-03	
Heterotrimeric G-protein signalling pathway-Gi alpha and Gs alpha mediated pathway	24	1.74E-03	
Heterotrimeric G-protein signalling pathway-Gq alpha and Go alpha mediated pathway	17	2.18E-03	
Oxytocin receptor mediated signalling pathway	13	5.86E-03	
Blood coagulation	4	9.26E-03	
Thyrotropin-releasing hormone receptor signalling pathway	13	1.18E-02	
Cadherin signalling pathway	19	1.40E-02	
Inflammation mediated by chemokine and cytokine signalling pathway	36	1.52E-02	↓ LPCs
Interleukin signalling pathway	23	1.57E-02	
Alzheimer disease-amyloid secretase pathway	13	1.66E-02	
VEGF signalling pathway	9	1.72E-02	
Wnt signalling pathway	45	2.05E-02	
JAK/STAT signalling pathway	4	2.57E-02	
Nicotinic acetylcholine receptor signalling pathway	10	2.58E-02	
FGF signalling pathway	22	2.75E-02	
PI3 kinase pathway	22	3.96E-02	
EGF receptor signalling pathway	29	4.04E-02	
TGF-beta signalling pathway	23	4.38E-02	
Ras Pathway	11	4.58E-02	
DNA replication	13	7.95E-03	↑ LPCs
De novo pyrimidine deoxyribonucleotide biosynthesis	9	2.34E-02	
Cell cycle	13	2.66E-02	

Table 3.5: Pathways significantly upregulated or downregulated in LPCs relative to LSCs.
The table summarises a list of pathways showing a significant (p<0.05) shift of distribution in a positive (LPC) or negative (LSC) direction relative to the overall distribution. The number of genes affected in each pathway and the calculated p-values (*U*-test) are also shown. The pathways are ranked based on their significance in each direction. The HSC identity pathways are highlighted in green.

Eight HSC identity pathways were shown to be significantly downregulated in LPCs. These included Wnt, TGF-β, PDGF, interleukin, and inflammation via

chemokine and cytokine signalling pathways, as well as serotonin signalling via 5HT2 receptors, oxytocin, and TRH signalling pathways. The downregulation of these pathways in the transition from LSCs to LPCs provides a second line of evidence that they may be important in maintaining the stem cell phenotype of LSCs. However, four HSC identity pathways did not show a significant up- or downregulation in LPCs, which include α-adrenergic receptor signalling, muscarinic acetylcholine receptors 1 and 3 signalling, histamine H1 receptor signalling and angiotensin II receptor signalling pathways. Therefore, it can be argued that these pathways may confer a less mature phenotype to LPCs.

Furthermore, three pathways associated with the HPC identity were upregulated in LPCs. These included DNA replication, the cell cycle, and de novo pyrimidine biosynthesis. These three pathways had also been upregulated in LSCs relative to HSCs, but were upregulated even further in the transition from LSCs to LPCs. Thus, the upregulation of these three pathways may also constitute key determinants in the transition from LSCs to LPCs.

Additionally, pathways involved in angiogenesis, blood coagulation, RAS pathway, heterotrimeric Gα protein signalling, integrin, cadherin, vascular endothelial growth factor (VEGF), JAK/STAT, FGF, epidermal growth factor (EGF), PI3K, and nicotinic acetylcholine signalling were uniquely downregulated in the transition from LSCs to LPCs. Therefore, these pathways could also be involved with the CML identity, since they were not associated with any significant changes during normal haemopoiesis.

Therefore, the results indicate that the exit from stem cell compartment in CML is associated with the loss of HSC identity in addition to acquiring further proliferative capacities by upregulating the pathways involved in cell cycle and DNA replication.

Overall, the three sets of pathway analyses described above identified three major classes of pathways, including HSC identity, HPC identity, and CML identity (Figure 3.10). The HSC identity pathways could be involved in the regulation of self-renewal and quiescence, whereas the HPC identity pathways were associated with proliferation mechanisms. The CML identity pathways could be linked to proliferation, differentiation, survival, cell-cell contact, migration, and vascularisation.
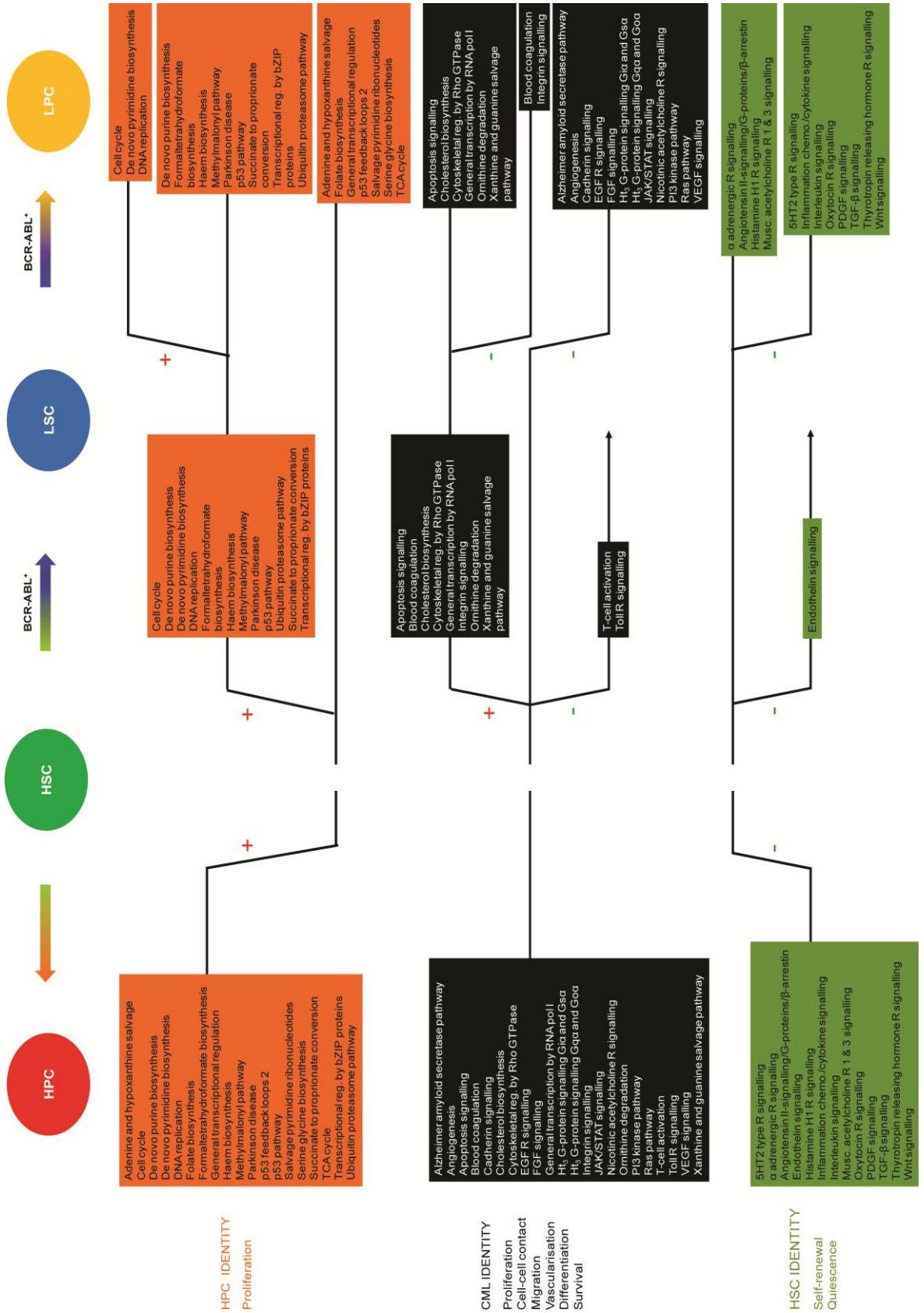
Figure 3.10: Pathways significantly affected during normal commitment and leukaemic transformations.

The schematic diagram shows pathways significantly (p<0.05) associated with HPC identity (red boxes), CML identity (black boxes), and HSC identity (green boxes). The significant up- and downregulation of pathways in each transition are marked with (+) and (-) signs, respectively. The pathways that were not significantly affected in a transition are connected by a straight line. 5HT, 5-hydroxytryptamine (serotonin); Musc., muscarinic; R., receptor; reg., regulation.

The above observations were further strengthen by investigating the biological pathways that were significantly up- or downregulated by the significant gene expression changes between HSCs and LPCs (Table 3.6). This analysis would allow expression levels of pathways to be measured accurately between the cell types.

Pathways	No. of genes in differential expression list	p-value	
Wnt signalling pathway	143	2.35E-03	
Histamine H1 receptor mediated signalling pathway	28	2.77E-03	
Thyrotropin-releasing hormone receptor signalling pathway	29	3.74E-03	
Oxytocin receptor mediated signalling pathway	30	5.16E-03	
Heterotrimeric G-protein signalling pathway-Gq alpha and Go alpha mediated pathway	60	5.92E-03	↓ LPCs
5HT2 type receptor mediated signalling pathway	31	6.95E-03	
Nicotine degradation	7	1.27E-02	
Endothelin signalling pathway	39	2.33E-02	
Angiogenesis	77	2.79E-02	
VEGF signalling pathway	37	3.48E-02	
Acetate utilization	2	4.77E-02	
Ubiquitin proteasome pathway	49	5.30E-08	
Cell cycle	20	1.11E-06	
Transcription regulation by bZIP transcription factor	32	3.67E-05	
General transcription regulation	24	3.32E-04	
De novo purine biosynthesis	24	7.23E-04	
Haem biosynthesis	9	1.06E-03	
Cholesterol biosynthesis	11	1.57E-03	
Parkinson disease	59	1.66E-03	
p53 pathway	67	1.72E-03	
DNA replication	20	3.49E-03	
General transcription by RNA polymerase I	9	5.84E-03	↑ LPCs
Xanthine and guanine salvage pathway	3	1.16E-02	
De novo pyrimidine deoxyribonucleotide biosynthesis	14	1.33E-02	
ATP synthesis	4	1.46E-02	
Serine glycine biosynthesis	4	1.99E-02	
Glycolysis	14	2.22E-02	
TCA cycle	12	2.28E-02	
5-Hydroxytryptamine degradation	9	2.89E-02	
Ascorbate degradation	2	3.50E-02	
Asparagine and aspartate biosynthesis	4	4.25E-02	
Pentose phosphate pathway	4	4.70E-02	

Table 3.6: Pathways significantly upregulated or downregulated in LPCs relative to HSCs. The table summarises a list of pathways showing a significant (p<0.05) shift of distribution in a positive (LPC) or negative (HSC) direction relative to the overall distribution. The number of genes

affected in each pathway and the calculated p-values (*U*-test) are also shown. The pathways are ranked based on their significance in each direction. The HSC identity pathways are highlighted in green.

The results showed that the formation of LPCs from HSCs was associated with significant downregulation of HSC identity pathways and the acquisition of proliferative capacities, which was in agreement with the earlier pathway analysis between LPCs and LSCs.

3.15 Analysis of expression changes of the genes in the HSC identity pathways

In order to further understand the fate of HSC identity pathways in LSCs and LPCs, their relative expression levels in the four cell types were established by performing pathway analysis for all the possible comparisons between the four cell types (Figure 3.11).
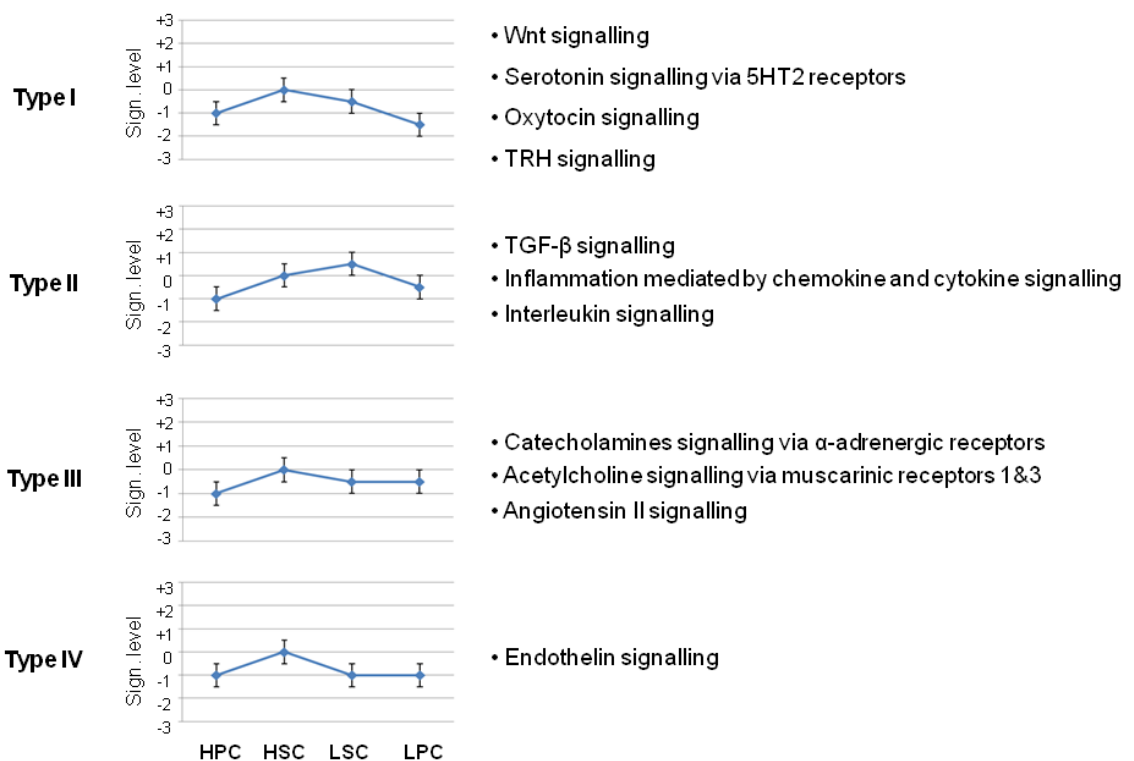
Figure 3.11: The expression levels of stem cell maintenance pathways in the four cell types. The expression levels of stem cell maintenance pathways were analysed in the four cell types in relation to each other. The pathways were categorised into four major classes based on the similarities in their expression level in the four cell types. The horizontal grey lines indicate significance levels (p<0.05). The error bars represent the 95% confidence intervals and the significant differences are assigned in the absence of overlap between the error bars.

The HSC identity pathways were by definition significantly downregulated in HPCs relative to HSCs. By performing pathway analysis in all possible comparisons between the four cell types the expression levels of each pathway in each cell type with relation to the other cell types could be distinguished. From 13 HSC identity pathways the expression levels of two pathways could not be reliably established between the four cell types, which include PDGF signalling and histamine H1 receptor signalling pathways. The fate of remaining 11 HSC identity pathways in LSCs and LPCs could be classified into four major classes (Figure 3.11).

Type I pathways included Wnt, serotonin via 5HT2 receptors, oxytocin, and TRH signalling pathways which demonstrated a non-significant downregulation in LSCs with further significant downregulation in the progenitor cells with LPCs representing the lowest expression levels of all. Type II pathways, which included TGF-β, PDGF, interleukins, and inflammation mediated by chemokine and cytokine signalling, exhibited a non-significant upregulation in LSCs, but were significantly downregulated upon commitment to LPCs. However, their expression level in LPCs was not significantly lower than HSCs which indicated a higher level of activity for these pathways in LPCs compared to their normal counterparts, HPCs. Type III pathways included α-adrenergic receptors signalling, muscarinic acetylcholine receptors 1 and 3 signalling, and angiotensin II signalling pathways. These pathways showed a non-significant downregulation in both LSCs and LPCs relative to HSCs. Therefore, their expression level in CML cells was in an intermediated state between HSCs and HPCs. Type IV category included endothelin signalling pathway that showed a uniform significant downregulation in HPCs, LSCs, and LPCs relative to HSCs.

The results indicated that the type I and type II HSC identity pathways showed no significant differential expression between LSCs and HSCs, but were significantly downregulated upon commitment to LPCs. Thus, it was plausible to associate them with the maintenance of stem cell-like features in LSCs. Additionally, type III pathways were not significantly different between LSCs and LPCs relative to HSCs. As a result, these pathways may confer a less mature phenotype to LPCs.

Similar to the earlier analysis that was performed for the genome-wide expression changes in the four studied comparisons, the gene expression differences of the genes in the HSC identity pathways were also examined in each comparison. In

total, 845 genes were associated with the 13 HSC identity pathways. Heatmaps were generated for all the gene expression changes, including both significant and non-significant changes, as well as the significant gene expression changes in order to investigate the concordance between the direction of gene expression changes between the four comparisons (Figure 3.12A). Similar to the earlier observations at the genome-wide level the direction of gene expression changes was concordant between the four comparisons despite differences in their significance level and their magnitude of change. This could suggest that the genes in the HSC identity pathway could be under a global gene regulation programme upon the exit from stem cell compartment. Furthermore, the numbers of significantly up- and downregulated genes in the HSC identity pathways in each comparison were identified and the overlap between the significant changes during normal commitment and the leukaemic transformations was established (Figure 3.12B). The significant gene expression changes in the HSC identity pathways identified 88 upregulated and 103 downregulated genes in the HSC-to-HPC transition. This indicated a preference towards downregulation, which could be expected from the nature of these pathways that were significantly downregulated in HPCs. Furthermore, the gene expression changes of the HSC identity pathways in the HSC-to-LSC transition were balanced between up- and downregulation with 76 and 79 genes, respectively.

The same analysis in the LSC-to-LPC transition identified 57 significantly upregulated genes as opposed to 46 significantly downregulated genes. This indicated a preference towards the upregulation events, although fewer genes were significantly changed relative to the other transitions. However, when the overall comparison was made between HSCs and LPCs, 188 genes were identified to be significantly downregulated which was slightly higher than 181 significantly upregulated genes. Therefore, the preference towards downregulation was restored similar to the normal commitment. Nevertheless, the number of significantly affected genes were doubled in either directions relative to the HSC-to-HPC transition. Furthermore, 80% (n=70) of the genes that were significantly upregulated in HPCs were also significantly upregulated in LPCs relative to HSCs; and 70% (n=73) of the significantly downregulated genes in HPCs were also significantly downregulated in LPCs relative to HSCs. These observations indicated that the downregulation of HSC identity pathways in LPCs was mediated by suppressing more genes in these pathways than HPCs.
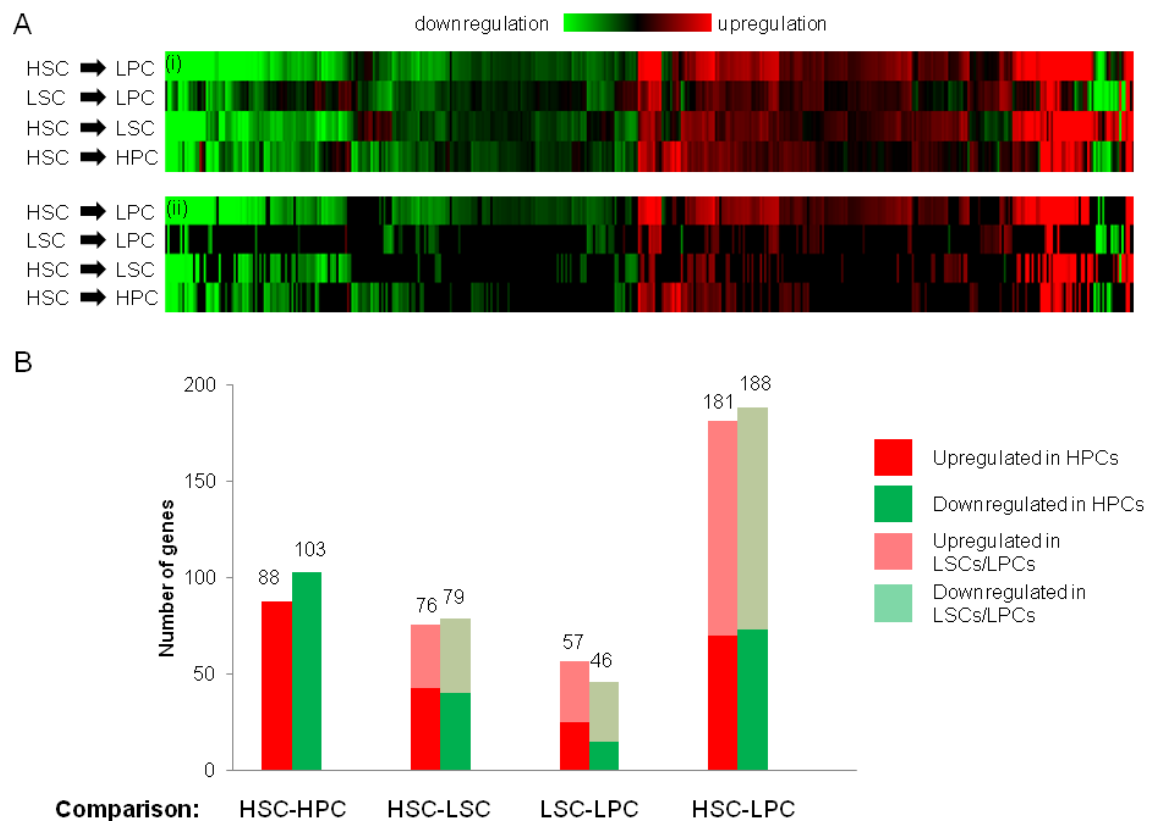
Figure 3.12: The significant expression changes of the genes in the stem cell pathways.
(A) The heatmaps represent the expression differences observed for the genes in the HSC identity pathways in each of the studied comparisons. (i) All expression changes (n=845) that include both significant and non-significant and (ii) the significant (p<0.05) expression changes are depicted. The green bars indicate downregulation events and the red bars indicate upregulation events. (B) The number of significantly (p<0.05) upregulated (red bars) and downregulated genes in the HSC identity pathways in each of the studied comparisons. The overlap between the significant changes in the HSC-to-HPC transition and the other three leukaemic transformations is shown as darker overlaid red and green bars. The total number of gene expression differences in each transition is also indicated on top of each bar.

Overall, it can be argued that the genes in the HSC identity pathways were regulated by a global programme upon the exit from HSC compartment, which may be similar to the global programming that regulate genome-wide gene expression changes. However, the scope of transcriptional repression in the genes of the HSC identity pathways was a lot bigger during LPC formation compared to the normal development. This could suggest potential differences in the global programming between CML cells and their normal counterparts.

3.16 Discussion

The work presented in this Chapter showed the gene expression patterns of CML cells and their normal counterparts. Furthermore, the study of biological pathways that were significantly up- or downregulated during normal commitment and

leukaemic transitions could help the understanding of underlying biology in the maintenance of LSCs and their exit from stem cell compartment during LPC formation. The key findings of this Chapter are discussed below.

3.16.1   *LSCs share similarities with both HSCs and HPCs.*

The PCA clustering of the expression profiles of the four studied cell types, demonstrated that LSCs were located between HSCs and LPCs. Moroever, the distance between LSCs and HSCs in the 3D environment was similar to the distance between HSCs and HPCs, although LSCs and HPCs were not overlapping. This observation provided the first line of evidence that LSCs and HSCs do not completely match with regards to their gene expression profiles and demonstrate similarities to the progenitor cell types, HPCs and LPCs. The second line of evidence was obtained from the analysis of genome-wide gene expression differences. The majority of significant changes between HSCs and LSCs were also found to be significantly changing between HSCs and HPCs. This observation provided further evidence that LSCs share similarities at the level of gene expression differences with HPCs. The third line of evidence was obtained from the pathway analysis performed for the significantly differentially expressed genes between HSCs and LSCs. The pathway analysis identified 12 HSC identity pathways to be expressed at similar levels between HSCs and LSCs. Furthermore, 12 HPC identity pathway were also significantly upregulated in LSCs relative to HSCs. This observation provided evidence at the pathway level that LSCs share similarities with both HSCs and HPCs.

Previous gene expression analysis of LSCs and CML progenitors also demonstrated significant similarities between LSCs and the progenitor cells based on the gene expression differences (Bruns et al., 2009, Graham et al., 2007). These studies identified fewer expression differences between LSCs and the early progenitors in CML than LSCs and their normal counterparts. However, these studies only investigated the significant gene expression differences between cell types without providing further evidence on the similarities at the level of biological pathways. Furthermore, it was reported that the genes associated with DNA replication, cell cycle and mitosis were significantly upregulated in LSCs relative to HSCs (Graham et al., 2007, Kronenwett et al., 2005), which was in agreement with the observations in this Chapter.

The observations in this Chapter were also in agreement with the phenotype of LSCs as reported by several studies. LSCs have been shown to reside in the G_0 cell cycle compartment and exhibited stem cell properties both *in vitro* and *in vivo* (Holyoake et al., 1999). Nevertheless, LSCs were demonstrated to exit quiescence state more readily than HSCs (Traycoff et al., 1998) and also exhibited reduced self-renewal capacities as shown by the LTC-IC assays (Udomsakdi et al, 1992). Furthermore, unlike their normal counterparts, LSCs demonstrate a growth factor-independent survival mechanism as a result of abnormal proliferative capacities (Bedi et al., 1994, Strife et al., 1988). These observations provide further evidence that show LSCs share similarities with both stem cell compartment and proliferative cells.

3.16.2   LPCs show a significant loss of HSC identity.

The pathway analysis for the genes that were significantly differentially expressed between LSCs and LPCs identified eight HSC identity pathways to be significantly downregulated in LPCs. However, only three proliferative pathway associated with HPC identity was significantly upregulated, such as cell cycle and DNA replication. This finding was further strengthened by performing pathway analysis between the genes that were significantly differentially expressed between HSCs and LPCs, which confirmed the above findings. Therefore, the exit from stem cell compartment in CML is mediated by the loss of pathways associated with stem cell maintenance such as Wnt, TGF-β and chemokine and cytokine signalling as well as acquiring more proliferative capacities. It can be argued that the HSC identity pathways block the derive of upregulated proliferative pathways in LSCs and upon their downregulation LSCs will start to proliferate and differentiate. However, It can be conversely argued that the derive for the exit from stem cell compartment in CML was mediated by the upregulation of the three upregulated proliferative pathways in LPCs or a combination of both.

In addition to the downregulation of eight HSC identity pathways, the expression of four HSC identity pathways remained unchanged between LSCs and LPCs. The persistent activity of pathways associated with stem cell maintenance could confer a less mature phenotype to LPCs.

3.16.3   The role of HSC identity pathways in the maintenance of LSCs

TGF-β and Wnt signalling pathways, which are the master regulators of HSC quiescence and self-renewal (Reya et al., 2003, Sitnicka et al., 1996), retained their expression levels in LSCs. These pathways were significantly downregulated in LPCs, which provide further evidence for their role in the maintenance of stem cell features in LSCs.

The role of TGF-β signalling in the maintenance of LSCs was recently reported (Naka et al., 2010). The activated TGF-β signalling in LSCs was shown to block the phosphorylation and, consequently, activation of the AKT protein despite BCR-ABL1 expression. As a result, FOXO3A transcription factor was not degraded and induced quiescence in LSCs. This would further highlight the possible interactions between the stem cell maintenance pathways and the upregulated proliferative pathways in LSCs. As mentioned earlier, one argument for the exit from stem cell maintenance could indicate an important role for the loss of HSC identity pathways that would release the block on the proliferative pathways to induce proliferation and the formation of LPCs.

Furthermore, the role of Wnt signalling in LSC maintenance has also been reported in recent studies. The loss of β-catenin in the murine models of CML resulted in reduced self-renewal capacities of LSCs and increased the survival rate of the mice (Zhao et al., 2007). Increased activity of Wnt signalling have also been reported in the case of self-renewing GMPs during blast crisis (Jamieson et al., 2004). Therefore, the downregulation of Wnt signalling in LPCs could indicate their importance in self-renewal regulation of LSCs.

The results in this Chapter also showed that the chemokine and cytokine signalling pathway retained its expression level in LSCs, but was subsequently downregulated in LPCs. However, previous reports suggested significant downregulation of the genes associated with chemokine ligands and receptors in LSCs relative to HSCs (Graham et al., 2007, Kronenwett et al., 2005). The expression levels of these genes were subsequently downregulated in the dividing CML population (Graham et al., 2007). These analyses were performed at the gene level and the expression changes of various members of the pathway were not taken into consideration. Therefore, although individual genes in each pathway

may be significantly up- or downregulated but such changes may not necessarily result in a significant expression change of the whole pathway.

3.16.4   Neurotransmitters signalling pathways may have a role in the maintenance of stem cell features in LSCs.

In addition to TGF-β and Wnt signalling whose roles in the maintenance of stem cells have been elucidated, several neurotransmitter signalling pathways were also found to be associated with the HSC identity. These pathways included acetylcholine, adrenergic (epinephrine and norepinephrine), serotonin, histamine, angiotensin II, oxytocin, and TRH signalling pathways. The neurotransmitter signalling pathways also maintained their expression level in LSCs, which could imply their role in the maintenance of LSCs. Subsequently, in the course of LPC development serotonin, oxytocin and TRH signalling pathways were significantly downregulated. However, acetylcholine, adrenergic, histamine H1 and angiotensin II signalling pathways were the only HSC identity pathways that did not show a significant downregulation in LPCs relative LSCs. This could suggest their role in conferring a less mature phenotype to LPCs in comparison to their normal counterparts.

The BM has been shown to be substantially innervated in close proximity to arterial smooth muscle cells, although no direct contact between HSCs and nerves has been reported (Mignini et al., 2003). The functional link between HSCs and neurons was established through observation of neurotransmitter receptors on these cell types, including dopamine, adrenergic, opioid, serotonin, adenosine, orexins, tachykinins, GABA-B, VIP, CRH, somatostatin, and angiotensin II receptors (Kalinkovich et al., 2009, Steidl et al., 2004). These neurotransmitter receptors were shown to be subsequently downregulated in the course of differentiation (Steidl et al., 2004). These observations was in agreement with the pathway analysis performed between HSCs and HPCs.

Furthermore, similar neurotransmitter pathways were also identified on the surface of CML CD34$^+$ cells which were expressed at higher levels relative to their normal counterparts (Kronenwett et al., 2005). Although the observation of neurotransmitter receptors on the surface of CML cells was in agreement with the findings of this work, they did not identify the same type of receptors.

Recent studies have reported the roles of neurotransmitter signalling in the maintenance of HSCs through interactions with other stem cell maintenance pathways. For instance, adrenergic receptors were identified on the surface of murine LSK population (Muthu et al., 2007) as well as human HSCs (CD34$^+$ CD38$^-$) (Spiegel et al., 2007). Furthermore, stimulation of adrenergic signalling pathway in human HSCs induced BM repopulation via activating the canonical Wnt signalling pathway (Spiegel et al, 2007). A recent report also suggested that norepinephrine induces the production of HSCs from their haemogenic endothelium precursors in AGM; thus, providing compelling evidence on the role of neurotransmitters in the formation of HSCs (Fitch et al., 2012). Serotonin receptors, in particular the 5HT2 type, were demonstrated to induce expansion of primitive cord blood CD34$^+$ cells in response to serotonin. This in turn enhanced their engraftment potential after transplantation into NOD/*scid* mice (Yang et al., 2007). Additionally, the nonmyelinating Schwann cells, which are a subset of the neural cells in BM niche, were shown to be required for the activation of the latent form of TGF-β and therefore, indirectly maintain the quiescence of HSCs (Yamazaki et al., 2011). These evidence, therefore, provide further support for the role of neurotransmitters in the maintenance of HSCs directly or indirectly.

However, the neurotransmitter signalling was also linked to the induction of HSC mobility and proliferation. Norepinephrine treatment of HSCs induced motility in response to both G-CSF and GM-CSF cytokines (Spiegel et al., 2007). Angiotensin II, was also demonstrated to be capable of stimulating the expansion of myeloid (CFU-GM) and erythroid (BFU-E) colonies through activation of angiotensin receptor II in immature human cord blood CD34$^+$ cells (Peng et al., 2005). Therefore, some may also play a role in mobilisation, maturation and proliferation of HSCs. This could be linked with the earlier observations in this Chapter that showed four neurotransmitter pathways retained their expression levels in LPCs similar to HSCs and LSCs. Thus, the role of neurotransmitters could be context-dependent.

Despite discovering the neurotransmitter receptors, their role in leukaemia and particularly in the maintenance of LSCs has not been reported previously. Thus, the identification of these neurotransmitters signalling pathways in this Chapter suggest they may contribute to stem cell identity along with known pathways that promote self-renewal and quiescence. Such preliminary data requires further *in*

vitro and *in vivo* validation to elucidate the roles of these pathways in maintaining a stem cell phenotype. Therefore, the role of these identified neurotransmitters in the LSC maintenance is going to be analysed *in vitro* in Chapter 6.

3.16.5   The CML identity pathways could be important in LSC maintenance

Several biological pathways were identified to be significantly up- or downregulated in the HSC-to-LSC and LSC-to-LPC transitions which were not significantly changed during normal commitment. These pathways were, therefore, linked to the CML identity. Integrin signalling and the cytoskeletal regulation by Rho GTPases were two CML identity pathways that were significantly upregulated in LSCs relative to HSCs. Integrin signalling pathway was subsequently downregulated in LPCs. Therefore, these pathways may also contribute to the regulation of other aspects of stem cell maintenance such as adhesion and cell-cell contact in the niche.

It was reported that BCR-ABL expression enhances fibronectin (extracellular matrix) adhesion via integrin signalling mediated by VLA-4 ($\alpha_4\beta_1$) and VLA-5 ($\alpha_5\beta_1$) (Bazzoni et al., 1996, Fierro et al., 2008). β_1-integrin has been associated with the cell adhesion-mediated-drug resistance (CAM-DR) that causes resistance to anti-leukaemic drugs (Hazlehurst et al., 2000, Tabe et al., 2007). The upregulation of β_1-integrins in CML cells could, therefore, result in their adhesion to protected regions of bone marrow in order to escape chemotherapeutics (Fierro et al., 2008). Furthermore, Rho kinases (ROCKs) are regarded as oncogenes by increasing migration and reducing adhesion (Greenman et al., 2007). It has been recently shown that the inhibition of ROCK in CML cells results in growth arrest and apoptosis. The mechanism of ROCK activation in CML stem cells is thought to involve the activation of upstream RhoA protein by the BCR-ABL-phosphorylated PI3 kinases (Mali et al., 2011).

The above observations are in agreement with the upregulation of integrin and Rho GTPase signalling pathways in LSCs and also their association with CML identity.

3.16.6   Genome-wide gene expression changes are globally regulated.

The results in this Chapter showed that the direction of expression changes, both significant and non-significant, between the studied comparisons was in concordant. In other words, a subset of genes were programmed to be downregulated upon the exit from HSC compartment, whereas another subset of genes were programmed to be upregulated during this transition. Furthermore, the presence of BCR-ABL1 oncoprotein in CML cells did not seem to alter the direction of expression changes. Nevertheless, the gene expression changes were not in agreement in their significance level or their magnitude of change between the studied comparisons. These observations could indicate that a global programming regulates the gene expression changes at the genome-wide level. Therefore, this global programming could control many biological pathways. This was further investigated for the gene expression changes of the HSC identity pathways. The results revealed that the direction of changes in the genes of these pathways was also concordant between the studied comparisons. Therefore, the HSC identity pathways could also be regulated by this global programming. Furthermore, it was observed that LPCs can downregulate the genes in the stem cell identity pathways more efficiently than their normal counterparts. This key finding demonstrated that although these genes are globally regulated, there are potential differences between leukaemic cells and their normal counterparts in implementing this programme.

These observations was in agreement with a recent report that demonstrated the role of epigenetic regulation in global programming of gene expression changes during haemopoiesis development (Cui et al., 2009). This report indicated that the subset of genes that were upregulated upon HSC differentiation were associated with accessible and open chromatin conformation in HSCs. Conversely, the subset of genes that were downregulated in the course of normal differentiation were associated with closed and inaccessible chromatin conformation in HSCs or acquired this chromatin conformation in the differentiated cells. Therefore, the report suggested that the genes that were upregulated or downregulated upon differentiation were epigenetically programmed in HSCs. Thus, the global programming of gene expression differences observed in this Chapter could be linked to such epigenetic programming mechanism.

3.16.7   Further improvements in studying gene expression signatures of LSCs

The source of HSCs in this study was PB instead of BM and therefore, the observations could be biased towards the features of G-CSF mobilised stem cells as opposed to real HSC gene expression signatures. However, the differences between the transcriptomes of BM-derived CD34$^+$ cells and G-CSF mobilised HSCs were studied previously (Steidl et al., 2002). This report showed that CD34$^+$ cells in BM upregulated several cell cycle, DNA replication, and pro-apoptotic genes. Conversely, the mobilised HSCs showed elevated levels of self-renewal and differentiation inhibitor transcription factors such as GATA2. The cell adhesion proteins did not show a significant differential expression between the two populations, except CXCR4 which was highly expressed in the BM population. Furthermore, the BM-derived CD34$^+$ populations showed more homogeneity than the mobilised HSCs (Steidl et al., 2002). Overall, the gene expression signatures of HSCs could vary in response to different environments and the events observed in PB may not fully reflect the signature of the HSCs in the BM niche. However, obtaining sufficient number of HSCs from BM samples is difficult.

The CML and non-CML samples in this study were cultured prior to purification and gene expression analysis (Chapter 2). However, this *in vitro* culture lacks the major components of the BM niche, such as osteoblasts, stromal cells, nerves, and endothelium. These supporting cells in the niche are responsible for secretion of many cytokines, chemokines, and neurotransmitters, which in turn affect the gene expression signatures of the haemopoietic cells in the niche. However, under a non-BM microenvironment culture, all procedures were performed similarly in all cell populations. Therefore, the effect of external cues on their expression profiles should not be reflected in the observed differences we detected between populations.

Overall, the statistical power of the analyses in this Chapter would be significantly increased using more biological replicates obtained from different patients, especially given the heterogeneity observed between LSC samples. This could potentially identify two subsets of LSCs with respect to their similarities to HSCs or progenitor cells.

Functional assays such as CFC and cell division tracking using CFSE dye will be required to monitor the effect of the novel pathways identified in the maintenance of LSCs (Chapter 6). This would further confirm the importance of the findings in this Chapter.

3.17 Conclusion

LSCs share similarities with both HSCs and HPCs at the level of gene expression differences and biological pathways. Therefore, LSCs have the characteristics of both stem cell compartment and the proliferative cells. Furthermore, the exit from stem cell compartment in CML is associated with the loss of HSC identity as well as acquiring further proliferative capacities. Several novel neurotransmitter signalling pathways have also been identified that potentially maintain stem cell characteristics of HSCs and LSCs. The genome-wide gene expression changes are globally regulated upon the exit from HSC compartment in both normal and CML cells. This global mechanism can also regulate the gene expression changes in many biological pathways as demonstrated for the HSC identity pathways. One of the major global mechanisms of gene regulation is epigenetic programming, which have been reported previously to have a key role in fate determination during normal haemopoiesis. Thus, the next Chapter investigated the role of epigenetic programming in the global gene expression regulation in both normal and CML cells.

4. Improvements of a low-cell ChIP method suitable for genome-wide analysis using microarrays or deep sequencing

4.1 Abstract

Chromatin immunoprecipitation in combination with high-throughput readouts is used routinely to study the epigenetic signatures of different cell types. However, the need for large numbers of cells ($>10^6$) per experiment has been a major limitation in studying rare primary cell populations. Since this thesis is focused at studying the epigenetic signatures of rare HSCs and LSCs, developing a low-cell ChIP method that could reproducibly generate genome-wide histone modification maps was a primary aim. This Chapter, therefore, demonstrates empirically validated adjustments to the ChIP procedure that allow it to be used to study histone modifications with as few as 10,000 cells in combination with genome-wide scale platforms including microarrays (ChIP-chip) and high throughput DNA sequencing (ChIP-seq). The low-cell ChIP method for a range of histone modifications was optimised initially in the context of murine embryonic stem cells and was subsequently tested using primary human HSCs. The results presented in this Chapter, demonstrate the successful application of ChIP method to small cell populations which can be robustly and reproducibly used in combination with downstream genome-wide readouts.

4.2 Introduction

Chromatin immunoprecipitation (ChIP) is a method of enriching for the DNA sequences which are associated with proteins in genomes. This method is powerful to identify *in vivo* protein-DNA interactions which are crucial for cellular processes such as DNA replication and repair, chromosomal stability, cell cycle succession, and epigenetic regulation through silencing and activating mechanisms. ChIP has been widely-used to study histones, transcription factors, and other protein binding sites (O'Neill and Turner, 1996, Orlando et al., 1997, Solomon and Varshavsky, 1985). Moreover, ChIP can be coupled to various downstream applications including PCR and qPCR, southern blot analysis (Orlando et al., 1997), western blot analysis (Wells and Farnham, 2002), cloning and sequencing (Loh et al., 2006), microarrays (ChIP-chip) (Hanlon and Lieb,

2004, Weinmann and Farnham, 2002), and high throughput DNA sequencing (ChIP-seq) (Barski et al., 2007). ChIP can be classified into two sub-types: crosslinked ChIP (X-ChIP) (Orlando et al., 1997) or native ChIP (N-ChIP) (O'Neill and Turner, 2003). In X-ChIP, DNA-bound protein complexes are usually reversibly crosslinked by formaldehyde and chromatin is then sheared by sonication to an approximate size range of 200-1000 bp, with an average length of 500 bp. N-ChIP uses unfixed chromatin that can be fragmented mainly by nuclease digestion. For both sub-types, the fragmented chromatin is then incubated with antibodies specifically raised against the protein of interest and antibody-protein-DNA complexes are precipitated by agarose or paramagnetic beads coated with protein A/G followed by a series of washes. After crosslink reversal and digestion of proteins, the associated DNA fragments are purified by phenol-chloroform treatment. The X-ChIP approach is schematically illustrated in Figure 4.1.
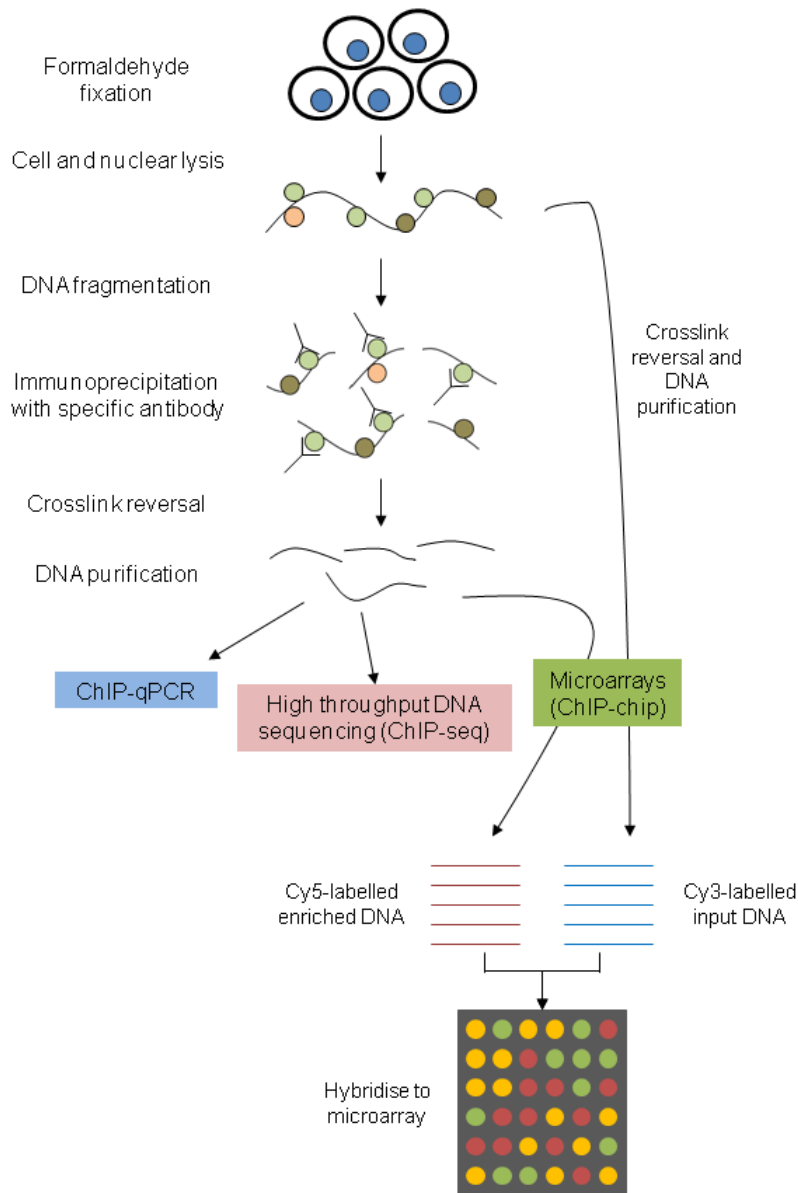
Figure 4.1: ChIP and downstream applications.
Schematic representation of ChIP steps and main downstream analysis approaches (coloured boxes). The concept of ChIP-chip is also illustrated. Red spots are indicative of lower levels of enriched DNA relative to input whereas the green spots show higher abundance of enriched DNA relative to input. Yellow spots are associated with array elements which have equal levels of enriched and input DNA.

ChIP, like many other molecular biology methods, has certain steps in the procedure that require some adjustment depending on the context of the cells under study and/or the protein-DNA interaction of interest. These parameters are briefly discussed below:

Amount of starting chromatin material – one of the major obstacles in ChIP is the requirement for large cell numbers as the starting material (~10^7-10^8 cells, ~200 µg chromatin), which is mainly due to the loss of cells and DNA material during the

ChIP procedure, and overall ChIP inefficiency being relatively low (150 ng ChIP output) (Acevedo et al., 2007). Therefore, this requirement has hampered the study of rare cell populations, such as small primary stem cell populations from embryos and adult stem cells or small tissue biopsies, where the starting number of cells available is quite often less than 10^6 cells. However, there have been some recent improvements to the ChIP procedure which would allow such cell samples to be analysed (see section 3.1.3).

Fixation – formaldehyde is the most commonly used fixation agent due to its ability to create biochemical bonds between molecules in close proximity (~2 angstroms) and fixation is readily reversible (Dedon et al., 1991). The length of formaldehyde treatment should be optimised to ensure an appropriate level of fixation. Histone proteins are known to crosslink faster than non-histone proteins (Das et al., 2004). As mentioned above, long formaldehyde treatment can complicate DNA fragmentation by sonication and mask the recognisable epitopes of the desired protein; therefore, reducing the ChIP yield and possibly biasing the sites that can be assayed. It is also reported that lysine residues are more likely to be fixed by formaldehyde (Orlando et al., 1997), therefore histone modifications involving a lysine residue can be affected in longer treatments.

DNA fragmentation – the strength and duration of sonication are the major factors to be optimised in this step to ensure the fragmentation of chromatin to the desired size range. Nuclease digestion is an alternative approach which is largely applicable in N-ChIP assays. Nuclease digestion is more targeted, though it poses the risk of over-digestion and fragmentation to sub-optimal levels (O'Neill and Turner, 2003). It is also a relatively longer procedure than sonication.

Pre-clearing – conventionally, the fragmented chromatin is pre-cleared prior to IP by using non-specific IgG antibody. Pre-clearing is believed to improve ChIP efficiency by removal of excess proteins, lipids, carbohydrates, and cellular debris present in the cell and nuclear lysate that may interfere with IP by sequestering the antibody or preventing antibody interaction with the right epitope (Becker, 1999). However, in low chromatin concentrations pre-clearing can potentially remove real protein-DNA interactions due to IgG's high affinity for certain epitopes.

Immunoprecipitation (IP) – Antibody quality is a key issue in any ChIP experiment as they determine the signal-to-noise ratio and the yield. Antibody specificity can be tested *in vitro* by western blotting and peptide competition assays for the epitope of interest (Suka et al., 2001). However, most of the antibodies used in ChIP assays are specifically developed to work under ChIP conditions and their specificity should be tested *in vivo* by verifying the genomic regions which are known or expected to be associated with the protein of interest (Spencer et al., 2003). Commercially available ChIP antibodies are mainly affinity purified which will reduce the risk of contamination by the source organism's sera and this eliminates the chance of generating background noise due to cross-reactivity. ChIP antibodies are available as either monoclonal or polyclonal. In the case of crosslinked ChIP, monoclonal antibodies may not be able to interact with the right epitope due to masking by the fixing agent and therefore, polyclonal antibodies are preferred as they are less limiting in their epitope recognition (Das et al., 2004). Furthermore, the optimum amount of antibody for maximum signal-to-noise ratio should be measured empirically through titration with the chromatin extract (Das et al., 2004). Antibody-to-chromatin ratio has to be appropriately adjusted to avoid non-specific interactions that would result in an elevated background noise in microarray and sequencing data.

The antibody complexes are commonly captured by using agarose/Sepharose beads coated with proteins A and/or G. The A or G preference is determined by the class of immunoglobulin and the source organism in which that antibody is raised; for example, protein G has a higher affinity for the IgG class of antibodies from different species (Harlow and Lane, 1999). Agarose beads have a high binding capacity and the beads-to-antibody ratio should be optimised as in the case of low beads-to-antibody ratio, the beads would be become limiting, and many antibody-bound complexes would not be precipitated. Therefore, some important interactions would be lost that result in lower yield and sensitivity. To determine the appropriate levels of protein A/G-coated beads, the chromatin extract-antibody mixture should be incubated with different concentrations of protein A/G solution. Alternatively, magnetic beads have recently been introduced as a substitute to agarose beads and although these are more expensive, they speed up the IP process due to their ease of handling and are suggested to reduce background noise (O'Neill and Turner, 2003).

4.3 Methods

4.3.1 Coupling ChIP to microarray hybridisation (ChIP-chip)

Using microarrays to investigate the nature of ChIP-purified DNA fragments has advanced the ChIP application over the last decade. The initial studies involving ChIP-chip were reported on *S. cerevisiae*'s transcription factor binding sites (Iyer et al., 2001, Lieb et al., 2001, Ren et al., 2000). For ChIP-chip assays, eluted ChIP DNA and non-immunoprecipitated input DNA are fluorescently labelled. After mixing the two samples, they are hybridised onto a microarray covered with oligonucleotide or PCR-amplified probes which may fully or partially cover the genome of interest. If the intensity of ChIP DNA signal exceeds the intensity of input on the array, the genomic region is regarded as a binding site of the protein of interest.

Coverage, density, and resolution are three major parameters in characterising microarrays. Coverage simply refers to the proportion of the genome that is represented on the array. The array density is determined by the number of individual probes, known as features or elements, on the array which are usually PCR-amplified DNA sequences or oligonucleotides of different lengths. The resolution is defined as the length of the features and the length of the intervening genomic regions between them on the array; therefore, a high resolution microarray contains very few gaps between features and the features themselves represent relatively short (approx. 20-50 bp) DNA sequences. Tiling path arrays have the highest resolution as there is little or no genomic spacing between the features on the array. The former is made up of PCR-amplified sequences that are spotted onto a glass slide and due to technical limitations their density cannot exceed 30,000 features (Hudson and Snyder, 2006). Oligonucleotide arrays on the other hand are synthesised *in situ* through synthetic oligonucleotide chemistry (Chee et al., 1996) and therefore, can be found in much higher densities (>1 million features). Thus, they are the ideal choice for studies requiring both high resolution and genome-wide coverage.

The major concerns associated with analysis of microarray platforms include sensitive scanning and analysis, background noise subtraction, standard normalisation, and statistical issues with the data analysis (reviewed in Buck and

Lieb, 2004). Microarray experiments are commonly performed in replicates to distinguish the real biological signals from the background noise. There are several data analysis methods to analyse ChIP-chip data including median percentile rank, the single array error model, and a sliding-window method (Buck and Lieb, 2004).

4.3.2 Coupling ChIP to deep sequencing (ChIP-seq)

ChIP-purified DNA fragments can also be coupled to deep sequencing (ChIP-seq) using the next generation sequencing (NGS) platforms. The major advantages of ChIP-seq over ChIP-chip include (i) base pair resolution, (ii) less expensive for whole genome analysis, (iii) greater genome coverage as the only limitation is the alignability of sequenced tags, (iv) the absence of noise generated as a result of cross-hybridisations or labelling artefacts, (v) larger dynamic range as the enrichment levels are independent of signal intensities which are affected by saturation points, (vi) requirement for significantly lower amounts of starting material (~10 ng); (vii) less amplification for genome-wide analysis which can even be completely avoided in the case of single-molecule sequencing (Park, 2009).

The ChIP-seq library is constructed from the immunoprecipitated DNA fragments that carry two adapter sequences at both ends. The PCR-amplified library is subjected to size selection in the ~200-300 bp range. The sequencing of ChIP DNA is normally performed from the 5′-end and is therefore, known as single-end (SE) sequencing. Nevertheless, it is possible to perform sequencing from both ends which is known as paired-end (PE) sequencing. The PE sequencing is advantageous in mapping the repetitive elements and long-range chromatin interactions (Fullwood and Ruan, 2009). Illumina Genome Analyzer is the most widely used NGS platform for the ChIP-seq studies, although other platforms such as Life Technologies' SOLiD and Helicos BioSciences' HeliScope have also been employed (Table 4.1) (Metzker, 2010, Park, 2009). The Illumina NGS platform was the only NGS technology available for ChIP-seq analysis in this study.

Platform	Library type	Template preparation method	Sequencing chemistry	Read length (bases)	Run time (days)	Advantages	Disadvantages	Application	Cost (US$)
Roche/454's GS FLX Titanium	SE PE	emPCR	Pyro-sequencing	330	8-10 hrs	Longer reads to facilitate mapping of the repetitive regions; short run times	High error rates; expensive reagents	Small genome *de novo* assemblies	500,000
Illumina/Solexa's Genome Analyzer II	SE PE	Solid-phase	Reversible terminators	75 or 100	4 (SE) 9 (PE)	The most widely used platform	Limited multiplexing capacities	Whole-genome sequencing, Seq-based methods	540,000
Life Technologies/ SOLiD 3	SE PE	emPCR	Cleavable probe sequencing by ligation	50	7 (SE) 14 (PE)	Error correction by internal two-base encoding	Lengthy run times	Whole-genome sequencing, Seq-based methods	595,000
Polonator G.007	PE	emPCR	Non-cleavable probe sequencing by ligation	26	5	Cheapest platform; can adapt other sequencing chemistries	Short read lengths; reagents require maintenance and quality control	Bacterial genome sequencing	170,000
Helicos BioSciences HeliScope	SE PE	Single molecule	Reversible terminators	32	8 (SE)	Unbiased representation of templates	High error rates; most expensive platform	Seq-based methods	999,000
Pacific Biosciences/ PacBio RS	SE	Single molecule	Real-time	964	N/A	Can generate >1 kb reads	Highest error rates	*De novo* assemblies, Targeted sequencing, Base modification detection	695,000

Table 4.1: The comparison of major NGS platforms.
Major NGS platforms are listed with their template preparation and sequencing chemistries as well as their advantages and disadvantages and the types of biological applications which they can be employed (Metzker, 2010). emPCR, emulsion PCR.

Illumina NGS platform requires amplified templates to be immobilised to a glass surface, known as the flow-cell. Clonal amplification of templates is necessary since most imaging systems are not designed to detect single fluorescent events. Illumina technology uses a solid-phase amplification approach in which forward and reverse primers are covalently attached to the flow-cell in very high density. After hybridisation of single-stranded template onto the flow-cell, each single-molecule template is extended and forms a bridge with immediately adjacent primers, which are attached to the flow-cell (Bentley et al., 2008, Fedurco et al., 2006). This bridge amplification leads to the formation of clusters which are spatially separated and therefore, allow 100-200 million of sequencing reactions to be performed simultaneously in a flow-cell. Size selection of the library around 200 bp was found to generate clusters more efficiently to obtain an optimum number of clusters on the flow-cell. Illumina sequencing uses a four-colour cyclic reversible termination method which includes fluorescently-labelled nucleotide incorporation, fluorescence imaging, and cleavage (Bentley et al., 2008). Figure 4.2 schematically illustrates the major steps in ChIP-seq library preparation for Illumina NGS platforms.
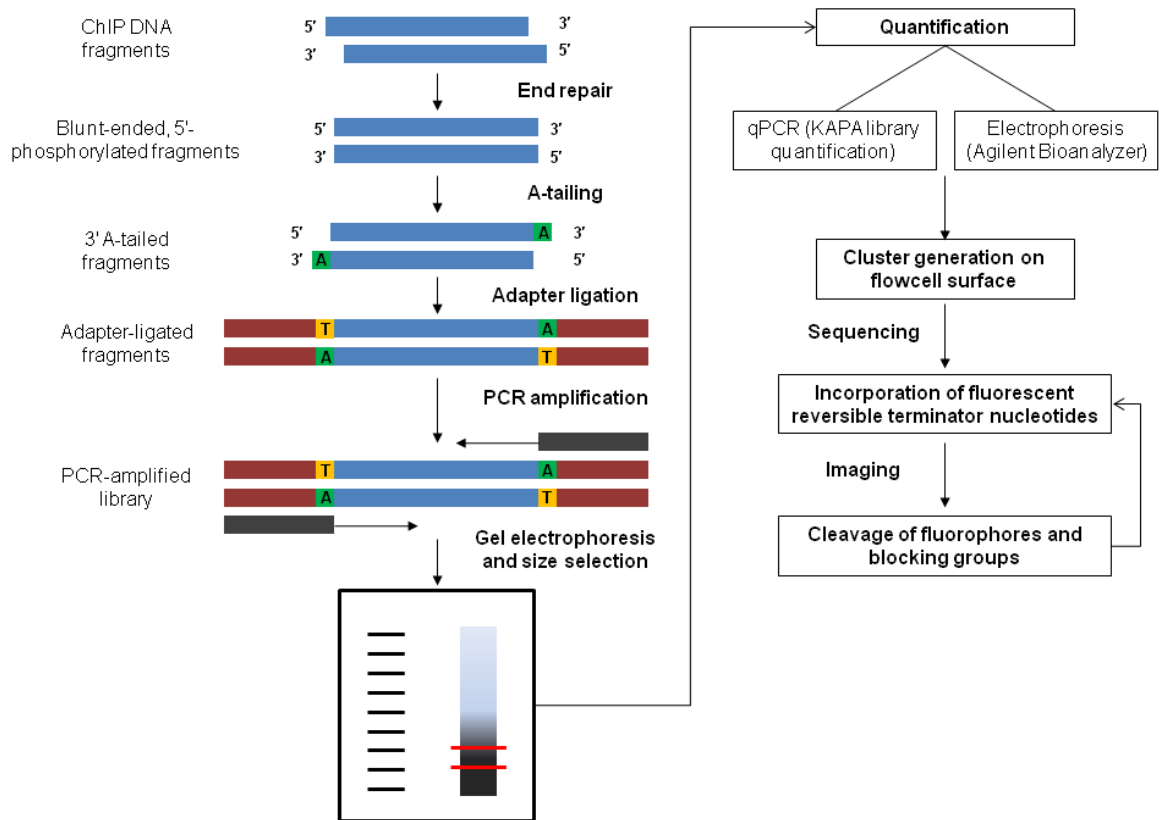
Figure 4.2: ChIP-seq library preparation for Illumina NGS platforms.
Schematic diagram showing the major steps of ChIP-seq library preparation for Illumina Genome Analyzer platform.

The requirement for 10 ng of starting material has made ChIP-seq a preferential method for studying abundant protein-DNA interactions in low cell number samples, as performing genome-wide studies using ChIP-chip demands whole-genome amplifications to generate >2 μg of DNA per array which could also introduce PCR bias (Park, 2009).

4.3.3 ChIP methods developed for study of small numbers of cells

In order to overcome the requirement for large numbers of cells in ChIP several ChIP methods have been improved recently for the detection of DNA-protein interactions or epigenetic features in a starting population of as few as 100 cells. These methods, and their suitability in genome-wide applications, are summarised below (Table 4.2):

1. *Carrier ChIP (CChIP).* This method used a "carrier" chromatin from a distantly related species which was mixed with the small numbers of the cell type of interest to drive the efficiency of the ChIP, facilitate DNA

precipitation and to increase the overall DNA yield. Large numbers of *Drosophila* cells (5×10^7 cells) were mixed with a small number of mammalian cells (100 - 1,000 cells). After mononuclease digestion of the isolated chromatin, histone modifications of interest were immunoprecipitated and the enriched DNA fragments were amplified using radioactively-labelled, species-specific PCR primers followed by phoshphorimaging to visualise the small quantities of enriched DNA (O'Neill et al., 2006). CChIP was used to study the active and repressive histone marks in cells isolated from the mouse blastocyst. The presence of *Drosophila* chromatin in CChIP means that high throughput analysis of the resultant ChIP DNA is not possible due to cross-reactivity and sensitivity issues during microarray analysis, and the abundance of sequence reads from the carrier material in ChIP-seq.

2. *MicroChIP.* MicroChIP was developed using as few as 10,000 cells to study genome-wide histone modifications and RNAPII binding sites in human cells using high-density oligonucleotide arrays (Acevedo et al., 2007). This method was developed by miniaturising the standard ChIP. microChIP did not require the use of foreign DNA as carrier chromatin and, therefore, could be used in ChIP-chip rather than being limited to the analysis of a few genomic loci. microChIP RNAPII ChIP-chip profiles over known gene promoters generated from 10,000 - 100,000 crosslinked cells were demonstrated to be highly reproducible between the microChIP bioreplicates and also could identify 70% of the promoters associated with RNAPII from a standard ChIP assay (10^7 cells) (Acevedo et al., 2007). The major drawback of this method was the requirement for whole-genome amplification of the ChIP DNA prior to microarray hybridisation which could introduce PCR amplification biases. In order to immunoprecipitate antibody complexes, *Staphylococcus aureus* protein-A expressing bacterial cells are used in combination with herring sperm DNA to reduce the risk of non-specific interactions during IP (Acevedo et al., 2007). However, the abundance of sequence reads from the herring sperm DNA would be an issue for ChIP-seq.

3. *Q^2ChIP and μChIP.* This method allowed a rapid precipitation from ChIP using 100,000 cells as starting material to analyse both histone

modifications and transcription factor binding sites (Dahl and Collas, 2007). Similar to the microChIP method, Q^2ChIP also required crosslinking and therefore, transcription factors and non-histone DNA binding proteins could be studied. The usage of paramagnetic beads, which speed up the IP step, and combining different steps of standard ChIP including reversal of crosslinking, protein digestion, and DNA elution into a single shorter (~2 hrs) step, shortened the overall procedure. Reducing the number of steps and limited handling as well as increasing the antibody-to-chromatin ratio enhanced the ChIP efficiency and increased signal-to-noise ratio in ChIP with limited cell numbers. Q^2ChIP was used to determine changes of histone methylations and acetylations in differentiating embryonic carcinoma cells through qPCR screens of developmentally regulated promoters (Dahl and Collas, 2007). Q^2ChIP was improved with slight modifications and reintroduced as µChIP (Dahl and Collas, 2008) which offered an even faster approach (1 day) with fewer cell numbers (100 - 1,000 cells as starting material) to study histone modifications and RNAPII binding sites. Furthermore, µChIP-enriched DNA could be amplified by whole-genome amplification kits in order to perform ChIP-chip assays (Collas, 2011). Although the Q^2ChIP method was demonstrated to be reproducible and sensitive in identifying histone modification enrichments in comparison to a standard ChIP assay, this has yet to be reported for the µChIP method.

4. *MiniChIP.* This method allowed investigation of histone modification enrichments from a range of 50,000 – 150,000 haemopoietic cells by miniaturising the standard protocol and the application of protein A-agarose beads in the presence of salmon sperm DNA (Attema et al., 2007). miniChIP's downstream application was initially limited to qPCR screens; however, recent improvements in the methodology through magnetic IP and whole-genome amplification have enabled miniChIP-chip assays from 10,000 murine HSCs on promoter tiling arrays (Weishaupt and Attema, 2010). miniChIP method has since been used in understanding the histone modification changes in the course of haemopoiesis and it was reported that approximately 40% of bivalent mouse promoters identified by miniChIP-chip in murine HSCs were also bivalent in multipotent human HSCs (CD133[+] CD34[+] CD38[+]) (Weishaupt et al., 2010).

5. *Low-cell ChIP*. It was reported that conventional ChIP with as few as 10,000 cells could be coupled to microarray hybridisation by adjusting the antibody-to-chromatin ratio and protein G-agarose beads concentration for several histone modifications (Dillon, PhD Thesis, 2008). The antibody-to-chromatin ratio was shown to be a key factor in generating reliable ChIP-chip data from 10,000 cells where antibody levels in the range of 0.5 – 5.0 µg produced the highest signal-to-noise ratio compared to the standard ChIP. Furthermore, reducing the concentration of protein G-agarose beads in ChIP with 100,000 cells showed an increase in signal-to-noise ratio. These results were obtained by aliquoting the chromatin that was prepared from 10^8 cells and therefore, the effect of the above adjustments are yet to be confirmed with a starting chromatin preparation that contains low numbers of cells. Furthermore, demonstrating a reliable and reproducible ChIP-chip without the requirement for whole-genome amplification could provide a platform to develop a low-cell ChIP method that can be used for ChIP-seq.

6. *Nano-ChIP*. Nano-ChIP is the latest ChIP method designed for small numbers of cells, in which histone modification maps were generated from as few as 10,000 mES cells and 20,000 mouse HSCs (Adli et al., 2010). Nano-ChIP was developed through miniaturising the conventional ChIP and required protein A-Sepharose beads for IP. Genome-wide histone modification maps was generated by nano-ChIP coupled to Illumina high throughput DNA sequencing (nano-ChIP-seq). The ChIP-seq library preparation was performed by random priming with a universal hairpin primer followed by 10-15 cycles of PCR amplification prior to a second round of PCR amplifications (18 cycles). The nano-ChIP-seq was demonstrated to identify 93% of the promoters marked with H3K4me3 in mES cells by standard ChIP-seq (Adli et al., 2010), despite the whole genome amplification of ChIP DNA which could potentially introduce PCR bias. Nano-ChIP-seq was also used to establish the epigenetic signatures of mouse HSCs and to identify the epigenetic changes involved in the course of HSCs development. The authors also emphasised the importance of antibody titration for small cell ChIP assays as well as establishing their sensitivity and efficiency with regards to the standard ChIP (Adli et al., 2010). Moreover, the application of this method was only demonstrated with

a few histone modifications while its efficiency for RNAPII and transcription factors has yet to be reported.

Method	Downstream applications	Advantages	Disadvantages
CChIP	ChIP-qPCR (O'Neill et al, 2006)	100 cells; histone modifications	Presence of *Drosophila* carrier chromatin; native ChIP only; no transcription factors; radioactively-labelled species specific qPCR primers; no genome-wide application; no reproducibility confirmation
microChIP	ChIP-chip (Acevedo et al, 2007)	10,000 cells; genome-wide application, histone modifications and RNAPII; no foreign carrier DNA; reproducible	Whole genome amplification; presence of deactivated bacterial cells and herring sperm DNA during immunoprecipitation; lengthy procedure; no transcription factors
Q^{2}ChIP	ChIP-qPCR (Dahl and Collas, 2007)	100,000 cells; histone modifications and transcription factors, no foreign carrier DNA; magnetic immunoprecipitation; reproducible; fast	No genome-wide application
µChIP	ChIP-qPCR (Dahl and Collas, 2008) ChIP-chip (Collas, 2011)	100 – 1,000 cells; histone modifications and RNAPII; no foreign carrier DNA; magnetic immunoprecipitation; genome-wide application; very rapid (1 day procedure)	Whole genome amplification; no transcription factors; no reproducibility confirmation
miniChIP	ChIP-qPCR (Attema et al, 2007) ChIP-chip (Weishaupt and Attema, 2010)	10,000 – 50,000 cells; histone modifications; no foreign carrier DNA; magnetic immunoprecipitation; genome-wide application	Whole genome amplification; no transcription factors; ambiguous reproducibility confirmation
Low-cell ChIP	ChIP-qPCR, ChIP-chip (Dillon, 2008)	10,000 – 100,000 cells; histone modifications and transcription factors; no foreign carrier DNA; agarose beads immunoprecipitation; genome-wide application; no whole genome amplification; reproducible	Aliquoting chromatin from 10^8 cells; no primary sample confirmation
nanoChIP	ChIP-seq (Adli et al, 2010)	10,000 – 20,000 cells; histone modifications; no foreign carrier DNA; agarose beads immunoprecipitation; genome-wide application; Illumina sequencing platform; reproducible	PCR amplification prior to ChIP-seq library preparation by using specific hairpin universal primers; no transcription factors

Table 4.2: ChIP methods designed for samples with limited cell numbers.
All the methods published to date with their major downstream applications and their advantages and disadvantages.

4.4 Aims of the Chapter

In order to obtain whole genome epigenetic profiles of both normal and leukaemic stem cell populations (Chapter 5), this Chapter focuses on further developing and improving the low-cell ChIP method developed previously in the laboratory of D. Vetrie (Dillon, PhD thesis, 2008) to address or overcome the following issues:

(i) ChIP from as few as 100,000 cells as starting material (or 10,000 cells/IP),

(ii) applicable to all genome-wide applications including ChIP-chip and ChIP-seq,

(iii) no requirement for whole genome amplification,

(iv) no contaminating source of carrier DNA,

(v) high signal-to-noise ratio,

(vi) reproducible between replicates and comparable to conventional ChIP profiles, and

(vii)   high sensitivity in identifying relevant genomic elements such as promoters.

To these ends, the aims of the work presented in this Chapter are as follows:

1. To establish the parameters required for generating histone modification maps from 10,000 cells using genome-wide applications such as ChIP-chip in the context of mES cells.

2. To assess the sensitivity of the improved low-cell ChIP method coupled to deep sequencing (ChIP-seq) for detecting the histone modification enrichment at the relevant genomic features such as promoters.

Results

4.5 Overall strategy

A number of steps in the ChIP procedure were examined as a means of optimising a low-cell ChIP procedure for looking at histone modification patterns in the mouse genome (Figure 4.3). These included:

1. the amount of antibody per IP condition,

2. the requirement for both a cell lysis and a nuclear lysis step,

3. the effect of chromatin pre-clearing.

Previous work in this laboratory had demonstrated that adjusting antibody levels (Dillon, PhD thesis, 2008) were important to obtaining the highest signal-to-noise ratio. Others had also adopted a low cell ChIP approach by performing a single cell and nuclear lysis step or by removing the pre-clearing step (Dahl and Collas, 2008, Weishaupt and Attema, 2010). All conventional ChIP was performed using 10^7 cells per IP condition derived from a chromatin preparation for 10^8 cells. All low-cell ChIP assays were performed using 10^4 cells per IP condition derived from a chromatin preparation for 10^5 cells.

In order to benchmark the performance of low cell ChIP against ChIP performed in the conventional way, it was necessary to compare results from a cell system which could provide large numbers of cells – aliquots of which could be used for conventional ChIP or for low-cell ChIP. mES cells (E14 clone) were chosen as they can be readily grown under experimental conditions and produce large cell populations in a short period of time. Additionally, their histone modification maps have been generated previously (Mikkelsen et al., 2007).

ChIP-chip was primarily used to assess the optimisation of the low-cell ChIP method. A mouse tiling path microarray was used in the present study to facilitate the development of a low-cell ChIP method. This array was designed to span regions of the mouse genome which corresponded to regions analysed in the human genome in the pilot phase of the ENCODE project (Birney et al., 2007). 6,448 PCR products (average size 1 kb) were spotted in triplicate to produce an array which represented 0.25% of the mouse genome (~8.5 Mb), and covered 12 different genomic regions (Table 4.3). This array used a spotting chemistry which utilised PCR-amplified single-stranded DNA features that were covalently attached to a glass surface through their 5′-aminolink modifications. This spotting chemistry had been previous demonstrated to provide high sensitivity in array-based ChIP assays and other applications (Dhami et al., 2005).

	Genomic coordinates (mm7)	Size (bp)	Biologically important loci
1	chr2: 30,276,337 - 30,671,423	395,086	-
2	chr2: 120,963,553 - 121,342,802	379,249	Stereocilin (Strc)
3	chr2: 155,482,001 - 155,979,495	497,494	Gdf5
4	chr6: 52,016,033 - 52,442,487	426,454	HoxA cluster
5	chr7: 98,867,518 - 100,722,200	1,854,682	β-globin
6	chr7: 138,780,265 - 139,450,798	670,533	Igf2
7	chr9: 45,982,594 - 46,396,529	413,935	Apo cluster
8	chr11: 53,345,522 - 54,347,642	1,002,120	Interleukins
9	chr16: 90,845,544 - 92,107,344	1,261,800	Interferon-α receptors
10	chr17: 45,797,727 - 46,260,264	462,537	Foxp4
11	chr19: 6,136,114 - 6,578,169	442,055	Neurexin 2 (Nrxn2)
12	chrX: 69,239,873 - 70,886,532	1,646,659	IKKγ, Mecp2

Table 4.3: Description of the mouse tiling path array regions.
The genomic coordinates (mm7 build) of 12 regions that are printed in triplicates on mouse tiling path arrays, which make up to approximately 8.5 Mb (~0.25%) of the mouse genome. These regions are orthologs of the biologically important human regions that were chosen for the ENCODE pilot project.

Additionally, the murine comparative studies between conventional and low-cell ChIP were also converted to ChIP-sequencing readouts. This would allow the robustness of low-cell ChIP to be tested independently of array-based methods and in a readout similar to that which would be used to examine histone modification patterns in both normal and leukaemic cells later in this thesis (Chapter 5).

Optimisation of low-cell ChIP was based on assessing the ability of the assays to detect biologically meaningful data in a reproducible way. These assessments were based on the following analyses:

a. *Sensitivity measurements:* Sensitivity was measured in a number of ways. Firstly, the ability of low-cell ChIP to obtain "peaks" or "blocks" of histone modifications across biologically meaningful features, such as gene promoters, or gene bodies, was assessed using the BLOCs algorithm (Pauler et al., 2009), ChIPOTle (Buck et al., 2005), SICER (Zang et al., 2009), or BayesPeak (Spyrou et a.l, 2009). BLOCs algorithm can identify broad local enrichments of histone modifications which are larger than 5 kb in breadth, typically for modifications such as H3K27me3. Additionally, the ChIPOTle and SICER algorithms were also used to identify histone modification peaks. ChIPOTle was designed to identify peaks in ChIP-chip data, whereas SICER and BayesPeak were designed for peaks found in ChIP-seq datasets. Collectively these algorithms were used to compare the "peaks" or "blocks" identified in histone modification patterns when comparing conventional ChIP with low-cell ChIP.

b. *Specificity measurements:* the BLOCs and ChIPOTle analyses were also used to examine the number of novel enrichment "blocks" or "peaks" that were identified by the low-cell ChIP assays. The novel islands could be a result of high background noise levels that were classified as real enrichments (false positives) or the rare events taking place in a subset of cells in the larger population.

c. *Reproducibility measurements:* This was measured between histone modification profiles generated using conventional ChIP and low-cell ChIP; for instance, H3K4me3 conventional ChIP versus H3K4me3 low-cell ChIP; and between different bioreplicates assaying the same histone modification using low-cell ChIP; for instance, between two H3K4me3 low-cell ChIP bioreplicates. Spearman's rank correlation coefficient was used for this analysis. These coefficients range from between -1 to +1. Values which approach +1 between two samples show a high positive correlation.
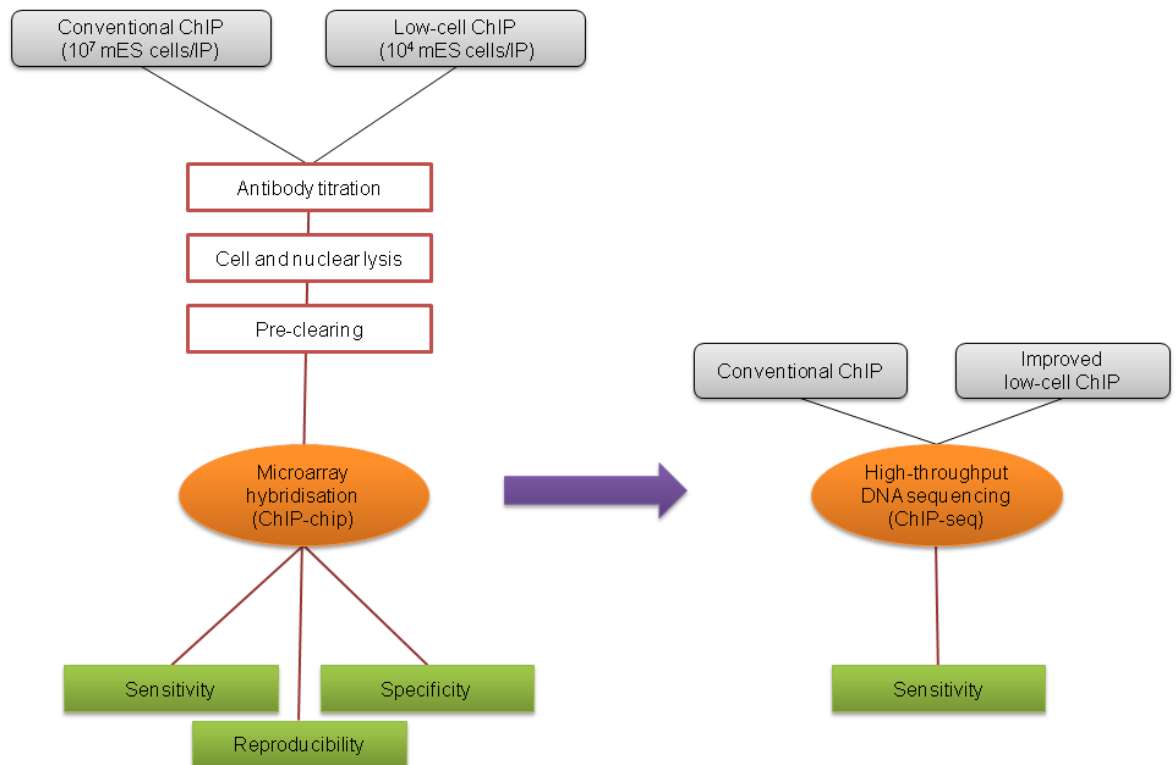
Figure 4.3: Overall strategy in improving the low-cell ChIP method.
Low-cell ChIP-chip was performed with 10,000 cells/IP and compared against the ChIP-chip profiles from the standard ChIP using 10^7 cells/IP. The ChIP steps boxed in red were modified or adjusted during optimisation. The sensitivity, specificity and reproducibility of low-cell ChIP were measured upon applying these adjustments. After optimising the low-cell ChIP on microarrays, the method was coupled to deep sequencing (ChIP-seq) and the sensitivity of the genome-wide approach was investigated.

4.6 Quality control assessment of ChIP material

Each ChIP assay performed as part of the study described in this Chapter was subjected to several quality control (QC) assessments prior to ChIP-chip or ChIP-seq analyses. These included (i) gel electrophoresis of purified ChIP DNAs, (ii) ChIP-qPCR to demonstrate that the ChIP material contained histone modification enrichments for genomic loci used as positive controls, (iii) gel electrophoresis of purified Cy-3 labelled ChIP prior to hybridisation onto microarrays along with Cy-5 labelled input DNA. Examples of these QC assessments are shown in Figures 4.4 and 4.5.

Typically, as shown in Figure 4.4, DNA smears for ChIP DNA purified from conventional ChIP were visible when subjected to gel electrophoresis. However, DNA smears from low-cell ChIP purified DNAs were not visible. Nevertheless, ChIP-qPCR analysis of gene promoters showed that there was enrichment for the histone modification of interest in DNAs derived from either conventional ChIP or

low-cell ChIP. The low-cell ChIP showed a similar pattern of enrichment at these loci, although the fold enrichment levels were lower relative to the conventional ChIP assay. Therefore, although the DNA smear could not be detected by gel electrophoresis, it could be shown at the level of qPCR that the IP had worked in low-cell ChIP, albeit at a lower efficiency.
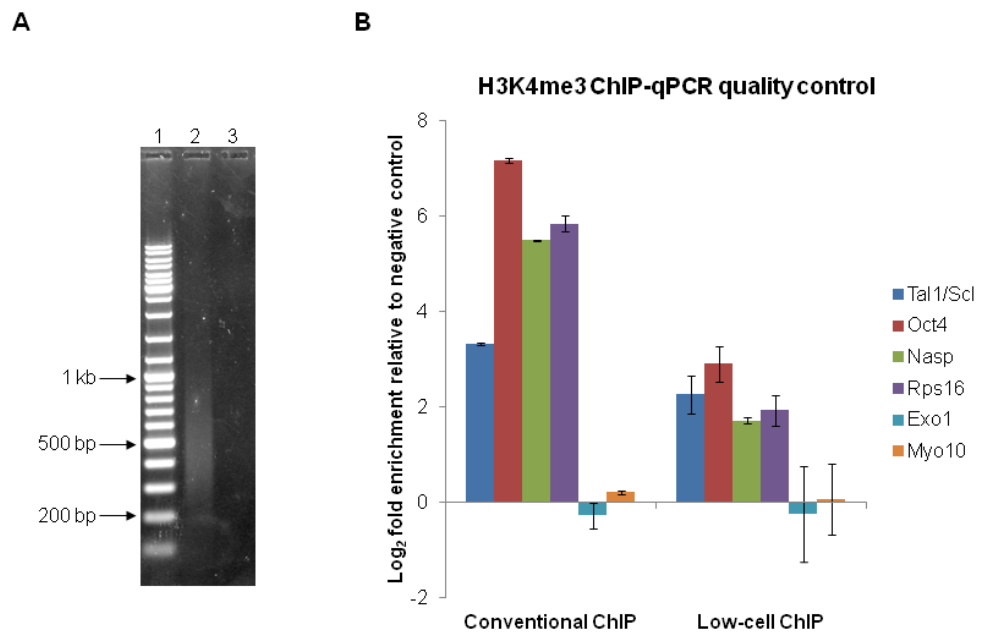
Figure 4.4: Gel electrophoresis of purified ChIP DNA and qPCR screens for quality control.
(A) Agarose gel electrophoresis of purified conventional and low-cell ChIP DNA. Lane 1 = 1 kb DNA marker ladder; lane 2 = H3K4me3 conventional ChIP, lane 3 = H3K4me3 low-cell ChIP. DNA smear is visible in conventional ChIP (lane 2) ranging from 200 bp to 1 kb which shows the efficiency of the sonication regime. The electrophoresis was performed on 1% agarose 1x TBE gels and visualised with SYBR Safe DNA gel stain (Invitrogen). (B) ChIP-qPCR screens of the genomic loci known to be associated with the histone modification of interest for low-cell ChIP quality control. The H3K4me3 conventional and low-cell ChIP-qPCR results are shown here with four positive control designed at the promoters of active genes, including Tal1/Scl, Oct4, Nasp, and Rps16; and two negative control primer pairs designed at the promoters of repressed genes, including Exo1 and Myo10. The height of the bars shows the Log_2 fold enrichment relative to the negative control enrichments. The error bars represent 95% confidence intervals.

Despite the relatively low recovery of DNA from low-cell ChIP (Figure 4.4), labelled DNA fragments were observed as a smear when subjected to agarose gel electrophoresis. These smears showed similar size distributions as labelled material derived from conventional ChIP (Figure 4.5), although there did appear to be less material in labelled samples obtained from low-cell ChIP (see lane 3 in Figure 4.5).
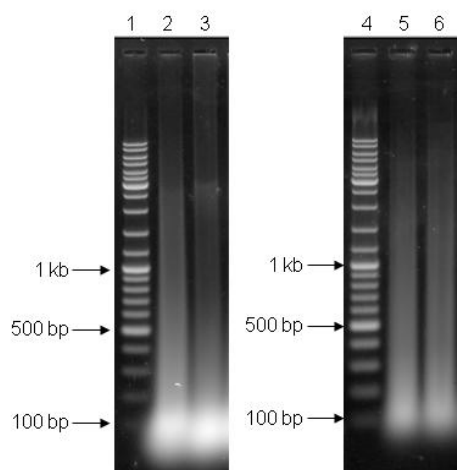
Figure 4.5: Gel electrophoresis of fluorescently labelled ChIP and input DNA.
Left: Gel electrophoresis of Cy3-labelled ChIP DNA samples. Lane 1 = 1 kb DNA marker ladder; lane 2 = H3K27me3 conventional ChIP; lane 3 = H3K27me3 low-cell ChIP. Labelling reactions have worked with similar efficiency in both conventional and low-cell ChIP assays, as DNA smears are visible in both cases (lanes 2 and 3). Majority of labelled DNA can be seen in the size range of 85-150 bp which is an outcome of random priming amplification. Right: Gel electrophoresis of Cy5-labelled input DNA. Lane 4 = 1 kb DNA marker ladder; lane 5 = input DNA for conventional ChIP; lane 6 = input DNA for low-cell ChIP. Input DNA is required as a reference in ChIP-chip assays which is directly purified from an unenriched sonicated chromatin source that is used for the ChIP assays. The labelling has worked with similar efficiency to the ChIP samples. The brighter intensity of DNA smears in ChIP samples (left) is due to the biochemical reaction between SYBR Safe gel stain and the Cy3 dyes. The electrophoresis was performed on 1% agarose 1x TBE gels and visualised with SYBR Safe DNA gel stain.

4.6.1 Examining the effect of antibody levels on low-cell ChIP

4.6.1.1 Titration of the anti-H3K27me3 antibody in low-cell ChIP

As previously reported, titrating antibody levels is a key factor in improving signal-to-noise levels and increasing the ChIP efficiency (Dillon, PhD thesis, 2008). It had been reported for H3K4me3 and H3 acetylation that antibody levels between 0.5 and 5.0 µg per IP produce the highest signal-to-noise ratio on average (Dillon, PhD thesis, 2008). Therefore, a serial dilution of antibodies at 5.0, 1.0, and 0.5µg was performed with low-cell ChIP using 10,000 cells/IP. These were benchmarked against conventional ChIP which was performed using an optimised 10.0 µg of antibody per IP (Dhami et al., 2010, Dillon, 2008). In every other respect, the low-cell ChIP antibody titration experiments were performed under the same conditions as the conventional ChIP with changes in antibody levels as the only variable.

The histone modification H3K27me3 is associated with repressed loci. This histone mark appears as either enrichment peaks at the promoters of repressed or

poised genes (Bernstein et al., 2006, Bracken et al., 2006), or large blocks of enrichments covering between 10 – 140 kb spanning entire genes or gene clusters as observed in the case of mammalian HOX gene clusters (Ringrose, 2007). The anti-H3K27me3 antibody used here was a polyclonal antibody raised in rabbit which was reported previously in genome-wide ChIP applications (Pan et al., 2007, Zhao et al., 2007). The antibody titration was performed with low-cell ChIP assays and these were used in ChIP-chip assays with the mouse tiling array (described in section 4.3) as the readout. Figure 4.6 shows ChIP-chip histone modification profiles from four different regions of the mouse genome that were chosen to represent the overall variation in histone modification enrichments that were observed in these experiments.
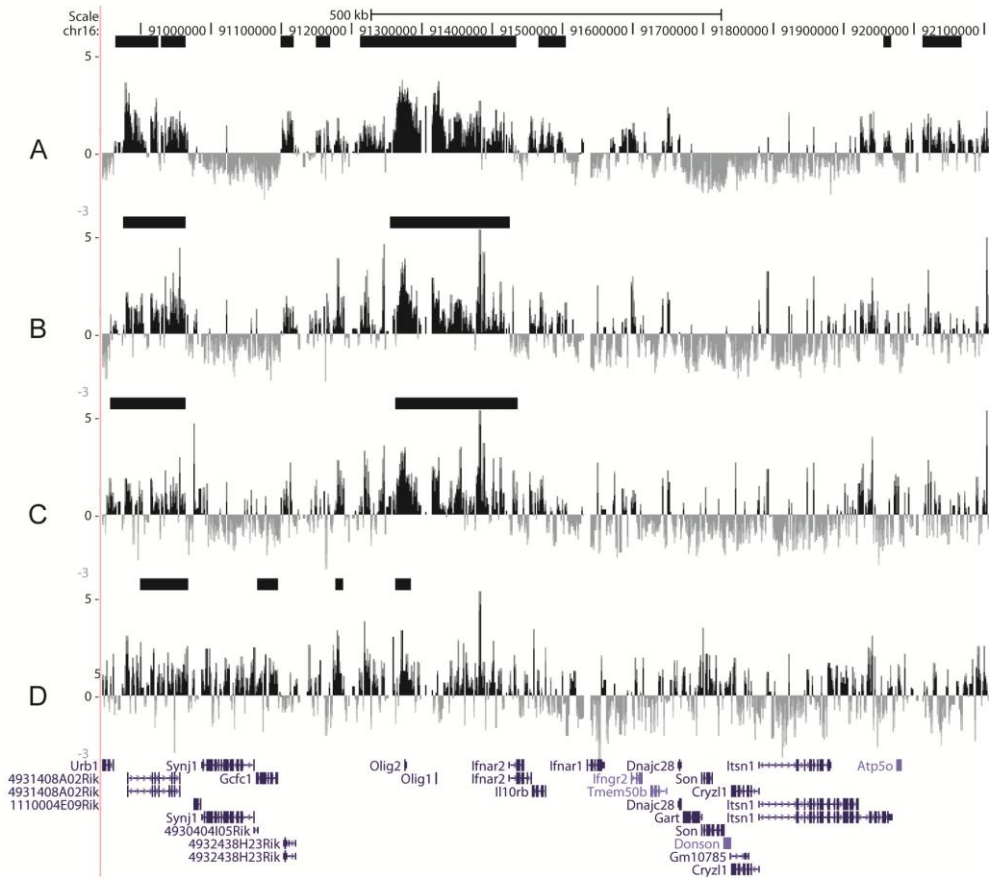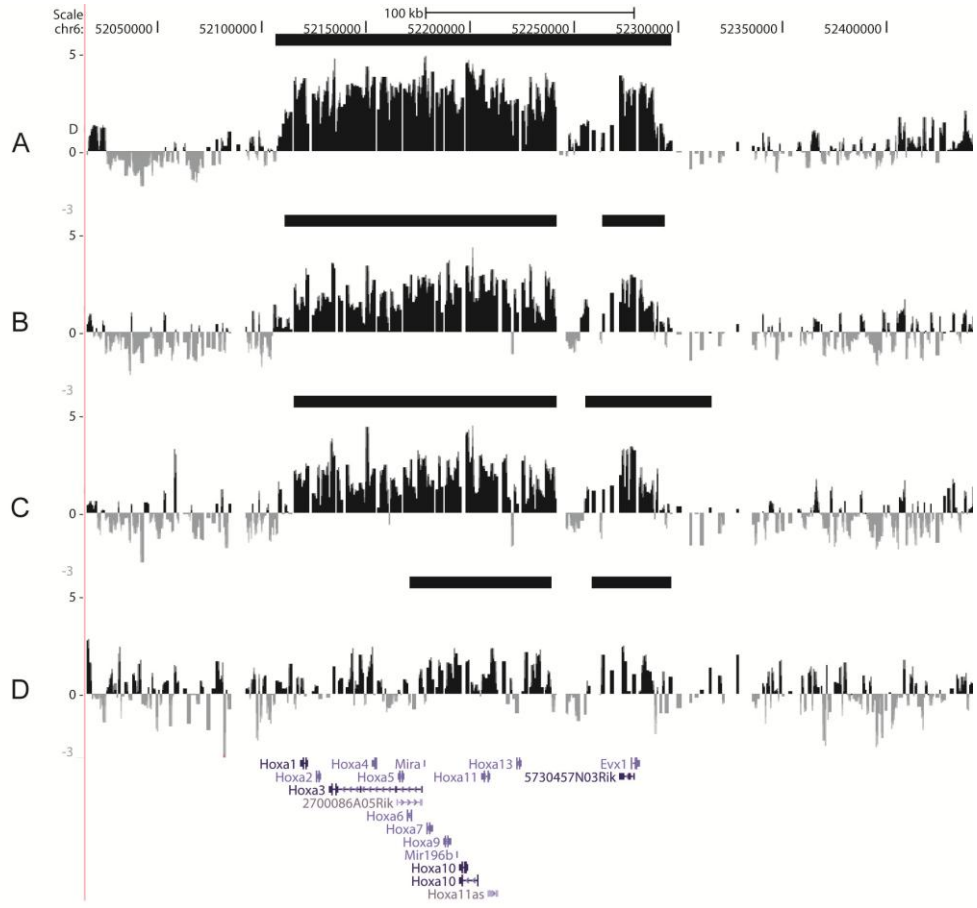
Figure 4.6: Anti-H3K27me3 antibody titration.
Screenshot views of the UCSC genome browser (Kuhn et al., 2007) of four different regions represented on the mouse arrays. For each region the ChIP-chip profiles are shown in the following

order: (A) 10.0 µg H3K27me3 conventional ChIP; (B) 5.0 µg; (C) 1.0 µg and (D) 0.5 µg H3K27me3 low-cell ChIP. The black horizontal bars on top of each track show the significantly enriched domains identified by the BLOCs algorithm. The bottom track shows the RefSeq genes (Pruitt et al., 2007) with transcriptional orientation indicated by arrows. The ChIP-chip data is displayed in the four intervening tracks as the ratio of ChIP sample fluorescence to input DNA fluorescence. Each black vertical bar is the enrichment measured at a single array element with the relative $\log_2$ fold enrichments represented at the height of the bar. The scale and the chromosome coordinates are demonstrated on top of each panel.

When compared against the results of conventional ChIP across all 12 murine regions represented on the tiling path array, low-cell H3K27me3 ChIP performed with 5.0 µg of antibody showed the highest Spearman's correlation coefficient (0.74). 1.0 µg and 0.5 µg of antibody showed coefficients of 0.66 and 0.07 respectively. Using the BLOCs algorithm, 36 H3K27me3 "blocks" or "islands" were identified significantly above the background in conventional ChIP-chip from which, 19 islands were also recognised in the 5.0 µg low-cell ChIP condition. However, low-cell ChIP with 1.0 µg of antibody could identify 16 of these 36 islands, whereas 0.5 µg could distinguish only 5 of 36. On the other hand, 0.5 µg and 1.0 µg of antibody identified 18 and eight novel islands, respectively, whereas 5.0 µg of antibody located only three novel domains. The novel islands could be a result of high background noise levels that were classified as real enrichments (false positives) or the rare events taking place in a subset of cells in the larger population. All of this data, when taken together, suggested that the low-cell ChIP-chip data derived from IPs which used 5.0 µg of anti-H3K27me3 antibody showed the highest degree of similarity to the results of conventional ChIP. This provided a first line of evidence that H3K27me3 antibody levels used in IPs were a key determinant of IP efficiency in low-cell ChIP.

4.6.1.2 Titration of the anti-H3K4me3 antibody in low-cell ChIP

The histone modification H3K4me3 is prevalent in open chromatin regions and the promoters of active or poised genes (Barski et al., 2007, Bernstein et al., 2006). The anti-H3K4me3 antibody used here was a polyclonal antibody raised in rabbits which had been used extensively in genome-wide studies (Bernstein et al., 2006, Santos-Rosa et al., 2002). Similar to those experiments described above, the antibody levels of the anti-H3K4me3 antibody was titrated in low-cell ChIP experiments and the resultant ChIP material hybridised to murine tiling path arrays. The conventional and low-cell ChIP-chip profiles that were derived across four different regions of the mouse genome are shown in Figure 4.7.
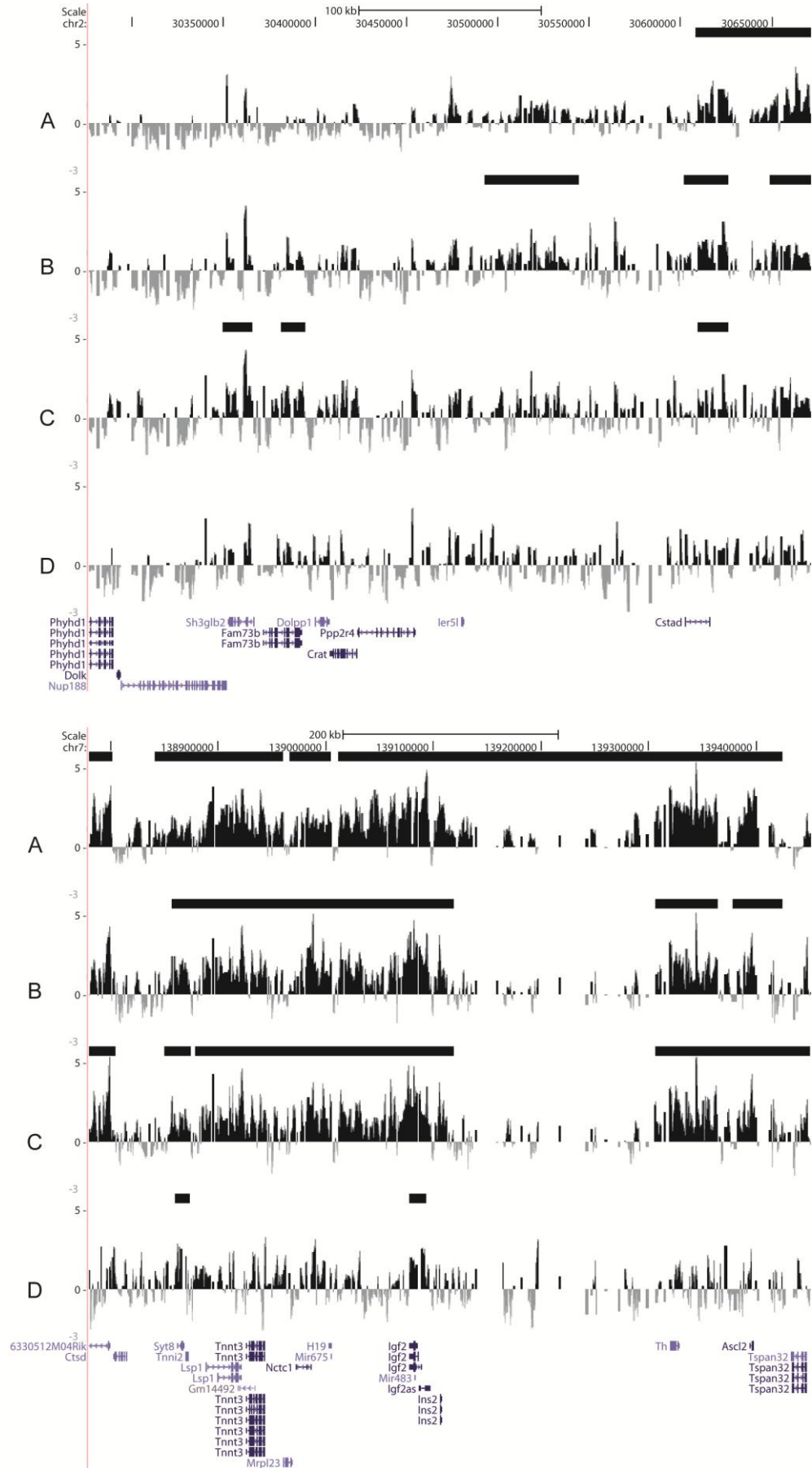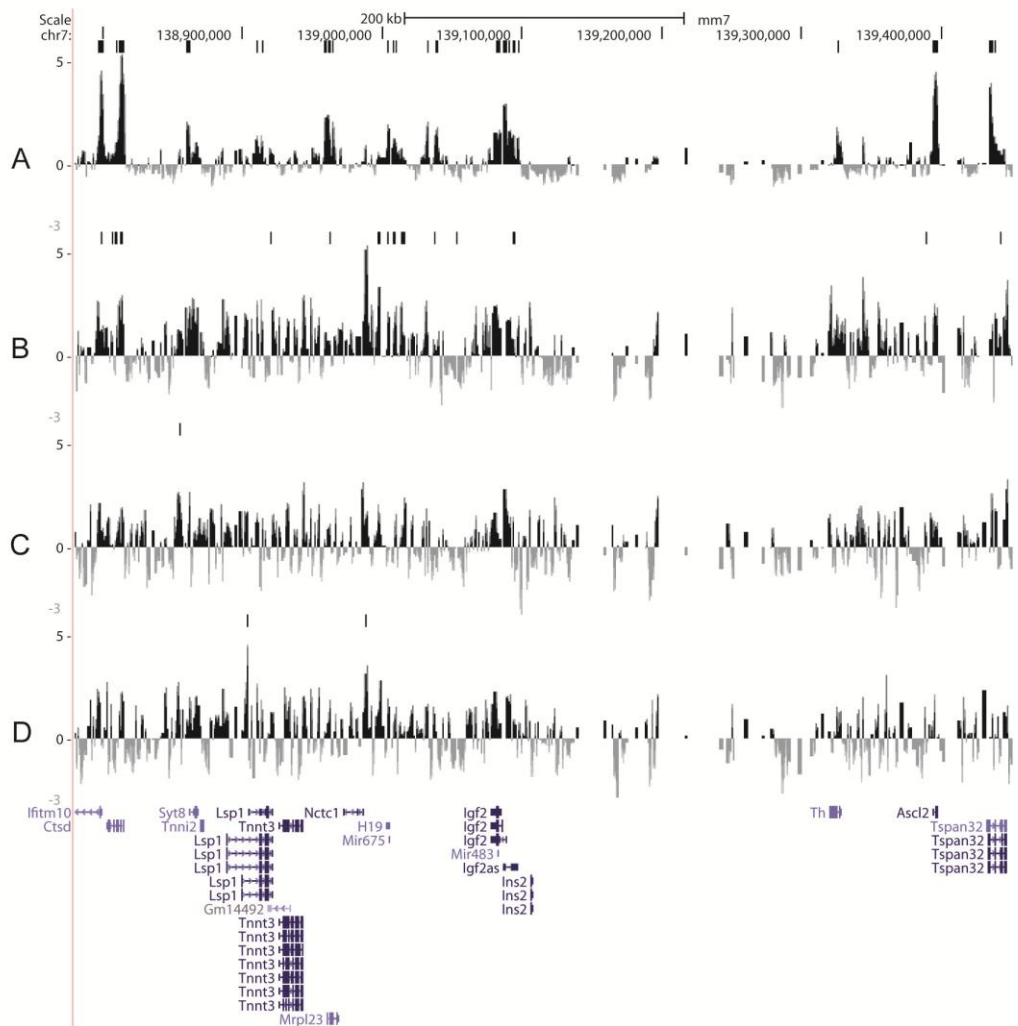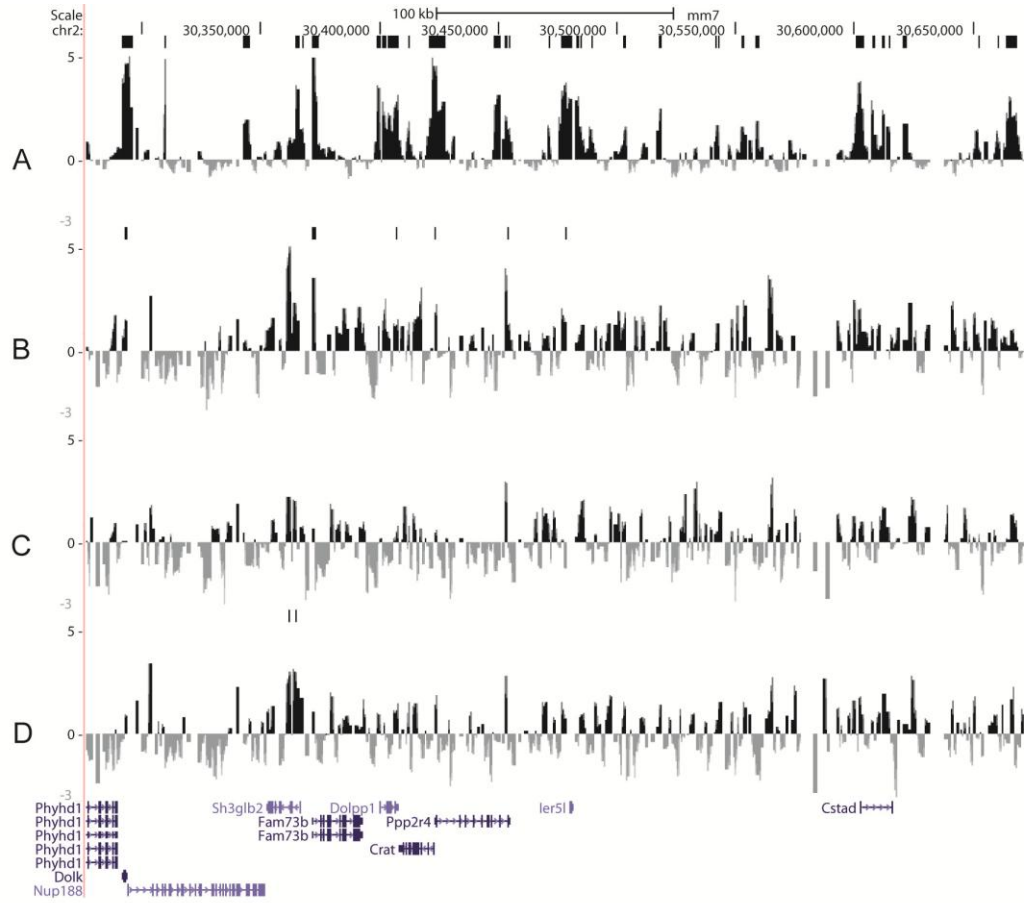
Figure 4.7: Anti-H3K4me3 antibody titration.
Screenshot views of the UCSC genome browser (Kuhn et al., 2007) of four different regions represented on the mouse arrays. For each region the ChIP-chip profiles are shown in the following order: (A) 10.0 μg H3K4me3 conventional ChIP; (B) 5.0 μg; (C) 1.0 μg; and (D) 0.5 μg H3K4me3 low-cell ChIP. The black bars on top of each track show the significantly enriched peaks identified by ChIPOTle. The bottom track shows the RefSeq genes (Pruitt et al., 2007) with transcriptional orientation indicated by arrows. The ChIP-chip data is displayed in the four intervening tracks as the ratio of ChIP sample fluorescence to input DNA fluorescence. Each black vertical bar is the enrichment measured at a single array element with the relative $\log_2$ fold enrichments represented at the height of the bar. The scale and the chromosome coordinates are demonstrated on top of each panel.

When compared against the results of conventional ChIP across all 12 murine regions represented on the tiling path array, low-cell H3K4me3 ChIP performed with 5.0 μg of antibody showed the highest Spearman's correlation coefficient (0.35). 1.0 μg and 0.5 μg of antibody showed coefficients of 0.22 and 0.26 respectively. Using the ChIPOTle analysis, 462 H3K4me3 enrichment "peaks" were identified significantly above the background in conventional ChIP-chip from which, 64 peaks were also recognised in the 5.0 μg low-cell ChIP condition. However, low-cell ChIP with 1.0 μg of antibody could identify 13 of these 462 peaks, whereas 0.5 μg could distinguish only 10 of 462. On the other hand, 0.5 μg and 1.0 μg of antibody identified 28 and 35 novel peaks, respectively. However, 5.0 μg of antibody located 43 novel peaks. The novel peaks could be a result of high background noise levels that were classified as real enrichments (false positives) or the rare events taking place in a subset of cells in the larger population. All of this data, when taken together, suggested that the low-cell ChIP-chip data derived from IPs which used 5.0 μg of anti-H3K4me3 antibody showed the highest degree of similarity to the results of conventional ChIP. This provided a first line of evidence that H3K4me3 antibody levels used in IPs were a key determinant of IP efficiency in low-cell ChIP.

Overall, the antibody titration was proved to be important in improving the low-cell ChIP sensitivity where 5.0 μg of anti-H3K4me3 and anti-H3K27me3 antibodies were shown to reproduce the conventional ChIP better than the other concentrations within the titration range.

4.6.2 Examining the effect of simultaneous cell and nuclear lysis on low-cell ChIP

The conventional ChIP protocol is comprised of an initial cell lysis step, which disrupts the cell membrane while maintaining the cell nuclei intact. Once the nuclei

are isolated, a second lysis step is then used to disrupt the nuclear membrane. This two-step approach facilitates the removal of proteins, lipids, and carbohydrates present in the cytoplasm from the intact nuclei – thus removing potential non-specific targets which could bind antibody during IP and thus affecting ChIP efficiency. However, simultaneous cell and nuclear lysis was employed in various low-cell ChIP methods with small cell numbers in order to minimise handling steps during ChIP and thus maximising the chromatin recovery and yield (Acevedo et al., 2007, Dahl and Collas, 2008, Weishaupt and Attema, 2010).

Therefore, the effect of simultaneous cell and nuclear lysis was tested under the low-cell ChIP conditions and was compared to conventional ChIP, which employed a two-step lysis, to determine the improvement in the efficiency of the assays. Low-cell ChIP was performed with anti-H3K4me3 and anti-H3K27me3 antibodies at 5.0 µg. The only variable between the low-cell ChIP conditions were the cell and nuclear lysis step that was performed either in two sequential steps or simultaneously. The resultant low-cell ChIP material was hybridised onto the murine tiling path array and the results compared to those obtained for conventional ChIP for these two histone modifications (Figures 4.8 and 4.9).
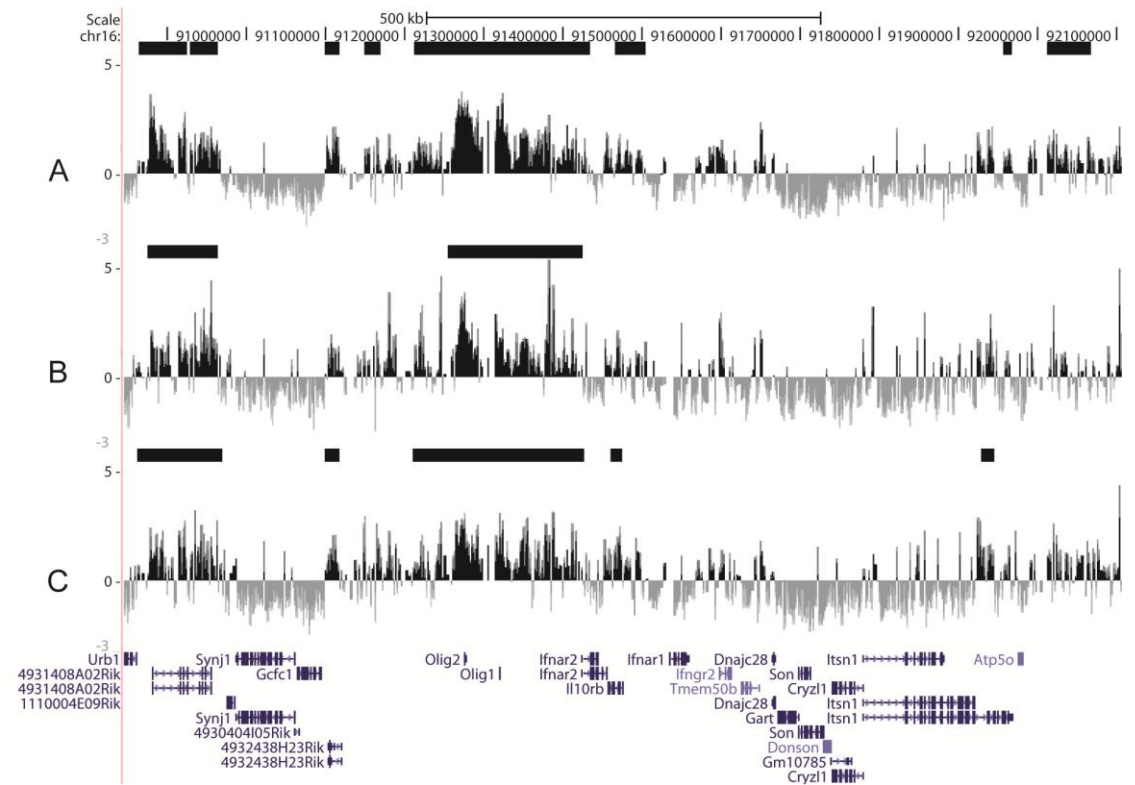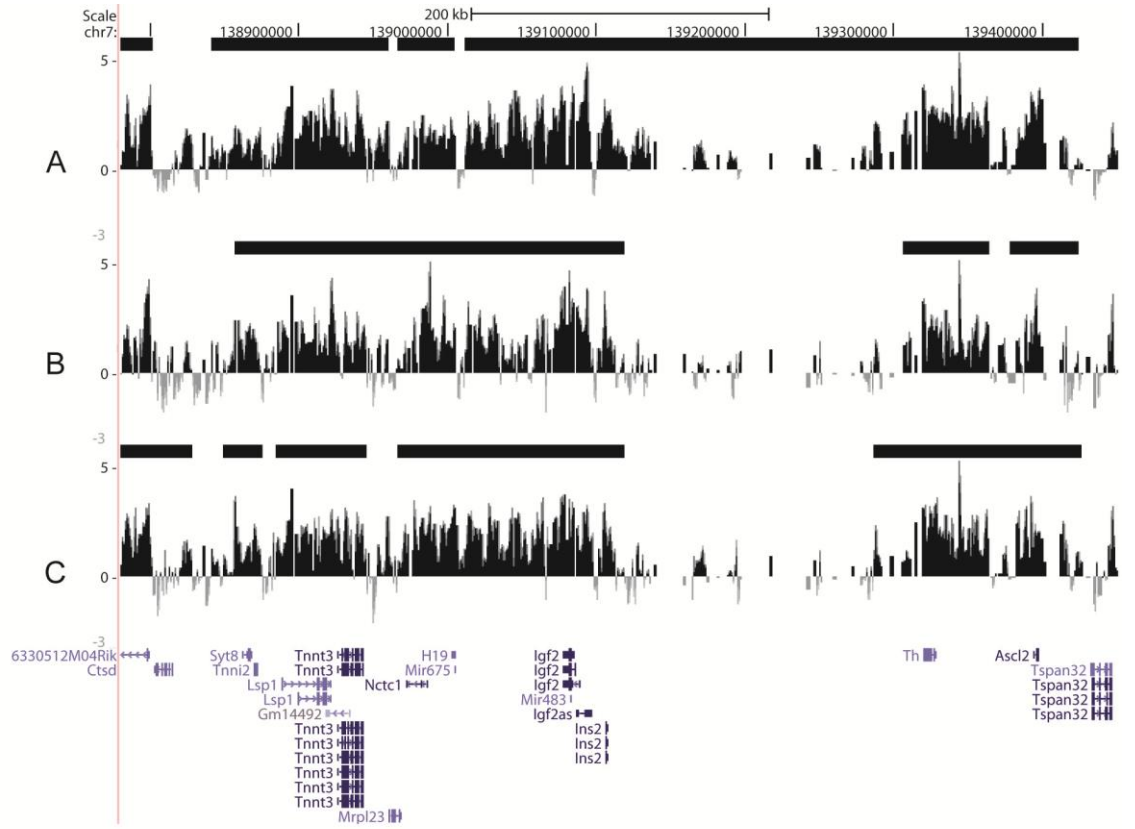
Figure 4.8: The effect of simultaneous cell and nuclear lysis on low-cell H3K27me3 ChIP.
Screenshot views of the UCSC genome browser (Kuhn et al., 2007) of four different regions represented on the mouse arrays. For each region the ChIP-chip profiles are shown in the following order: (A) 10.0 µg H3K27me3 conventional ChIP; (B) two-step lysis; and (C) simultaneous lysis H3K27me3 low-cell ChIP. The black horizontal bars on top of each track show the significantly enriched domains identified by the BLOCs algorithm. The bottom track shows the RefSeq genes (Pruitt et al., 2007) with transcriptional orientation indicated by arrows. The ChIP-chip data is displayed in the three intervening tracks as the ratio of ChIP sample fluorescence to input DNA fluorescence. Each black vertical bar is the enrichment measured at a single array element with the relative $\log_2$ fold enrichments represented at the height of the bar. The scale and the chromosome coordinates are represented on top of each panel.
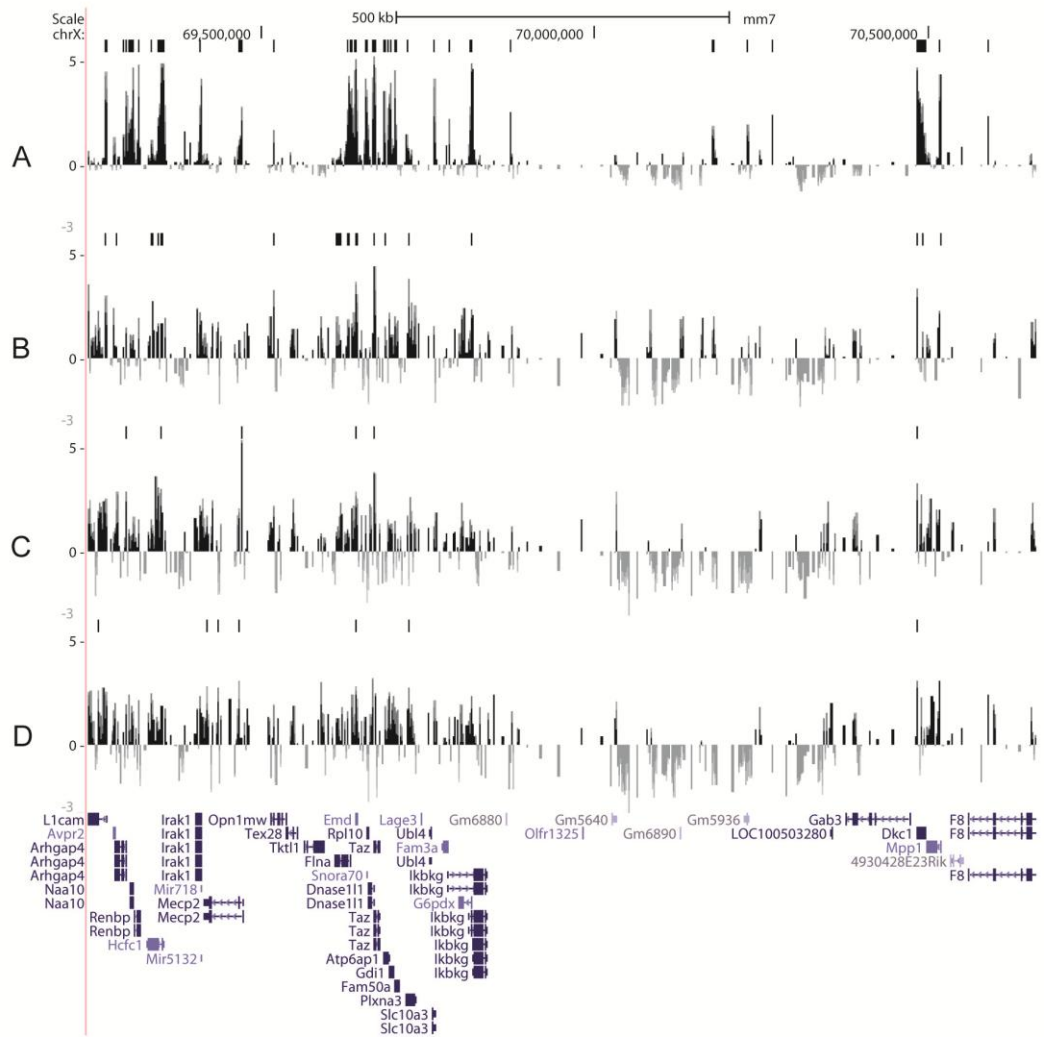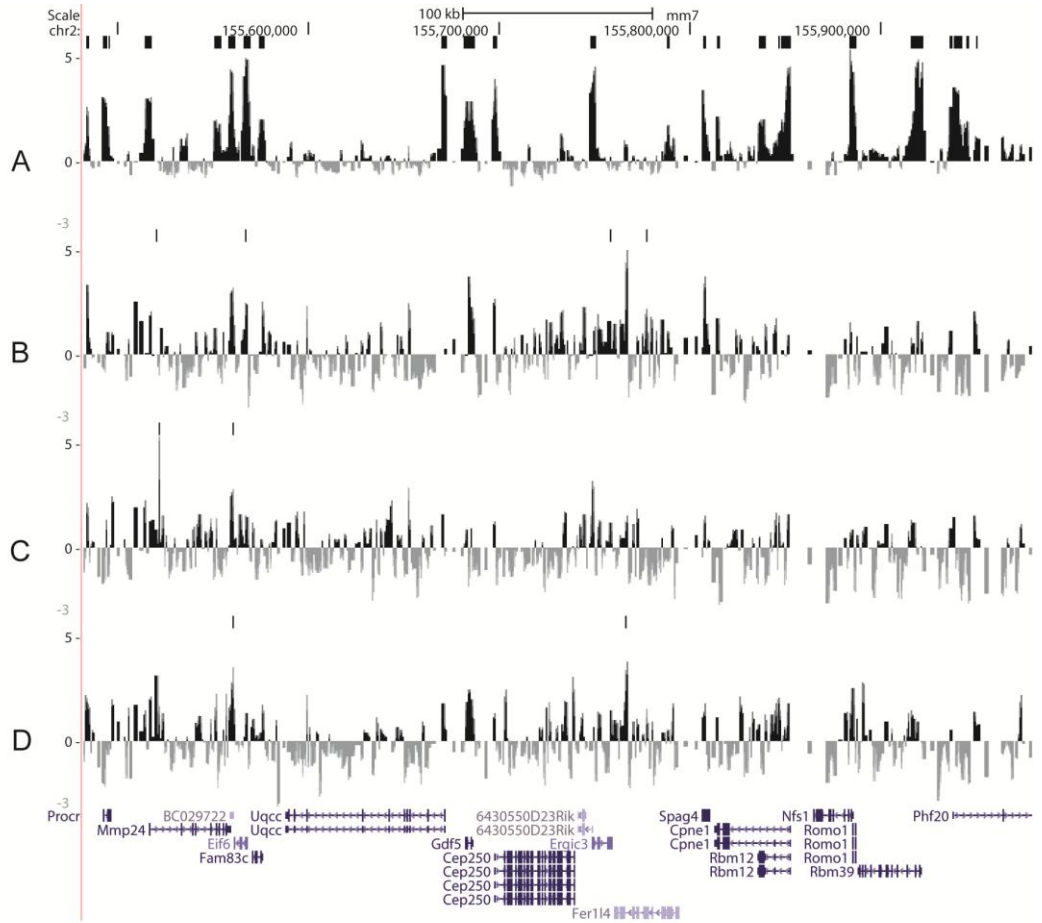
Figure 4.9: effect of simultaneous cell and nuclear lysis on low-cell H3K4me3 ChIP.
Screenshot views of the UCSC genome browser (Kuhn et al., 2007) of four different regions represented on the mouse arrays. For each region the ChIP-chip profiles are shown in the following order: (A) 10.0 μg H3K4me3 conventional ChIP; (B) two-step lysis; and (C) simultaneous lysis H3K4me3 low-cell ChIP. The black bars on top of each track show the significantly enriched peaks identified by ChIPOTle. The bottom track shows the RefSeq genes (Pruitt et al., 2007) with transcriptional orientation indicated by arrows. The ChIP-chip data is displayed in the three intervening tracks as the ratio of ChIP sample fluorescence to input DNA fluorescence. Each black vertical bar is the enrichment measured at a single array element with the relative $\log_2$ fold enrichments represented at the height of the bar. The scale and the chromosome coordinates are demonstrated on top of each panel.

When compared against the results of conventional ChIP across all 12 murine regions represented on the tiling path array, low-cell H3K27me3 ChIP performed with simultaneous cell and nuclear lysis showed higher Spearman's correlation coefficient (0.82) than the two-step lysis (0.74) (Figure 4.8). Using the BLOCs algorithm, simultaneous cell and nuclear lysis low-cell H3K27me3 ChIP could identify 29 of the 36 islands identified in the conventional assay. Furthermore, low-cell H3K27me3 ChIP with simultaneous cell and nuclear lysis identified four novel domains. The novel peaks could be a result of high background noise levels that were classified as real enrichments (false positives) or the rare events taking place in a subset of cells in the larger population. All of this data, when taken together, suggested that the H3K27me3 low-cell ChIP-chip data derived from IPs with simultaneous cell and nuclear lysis showed higher degree of similarity to the results of conventional ChIP and improved sensitivity in comparison to the two-step lysis approach.

Moreover, low-cell H3K4me3 ChIP performed with simultaneous cell and nuclear lysis showed higher Spearman's correlation coefficient (0.37) than the two-step lysis (0.35) in comparison to the conventional ChIP assay across all 12 murine regions represented on the tiling path array (Figure 4.9). Simultaneous cell and nuclear lysis low-cell H3K4me3 ChIP could identify 31 of the 462 peaks identified in the conventional assay as determined by ChIPOTle. Furthermore, low-cell H3K4me3 ChIP with simultaneous cell and nuclear lysis identified 37 novel peaks. The novel peaks could be a result of high background noise levels that were classified as real enrichments (false positives) or the rare events taking place in a subset of cells in the larger population. All of this data, when taken together, suggested that the H3K4me3 low-cell ChIP-chip data derived from IPs with simultaneous cell and nuclear lysis showed slightly higher degree of similarity to the results of conventional ChIP. Although the sensitivity of low-cell H3K4me3 ChIP was not improved using simultaneous cell and nuclear lysis, its specificity was increased in comparison to the two-step lysis approach.

Overall, although simultaneous cell and nuclear lysis was not detrimental to low-cell ChIP, it minimised the handling of cells which may improve yields. Therefore, simultaneous cell and nuclear lysis would present an improvement in terms of simplifying the method.

4.6.3 Analysing the effect of pre-clearing on low-cell ChIP

The pre-clearing step is conventionally performed prior to IP step to increase the ChIP efficiency, as described earlier (section 4.1.1). However, some ChIP protocols suggested removing this step when dealing with small cell numbers to minimise the loss of DNA-protein complexes as a result of non-specific IgG interaction in low chromatin concentration conditions (Dahl and Collas, 2008, Weishaupt and Attema, 2010).

In order to test the effect of pre-clearing on low-cell ChIP, pre-cleared (PC) and non-pre-cleared (NPC) conditions were examined and the low-cell ChIP assays were performed with 5.0 µg of antibody and simultaneous cell and nuclear lysis. Therefore, the only variable in these assays was the pre-clearing step. The resultant low-cell ChIP material was hybridised onto the murine tiling path array and the results compared to those obtained for conventional ChIP for these two histone modifications (Figures 4.10 and 4.11).
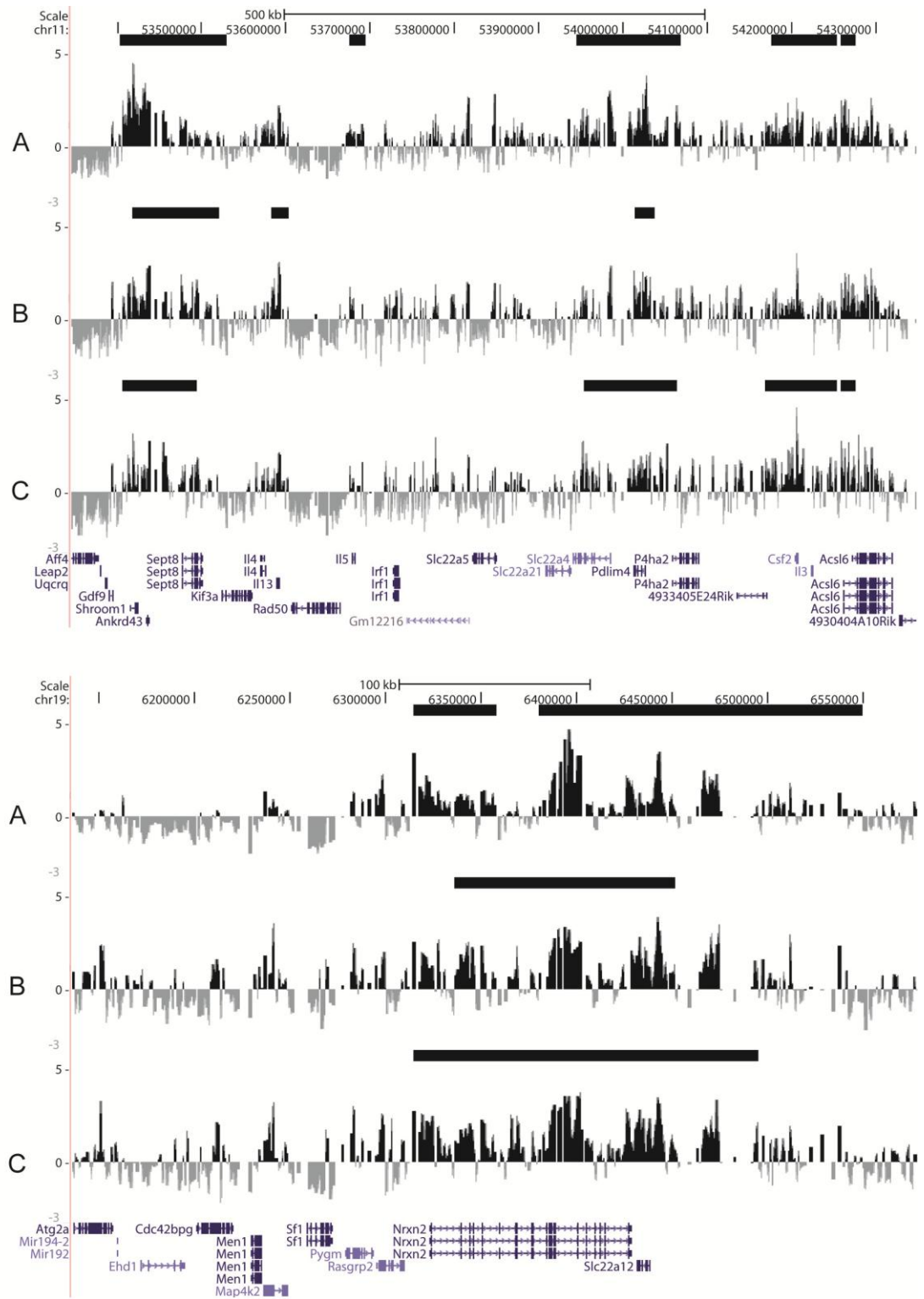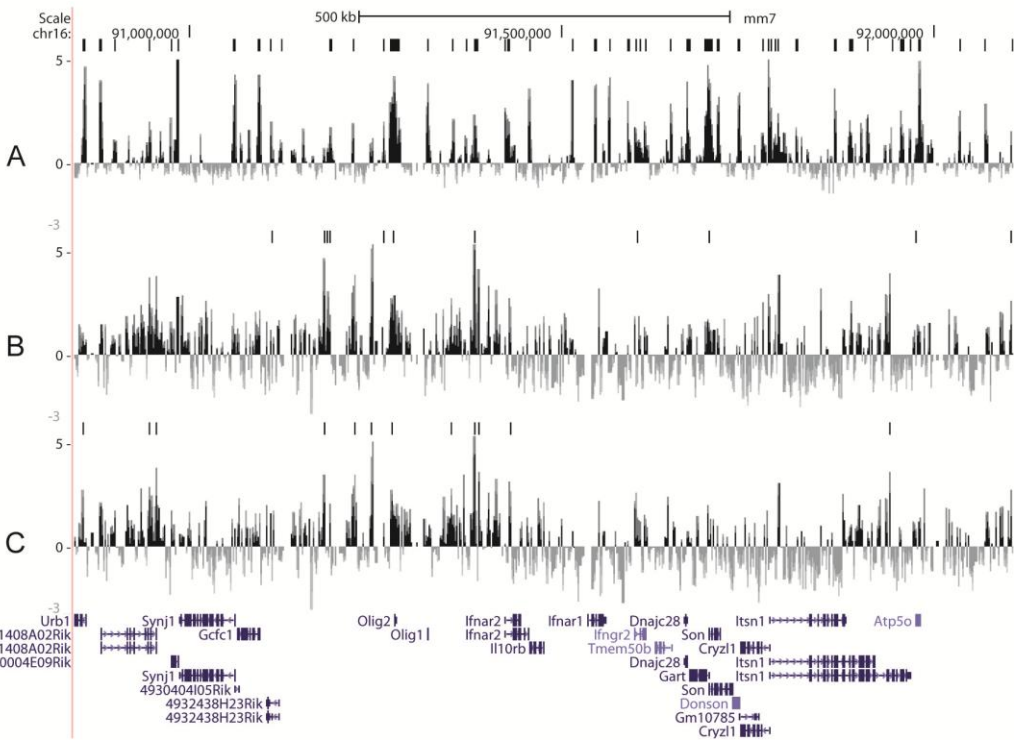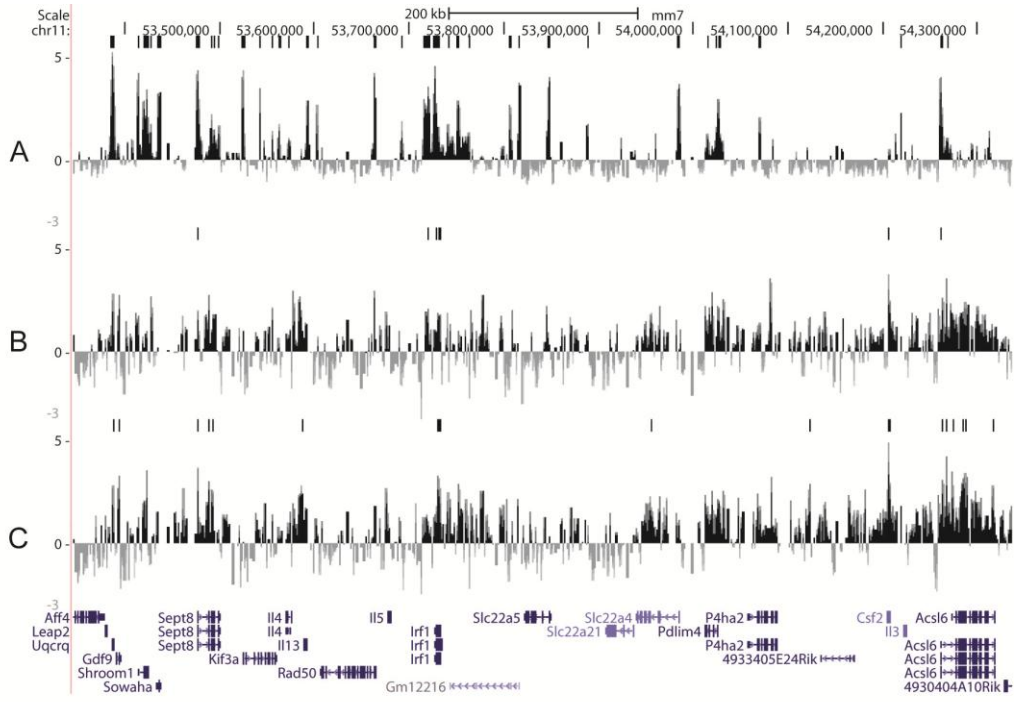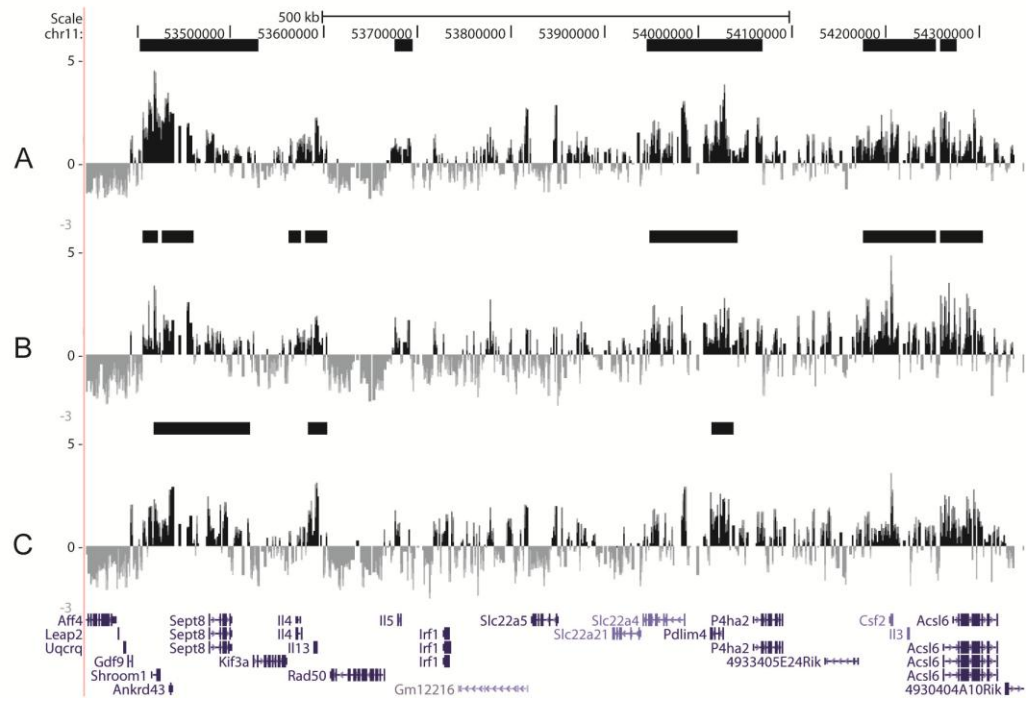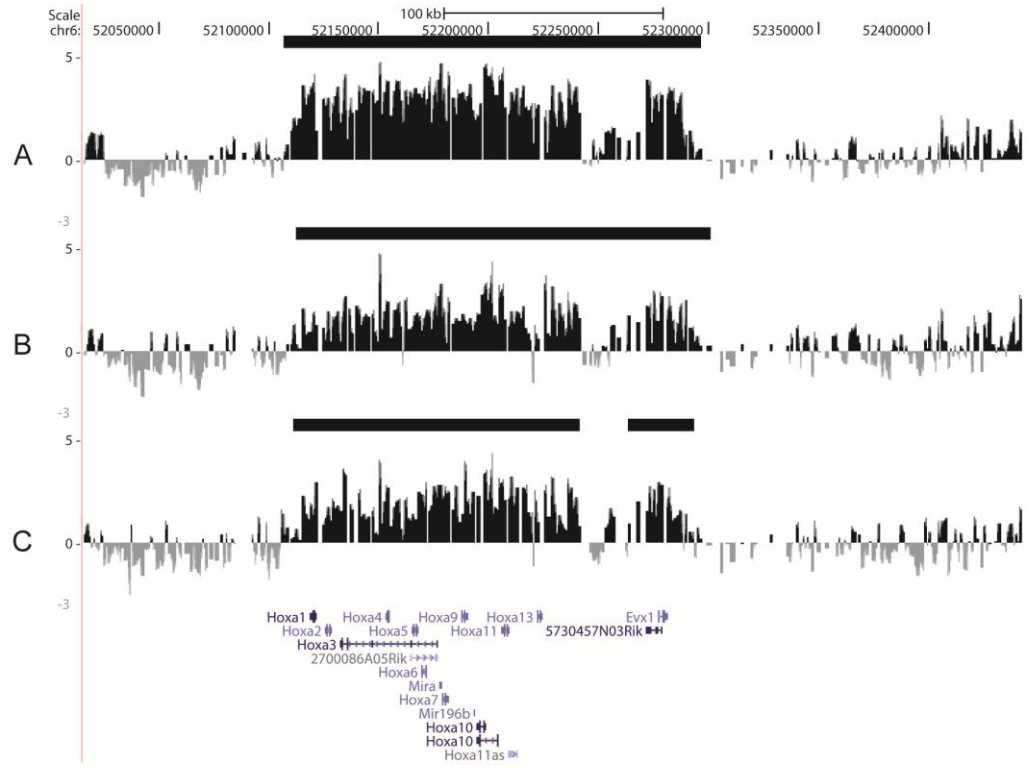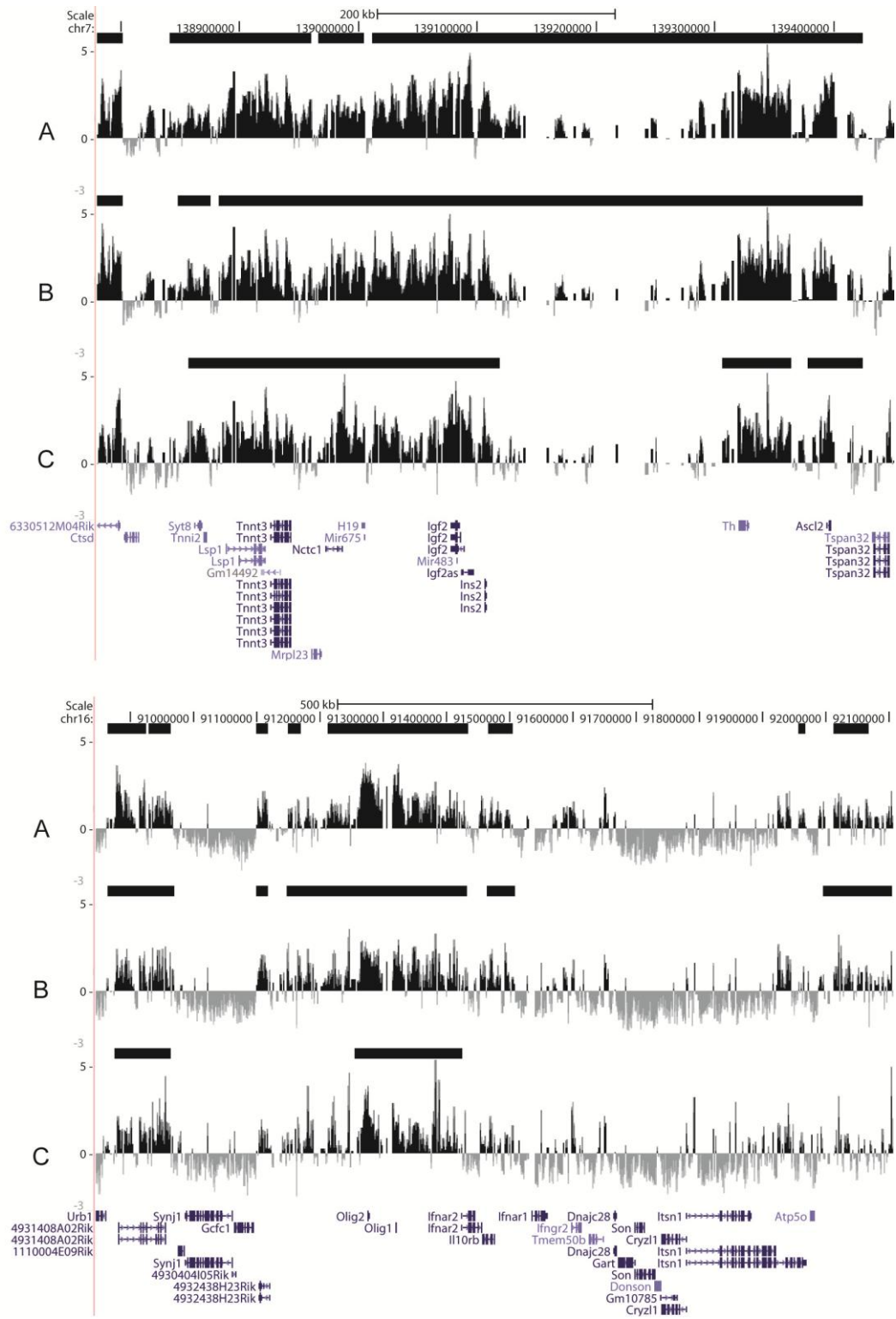
Figure 4.10: The effect of pre-clearing on low-cell H3K27me3 ChIP.
Screenshot views of the UCSC genome browser (Kuhn et al., 2007) of four different regions represented on the mouse arrays. For each region the ChIP-chip profiles are shown in the following order: (A) 10.0 µg H3K27me3 conventional ChIP; (B) NPC; and (C) PC H3K27me3 low-cell ChIP. The black horizontal bars on top of each track show the significantly enriched domains identified by the BLOCs algorithm. The bottom track shows the RefSeq genes (Pruitt et al., 2007) with transcriptional orientation indicated by arrows. The ChIP-chip data is displayed in the three intervening tracks as the ratio of ChIP sample fluorescence to input DNA fluorescence. Each black vertical bar is the enrichment measured at a single array element with the relative $\log_2$ fold enrichments represented at the height of the bar. The scale and the chromosome coordinates are represented on top of each panel.
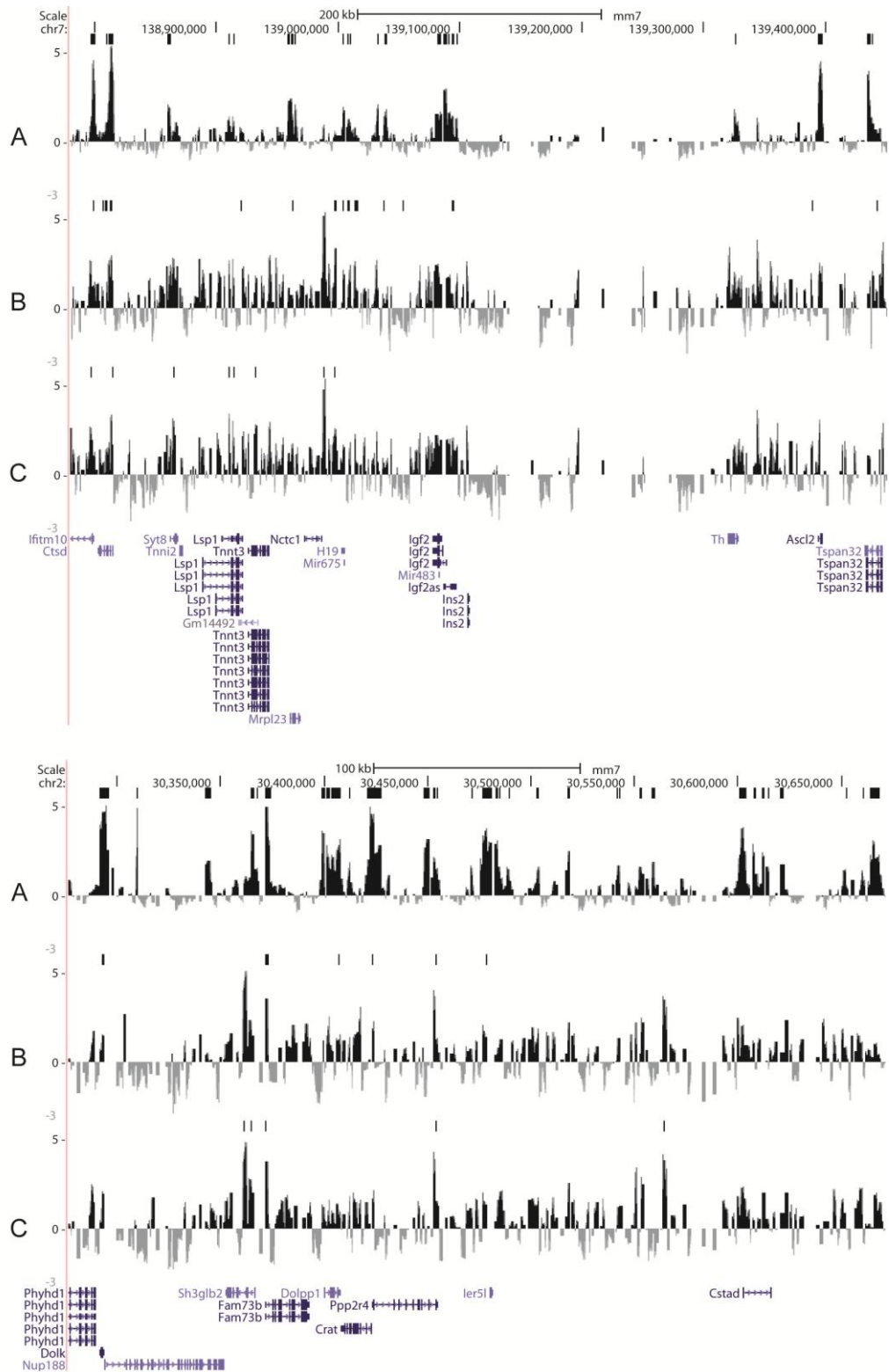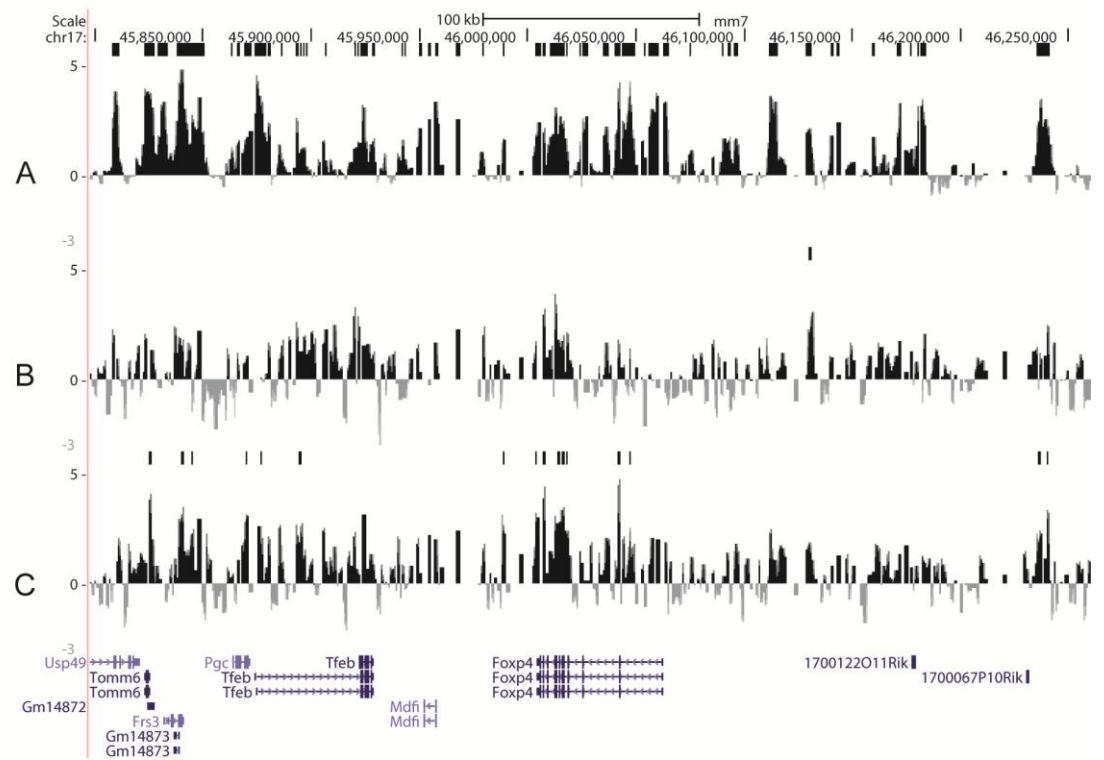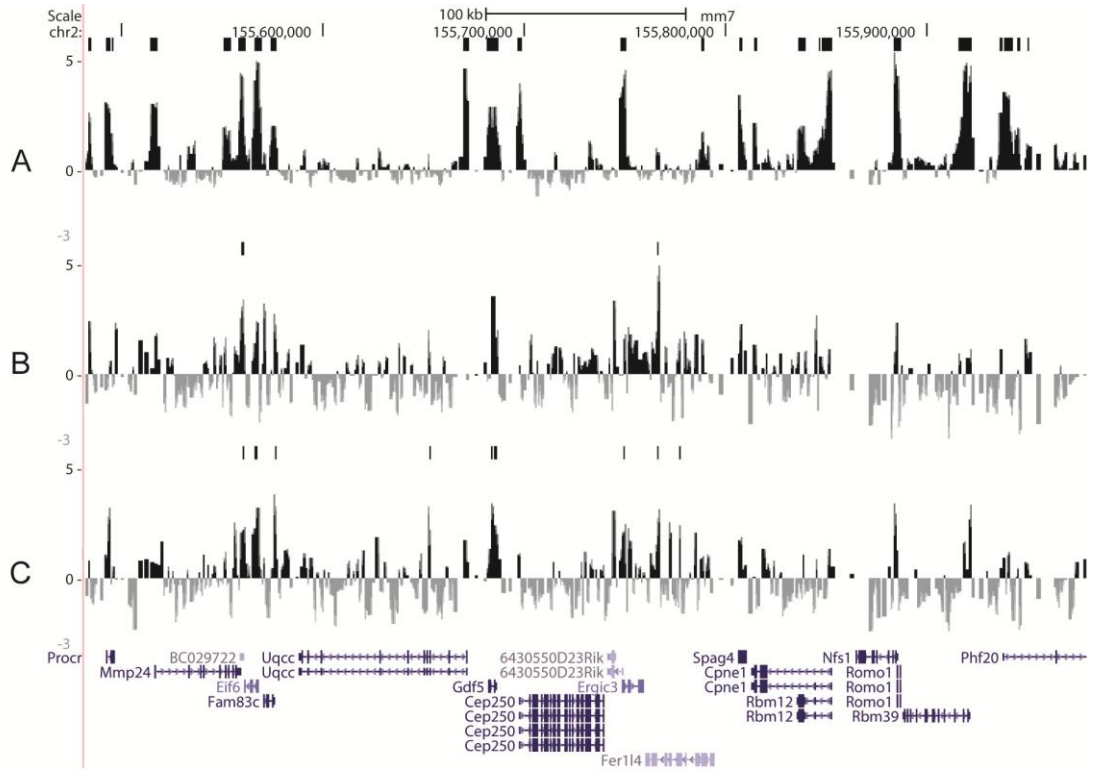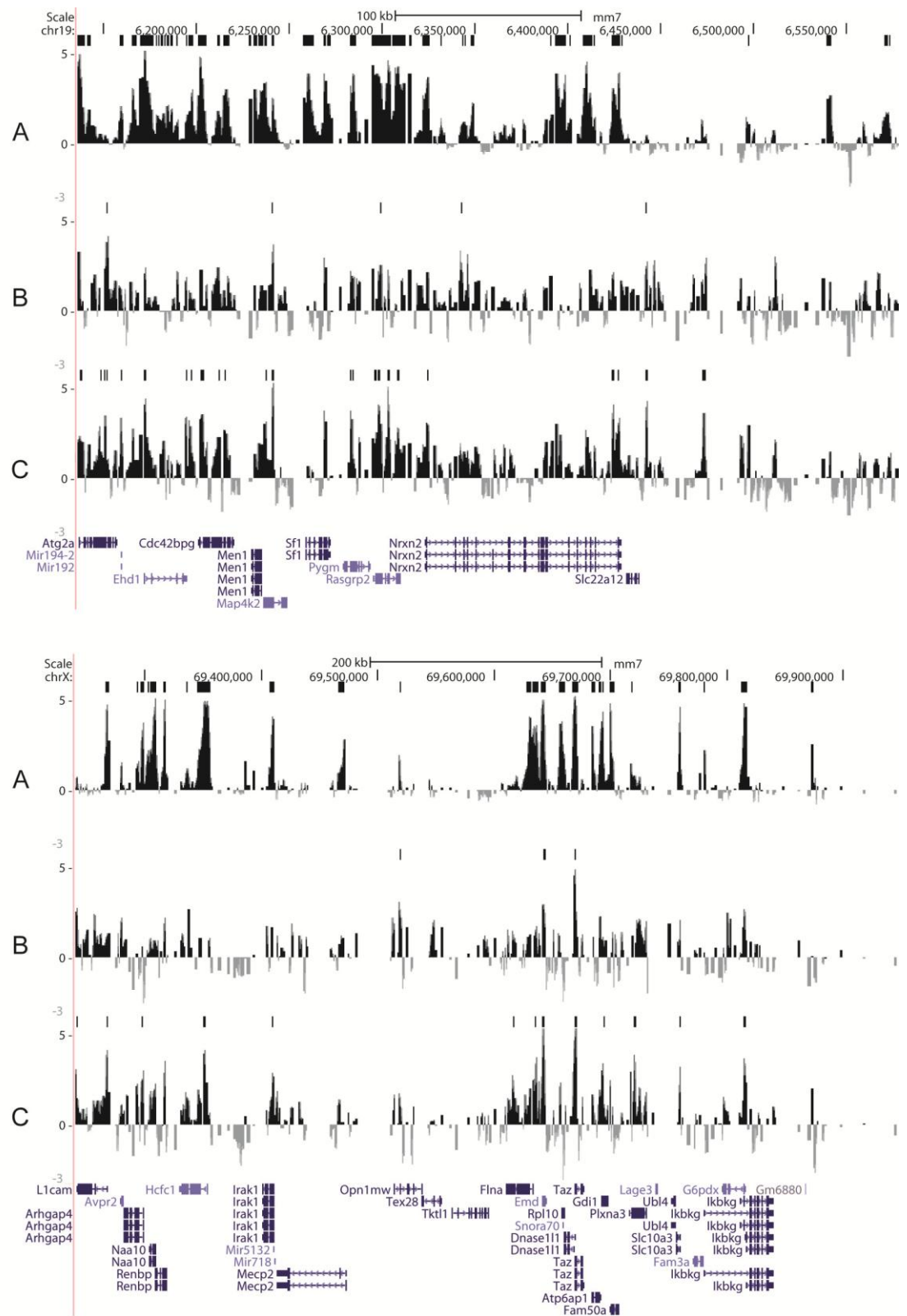
Figure 4.11: The effect of pre-clearing on low-cell H3K4me3 ChIP.
Screenshot views of the UCSC genome browser (Kuhn et al., 2007) of four different regions represented on the mouse arrays. For each region the ChIP-chip profiles are shown in the following order: (A) 10.0 µg H3K4me3 conventional ChIP; (B) PC H3K4me3 low-cell ChIP; and (C) NPC;. The black bars on top of each track show the significantly enriched peaks identified by ChIPOTle. The bottom track shows the RefSeq genes (Pruitt et al., 2007) with transcriptional orientation indicated by arrows. The ChIP-chip data is displayed in the three intervening tracks as the ratio of ChIP sample fluorescence to input DNA fluorescence. Each black vertical bar is the enrichment measured at a single array element with the relative $\log_2$ fold enrichments represented at the height of the bar. The scale and the chromosome coordinates are demonstrated on top of each panel.

When compared against the results of conventional ChIP across all 12 murine regions represented on the tiling path array, low-cell H3K27me3 ChIP performed under the NPC condition showed higher Spearman's correlation coefficient (0.86) than the assay under the PC condition (0.82) (Figure 4.10). Using the BLOCs algorithm, low-cell H3K27me3 ChIP under the NPC condition could identify 29 of the 36 islands identified in the conventional assay. Furthermore, low-cell H3K27me3 ChIP under the NPC condition identified five novel domains. The novel peaks could be a result of high background noise levels that were classified as real enrichments (false positives) or the rare events taking place in a subset of cells in the larger population. All of this data, when taken together, suggested that the H3K27me3 low-cell ChIP-chip data derived from IPs under the NPC condition showed higher degree of similarity to the results of conventional ChIP. However, the level of sensitivity and specificity did not show further improvements.

Moreover, low-cell H3K4me3 ChIP performed under the NPC conditions showed higher Spearman's correlation coefficient (0.54) than the assay under the PC condition (0.37) in comparison to the conventional ChIP assay across all 12 murine regions represented on the tiling path array (Figure 4.11). H3K4me3 ChIP under the NPC condition could identify 94 of the 462 peaks identified in the conventional assay as determined by ChIPOTle. Furthermore, low-cell H3K4me3 ChIP under the NPC condition identified 29 novel peaks. All of this data, when taken together, suggested that the H3K4me3 low-cell ChIP-chip data derived from IPs under the NPC condition showed higher degree of similarity to the results of conventional ChIP. Furthermore, the sensitivity and the specificity of low-cell H3K4me3 ChIP was considerably improved under the NPC condition in comparison to the assays performed under the PC condition.

Overall, H3K4me3 and H3K27me3 low-cell ChIP under NPC conditions showed higher degree of similarity to the conventional ChIP-chip results, in addition to the noticeable improvements of sensitivity and specificity in the case of low-cell H3K4me3 ChIP.

4.6.4 Summary of optimisation of low-cell ChIP

The low-cell ChIP method showed a noticeable improvement by adjusting the antibody-to-chromatin ratio, performing simultaneous cell and nuclear lysis, and

excluding the pre-clearing step. This improvement was demonstrated by the increase in statistical correlation with the conventional method (reproducibility), improved sensitivity which was shown by the number of islands or peaks identified significantly above background that were overlapping with the results of conventional ChIP assays, and improved specificity by showing the reduced numbers of novel peaks or islands identified significantly above background. Table 4.4 summarises the findings for each optimisation step.

Histone modification	Optimisation step		Statistical correlation with the conventional ChIP		No. of significant islands/peaks	
			Spearman's rank correlation	P-value	Overlapping conventional ChIP	Novel islands/peaks
H3K27me3	Antibody titration	0.5µg	0.073	0.000000	5/36	18
		1.0µg	0.665	0.000000	16/36	8
		5.0µg	0.748	0.000000	19/36	3
	Simultaneous lysis		0.821	0.000000	29/36	4
	No pre-clearing		0.867	0.000000	29/36	5
H3K4me3	Antibody titration	0.5µg	0.261	0.000000	10/462	28
		1.0µg	0.22	0.000000	13/462	35
		5.0µg	0.357	0.000000	64/462	43
	Simultaneous lysis		0.369	0.000000	31/462	37
	No pre-clearing		0.545	0.000000	94/462	29

Table 4.4: Summary of statistical tests performed for each optimisation step.
The values highlighted in green show the antibody concentration that produced highest statistical correlation and overlapping number of islands or peaks with the conventional ChIP. The values highlighted in red represent the optimisation steps that showed an improved statistical correlation, sensitivity (number of islands or peaks identified with BLOCs or ChIPOTle, respectively), and specificity (number of novel islands or peaks identified with BLOCs or ChIPOTle, respectively) in low-cell ChIP assays performed with 5.0 µg of antibody. All the correlation coefficients were highly significant (p-value=0).

The low-cell ChIP assays performed with anti-H3K27me3 antibody demonstrated higher statistical correlation with the conventional assay, noticeably better sensitivity and specificity in comparison with the low-cell ChIP assays performed with the anti-H3K4me3 antibody. Therefore, the efficiency of low-cell ChIP is largely determined by the sensitivity and specificity of the antibody under the low-cell conditions.

4.6.5 Investigating the reproducibility of the improved low-cell ChIP across bioreplicates

In order to determine the reproducibility of low-cell ChIP across bioreplicates that were obtained from different chromatin preparations, low-cell H3K4me3 and H3K27me3 ChIP assays were performed under the improved conditions for two bioreplicates. Similarly conventional H3K4me3 and H3K27me3 ChIP assays were performed for two bioreplicates. The resultant ChIP material was hybridised to murine tiling path arrays. The conventional and low-cell ChIP-chip profiles that were derived across four different regions of the mouse genome are shown in Appendix 5.

The comparison of two low-cell H3K27me3 ChIP bioreplicates across all 12 murine regions represented on the tiling path array, showed a significantly high Spearman's correlation coefficient (0.87) (Table 4.5). Similarly the comparison of two low-cell H3K4me3 ChIP bioreplicates across all 12 murine regions represented on the tiling path array, also demonstrated a significantly high Spearman's correlation coefficient (0. 73).

Furthermore, the Spearman's correlation coefficient between the two conventional H3K27me3 ChIP assays across all 12 murine regions represented on the tiling path array, showed a significantly high Spearman's correlation coefficient (0.83) (Table 4.5). Additionally, the comparison of two conventional H3K4me3 ChIP assays across all 12 murine regions represented on the tiling path array also demonstrated a significantly high Spearman's correlation coefficient (0.84).

The above results, therefore, indicated that the low-cell ChIP assays reproducibility between the bioreplicates obtained from different chromatin preparations was comparable to the degree of similarity observed between the bioreplicates in the conventional ChIP assays.

Histone modification	Statistical correlation between the low-cell ChIP bioreplicates		Statistical correlation between the conventional ChIP bioreplicates	
	Spearman's rank correlation	P-value	Spearman's rank correlation	P-value
H3K27me3	0.877	0.000000	0.834	0.000000
H3K4me3	0.737	0.000000	0.842	0.000000

Table 4.5: Statistical correlation between the bioreplicates in low-cell and conventional ChIP assays.
Highly significant (p-value=0) positive Spearman's rank correlations between the bioreplicates indicate a high reproducibility rate within low-cell ChIP assays. The observed correlation values between bioreplicates of conventional ChIP were also presented to show the level of reproducibility and the degree of variation within the ChIP assays.

4.7 Analysing the performance of low-cell ChIP coupled to deep sequencing (low-cell ChIP-seq)

ChIP-chip analysis using the mouse microarray was used to improve the low-cell ChIP conditions and to characterise the antibody levels in a smaller scale genomic platform, which covered only 0.25% of the mouse genome. However, a thorough analysis of genome-wide distribution of histone modifications was mainly possible by ChIP-seq. Except for the nano-ChIP method which has been recently reported in association with Illumina deep sequencing platform (Adli et al., 2010), no other ChIP assays designed for the small cell numbers were reported in this context. Therefore, to investigate the performance of low-cell ChIP-seq method, sequencing libraries were prepared for both conventional and low-cell H3K4me3 and H3K27me3 ChIP assays, as described in Chapter 2.

The sensitivity of low-cell ChIP-seq was examined by investigating the overlap of identified enrichment peaks by low-cell ChIP-seq with the ones identified by conventional ChIP-seq. The promoters of the 'on' genes (n=12,677), which were defined as the top 50% of expressed genes in the mES genome (E14 clone), were searched for the presence of H3K4me3 enrichment domains using BayesPeak algorithm. The total number of uniquely aligned sequencing reads in conventional H3K4me3 ChIP-seq was 2.5 times greater than the low-cell H3K4me3 ChIP-seq. Therefore, in order to reduce a potential bias towards the conventional ChIP-seq which had the bigger library size, random sampling of equivalent to low-cell ChIP-seq library size was performed on the conventional ChIP-seq library. On average 10,718 peaks were identified at the promoters of the 'on' genes by the random

sampling of conventional ChIP-seq library from which 64% could be identified at these promoters by the low-cell ChIP-seq (Table 4.6).

To evaluate the performance of low-cell H3K27me3 ChIP-seq data, the promoters of bivalent genes (n=2,706) reported in mES cells (Mikkelsen et al, 2007) were used as a reference to search for enrichment peaks by SICER algorithm. The total number of uniquely aligned sequencing reads in low-cell H3K27me3 ChIP-seq was 1.2 times greater than the conventional H3K27me3 ChIP-seq. Therefore, in order to reduce a potential bias towards the low-cell ChIP-seq which had the bigger library size, random sampling of equivalent to conventional ChIP-seq library size was performed on the low-cell ChIP-seq library. On average 2,145 peaks were identified at the promoters of the bivalent genes by the conventional ChIP-seq library from which 80% could be identified at these promoters by the low-cell ChIP-seq (Table 4.6).

The above observations indicated that the low-cell H3K4me3 and H3K27me3 ChIP-seq assays demonstrated a large degree of overlap with the conventional H3K4me3 and H3K27me3 ChIP-seq assays in terms of identifying biologically important features such as active or bivalent promoters, respectively. Low-cell H3K4me3 ChIP-seq assays, in particular, demonstrated a noticeably improved sensitivity when compared to the earlier observations in the low-cell H3K4me3 ChIP-chip assays.

Histone modification	Library size (uniquely aligned reads)		Low-cell ChIP-seq peaks	Conventional ChIP-seq peaks	Random sampling peaks	Low-cell/conventional peaks overlap
	Low-cell ChIP-seq	Conventional ChIP-seq				
H3K4me3	7,681,623	19,831,558*	6,875	10,454	10,725 10,721 10,708	64%
H3K27me3	6,475,810*	5,397,887	2,275	2,685	2,151 2,128 2,157	80%

Table 4.6: The comparison of low-cell and conventional ChIP-seq in mES cells.
ChIP-seq analyses for H3K4me3 and H3K27me3 in mES cells are summarised. The total number of uniquely aligned sequencing reads marked with the asterisks specifies the bigger library size for each histone modification which was used for the random sampling of equivalent to the size of smaller library. Random sampling was performed three times and the average number of islands was used to measure the percentage of overlap between the low-cell and conventional ChIP-seq assays.

4.8 Discussion

4.8.1 Improvements to low-cell ChIP method

The focus of this Chapter was to develop a sensitive and reproducible ChIP assay that can use as few as 10,000 cells through modifying various steps of the conventional ChIP. Adjustments in three steps of conventional ChIP assay resulted in increased reproducibility, sensitivity and specificity of low-cell ChIP. The adjusted steps included (i) antibody titration, (ii) simultaneous cell and nuclear lysis, and (iii) omitting the pre-clearing step.

The amount of antibody used in low-cell ChIP assays was shown to be an important factor in improving the sensitivity and specificity of the low-cell ChIP assays for both H3K4me3 and H3K27me3 antibodies. This was due to the reduced chromatin amounts in low-cell ChIP assays. It was demonstrated that 5 μg of anti-H3K4me3 and anti-H3K27me3 antibodies can reproduce the conventional ChIP-chip profiles more efficiently than the lower tested amounts of antibodies. Additionally, 5 μg of anti-H3K27me3 antibody increased both sensitivity and specificity of low-cell ChIP, whereas 5 μg of anti-H3K4me3 antibody could increase the sensitivity but not specificity. Therefore, it was noticeable that the optimal antibody to chromatin ratio could be assay-specific as H3K4me3 antibody performed less efficiently under the low-cell ChIP conditions than the H3K27me3 antibody under similar conditions based on the microarray readouts.

Furthermore, performing low-cell ChIP assays with simultaneous cell and nuclear lysis as opposed to the conventional two-step cell and nuclear lysis increased the low-cell ChIP efficiency in reproducing the conventional ChIP-chip profiles. This adjustment also increase the sensitivity of low-cell H3K27me3 and the specificity of low-cell H3K4me3 ChIP assays. Therefore, minimising the handling of cells during low-cell ChIP as a result of simultaneous cell and nuclear lysis would simplify the ChIP method and improve the low-cell ChIP readouts on microarrays.

Chromatin pre-clearing prior to IP during conventional ChIP assays may improve ChIP efficiency by removal of excess proteins and cellular debris present in the cell and nuclear lysate that may sequester the antibody or may prevent antibody interaction with the right epitope. However, when using reduced chromatin amounts in low-cell ChIP assays, the pre-clearing step may result in the removal

of real DNA-protein interactions and in turn true signals as a result of non-specific interactions with IgG. Consequently, by omitting from the low-cell ChIP assays, their sensitivity and specificity were improved for both H3K4me3 and H3K27me3 assays. Furthermore, low-cell ChIP was a lot more efficient to reproduce the conventional ChIP-chip profiled in the absence of pre-clearing step.

The improved low-cell ChIP was coupled to deep sequencing to investigate the sensitivity of low-cell ChIP assays at the genome-wide level in identifying biologically important features such as promoters. The low-cell H3K27me3 ChIP-seq assay could identify 80% of the H3K27me3 peaks that were also identified by the conventional H3K27me3 ChIP-seq at the bivalent promoters in mES cells. This was in agreement with the percentage of conventional H3K27me3 ChIP-chip peaks that were identified significantly above background by the low-cell H3K27me3 ChIP-chip (80%). Furthermore, the low-cell H3K4me3 ChIP-seq assay could identify 64% of the H3K4me3 peaks that were also located by the conventional H3K4me3 ChIP-seq assay at the promoters of 'on' genes in mES cells. This was significantly higher than the percentage of conventional H3K4me3 ChIP-chip peaks that were identified significantly above background by the low-cell H3K4me3 ChIP-chip (20%). The differences between the microarray and sequencing readouts could be associated with the absence of noise generated as a result of cross-hybridisations or labelling artefacts in the sequencing assays. Additionally, the requirement for significantly lower amounts of starting material for sequencing could suggest a more suitable readout for the low-cell ChIP assays. Furthermore, the sequencing results were not limited in their coverage, which can introduce bias in the microarray analyses as a result of variation in the association between the protein of interest and the regions presented on the arrays.

4.8.2 Comparison of low-cell ChIP method with other related ChIP methods

The advantages of this improved method over the previously reported ChIP assays designed for the small number of cells (Table 4.2) are discussed briefly below. The improved low-cell ChIP method presented in this Chapter could be coupled to genome-wide applications such as microarray and sequencing in the absence of whole genome PCR amplifications that could introduce bias to the results. Conversely, CChIP and Q^2ChIP were reported mainly in association with

qPCR readouts and genome-wide application of these methods were not reported (Dahl and Collas, 2007, O'Neill et al., 2006). Moreover, microChIP, miniChIP and µChIP that could be coupled with the microarray analyses required genome-wide amplification of ChIP-purified DNA prior to array hybridisations (Acevedo et al., 2007, Collas, 2011, Weishaupt and Attema, 2010).

Furthermore, the improved low-cell ChIP method presented in this Chapter did not require carried DNA or chromatin from other species, such as CChIP (O'Neill et al., 2006) and microChIP methods (Acevedo et al., 2007), to increase the yield and efficiency of the ChIP assays with small cell numbers. The abundance of sequence reads from foreign DNA material would be an issue for ChIP-seq.

In a more recent method, nano-ChIP-seq (Adli et al., 2010), ChIP-purified DNA from 10,000 mES cells was coupled to Illumina deep sequencing. This method also demonstrated the importance of antibody titration in increasing the efficiency of ChIP assays using small cells numbers. The ChIP-purified DNA by nan-ChIP was subjected to four cycles of random priming that required specialised random hairpin primers. Subsequently, the hairpin primers were subjected to restriction digestion and PCR amplification for 10-15 cycles prior to preparation of Illumina sequencing libraries, which require an extra 18-cycles of amplification. Random priming method could reduce the bias generated by the whole-genome amplification methods since the template DNA fragment could give rise to amplified fragments varying in size and sequence. Consequently, this would increase the complexity of the ChIP purified DNA library before PCR amplifications. Nano-ChIP-seq H3K4me3 was shown to identify 93% of the enriched promoters by the conventional ChIP-seq with the true positive rate of 68-83% and the true negative rate of 90% as reported by the receiver operating characteristic (ROC) curve (Adli et al., 2010). However, the improved low-cell ChIP method presented in this Chapter could be coupled to Illumina sequencing with a comparable level of sensitivity in the absence of universal hairpin primers and extra 10-15 rounds of amplification. Consequently, the improved low-cell ChIP method is simpler, but as sensitive as the nano-ChIP-seq approach.

The improved low-cell ChIP method is a lengthy procedure that could last four days. Conversely, µChIP method is a rapid approach as the whole procedure can be done in a day (Dahl and Collas, 2008). The µChIP method was developed by

combining different steps of conventional ChIP including reversal of crosslinking, protein digestion, and DNA elution into a single shorter (~2 hrs) step as well as using paramagnetic beads which speed up the IP step.

4.8.3 Further improvements of low-cell ChIP

Low-cell ChIP method can be further improved through implementing other adjustments. Although the handling of small cell numbers has been significantly improved by performing the fixation and sonication in small tubes with only one-step cell and nuclear lysis, replacing the probe sonication with waterbath sonication could significantly reduce the loss of material, cross-contamination, and non-uniform sonication bias. Besides, optimisation of the sonication regime is essential to maintain an optimised size range for the DNA fragments at 200-500 bp as smaller and larger DNA fragments could affect the downstream applications.

The improved low-cell ChIP method was not tested for the transcription factors and less abundant chromatin-bound proteins. Non-histone proteins can be difficult to ChIP under low-cell number ChIP conditions due to the reduced signal levels which are lost in the background noise. The concentration and incubation of the fixing agent should be optimised for such assays and antibody levels should be titrated in a wider range covering lower concentrations. Nevertheless, microChIP, Q^2ChIP and µChIP methods reported non-histone protein ChIP assays, particularly for RNAPII, which were mainly limited to qPCR readouts (Q^2ChIP and µChIP) (Dahl and Collas, 2007, Dahl and Collas, 2008) or required whole-genome amplifications for microarray analysis (Acevedo et al., 2007).

Nevertheless, low-cell ChIP assays with poor outcome could result from inefficient antibodies at the low-cell conditions due to the high antibody-to-chromatin ratio. Therefore, antibodies from different organisms or reactive against different epitopes can be tested. Moreover, N-ChIP could be substituted when the epitopes are feared to be masked by the fixation agent, although this would largely limit the application of low-cell ChIP to histone proteins. Finally, the usage of magnetic beads instead of agarose beads may have an impact on the efficiency of low-cell ChIP by increasing the affinity for the antibody-protein-DNA complexes and can also reduce the length of low-cell ChIP procedure.

4.9 Conclusion

The work presented in this Chapter described the improvements to a low-cell ChIP method (Dillon, PhD thesis, 2008) to generate reproducible and sensitive genome-wide histone modification maps from as low as 10,000 cells using ChIP-chip and ChIP-seq. These improvements were made to several steps of the conventional ChIP including (i) antibody titration, (ii) simultaneous cell and nuclear lysis, and (iii) omitting the pre-clearing step. This method can be used for identification of genome-wide histone modifications in primary samples with limited cell numbers such as HSCs whose epigenetic signature could not be established otherwise by ChIP-seq. In Chapter 5 the application of low-cell ChIP method to the primary human HSCs and CML stem cells was demonstrated to investigate the genome-wide distribution of histone modifications that would shed light on the underlying epigenetic mechanisms that are potentially important in the maintenance of LSCs.

5. Abnormal epigenetic programming in CML cells modulate gene expression patterns.

5.1 Abstract

The role of chromatin modifications in regulating gene expression has been widely investigated in the course of normal development and tumourigenesis (reviewed in Section 1.5). However, due to the limitations of working with rare adult stem cell populations such links have not yet been established in the context of cancer stem cells. Therefore, in this Chapter, the genome-wide distribution of histone modifications in LSCs was investigated using the low-cell ChIP-seq approach and compared to the histone modification profiles of HSCs, HPCs, and LPCs. It was shown that the HSC identity pathways (Chapter 3) were significantly associated with the bivalent class of promoters. Furthermore, it was demonstrated that an epigenetic reprogramming, which was associated with gene expression changes in CML cells, was initiated during LSC formation. Consequently, CML cells were shown to be more dependent on Polycomb proteins for modulating expression patterns in a different manner than that found in HPCs.

5.2 Introduction

There have been few studies investigating the genome-wide chromatin modifications in leukaemia. One studied investigated the genome-wide distribution of activation histone mark, H3K9 acetylation, in three cases of AML and two cases of Ph$^+$ ALL (Figueroa et al., 2008). The histone modification maps were combined with genome-wide DNA methylation and gene expression analysis. This study demonstrated an inverse relationship between promoter DNA methylation and H3K9 acetylation, which was found at the promoters of active genes. This report also suggested that the integration of H3K9 acetylation maps and gene expression profiles could increase the number of identified significantly differentially regulated genes between ALL and AML samples. Furthermore, the integration of gene expression signatures and chromatin modifications improved the identification of differentially regulated biological pathways in AML relative to ALL. This study showed that the integration of epigenetic analysis and gene expression profiles could enhance the identification of many novel genes and pathways that would have been missed with a less comprehensive approach.

MLL, one of the catalytic subunits of TrxG complex, is linked to chromosomal translocations that are found in 70% of the childhood leukaemias (Krivtsov and Armstrong, 2007). The MLL-rearranged leukaemias are shown to have different gene expression profiles than both ALL and AML leukaemias (Armstrong et al., 2002), which in indicative of a unique global programming in these cancers. In order to investigate the changes of chromatin conformation in MLL-rearranged leukaemias, a knock-in model of MLL-AF4 fusion oncogene have been developed (Krivtsov et al., 2008). The downstream targets of this fusion protein, such as HoxA9, RUNX1, ETV6 and RUNX2, showed significant increase of H3K79me2 histone modification, which is associated with transcribed genes. Genome-wide analysis of H3K79me2 in leukaemia cells identified ~1,200 gene promoters that were significantly associated with H3K29me2 marks compared to normal pre-B cells. Similar observations were also made in the cases of human MLL-rearranged ALL. Therefore, it was suggested that the MLL-AF4 fusion oncoprotein abnormally recruited DOT1L, H3K79 methyltransferase, to their target genes which resulted in their abnormal activation. Furthermore, similar association with DOT1L was also observed in the cases of MLL fusion proteins with AF9, AF10 and MLLT1 (Krivtsov et al., 2008) Interestingly, it was previously reported that the introduction of MLL-AF9 fusion oncoprotein to the haemopoietic progenitor cells could induce the formation of LSCs. This was associated with global reprogramming of gene expression which in turn caused increased self-renewal activity (Krivtsov et al., 2006). This finding, therefore, indirectly links the formation of LSCs from progenitor cells to the changes in chromatin structure (Krivtsov and Armstrong, 2007).

The role of PRC2 complex and in particular EZH2 in haematological malignancies have also been studied in recent years. Several reports have identified somatic mutations in EZH2 in several lymphoid or myeloid leukaemias. For instance, a gain-of-function mutation in the catalytic SET domain of EZH2 is found in association with diffuse large B-cell lymphoma. This observation suggested an oncogenic role for EZH2 (Morin et al., 2010, Sneeringer et al., 2010). However, loss-of-function mutations in EZH2 were also identified in T-ALL (Ntziachristos et al., 2012) and myeloid malignancies (Ernst et al., 2010). These observations, therefore, propose a tumour-suppressive role for EZH2. Consequently, the function of EZH2 as an oncogene or tumour-suppressor can be context-dependent which is determined by the haemopoietic lineage.

In addition to the identified mutations in EZH2, the overexpression of EZH2 is also reported in myeloid malignancies and AML (Grubach et al., 2008, Xu et al., 2011). A recent report showed that the EZH2 knockout in AML cells results in the differentiation of AML cells and blocks the progression of the disease (Tanaka et al., 2012). In this study, deletion of EZH2 in AML cells inhibited cell cycle progression and induced apoptosis, although their proliferative capacities were not affected. The EZH2-depleted AML cells failed to engraft immunodeficient mice as a result of significant reduction in the number of AML stem cells. Therefore, it was suggested that EZH2 was required for the maintenance of LSCs, especially against environmental stresses. The deletion of EZH2 was associated with significant reduction of H3K27me3 histone modifications, although many gene promoters maintained their H3K27me3 levels (Tanaka et al., 2012). This was linked to the potential role of EZH1 in partially complementing EZH2 similar to its role in normal HSCs, where EZH1 is the dominant catalytic subunit of PRC2 complex (Mochizuki-Kashio et al., 2011). The loss of EZH2 in normal HSCs did not affect their self-renewal capacities (Mochizuki-Kashio et al., 2011), but the loss of EZH2 in AML cells significantly impaired the self-renewal capacities of LSCs, suggesting that the AML cells are highly-dependent on EZH2 function (Tanaka et al., 2012). This report also showed that EZH2 targets were associated with differentiation and cell-fate determination in AML cells. Thus, EZH2 potentially reinforces the LSC compartment in AML by blocking differentiation mechanisms (Tanaka et al., 2012).

Overall, the above studies on the role of chromatin modifications in leukaemogenesis were performed on the bulk of leukaemia cells. The genome-wide analysis of chromatin modifications in leukaemia stem cells have not been performed to date due to the technical limitations with ChIP method, as discussed in Chapter 4. However, in order to understand the early events in global epigenetic programming during leukaemic transformation the analysis of chromatin structure in LSCs is crucial. Therefore, the development of the low-cell ChIP method (Chapter 4) could help elucidating the histone modification maps of LSCs in CML. These histone modifications can be further interpreted in association with the gene expression profiles that were established in Chapter 3 for a comprehensive analysis of differentially regulated genes and pathways in LSCs and LPCs.

5.3 Methods

As described in Chapter 4, ChIP-purified DNA fragments can be coupled to various downstream applications such as qPCR, microarrays (ChIP-chip), and deep sequencing (ChIP-seq). The advantages of ChIP-seq over other approaches can be summarised as higher resolution, enhanced coverage, fewer artefacts, and a greater dynamic range. Genome-wide histone modification profiling using ChIP-seq was first reported in human T cells (Barski et al., 2007) which used the Illumina/Solexa 1G platform to generate approximately 8 million tags per sample. ChIP-seq technology was also used in profiling histone modifications in mES cells and differentiated cells (Mikkelsen et al., 2007).

The requirement for 10 ng of starting material has made ChIP-seq a preferential method for studying abundant protein-DNA interactions in low cell number samples, as performing genome-wide studies using ChIP-chip demands whole-genome amplifications to generate >2 µg of DNA per array which could also introduce PCR bias (Park, 2009). Nonetheless, the amount of purified DNA in low-cell ChIP cannot be accurately quantified by NanoDrop and is estimated to be at picogramic quantities. Conventional ChIP-seq library preparation method requires at least 10 ng of DNA for the generation of high complexity libraries that can produce an optimum number of clusters and as a result can generate high quality and reliable sequencing data. On the contrary, a low complexity template, in the case of template quantities below 10 ng, poses various challenges that should be addressed in advance to reliably obtain and interpret the sequencing data. The challenges include (i) the presence of contaminating DNA material that can affect the sequencing even in the smallest amounts, and (ii) generation of PCR product duplicates which have previously been reported to reach up to 60% of the sequences (Quail et al., 2008).

The ChIP-seq technology also has certain limitations such as the sequencing errors that arise towards the end of each read (Minoche et al., 2011), or a bias towards GC-rich fragments caused by the amplification steps during ChIP-seq library preparation or during sequencing (Hillier et al., 2008, Quail et al., 2008). The cluster generation in Illumina platforms is an important step in determining the quality of output data, where too little sample will result in very few tags and too much sample results in the generation of clusters too close to one another which

reduces the data quality (Park, 2009). New improvements of the Illumina sequencing technology allow generation of 75 or 106 bp sequence reads to increase the alignability and therefore, specificity at a lower cost. However, it was shown that the gain in unique alignments declines after ~35 bp and is marginal beyond 70-100 bp (Nix et al., 2008).

The depth of sequencing is defined as the number of sequenced fragments in a ChIP-seq experiment. Initially, 4-6 million reads were generated per each sequenced sample; however, with the recent improvements of the NGS platforms the depth of sequencing has increased to 20-30 million reads per lane (Park, 2009). The depth of sequencing is subjective and largely dictated by the chromatin binding pattern of the protein of interest. For instance, in the case of histone modifications larger numbers of reads are required due to the greater genome coverage. It was demonstrated in a recent study that the number of protein binding sites detected increases steadily by increasing the depth of sequencing (Kharchenko et al., 2008). However, this report showed that the sequencing depth can reach a 'saturation point' after which additional reads will not identify any new binding sites.

It has been reported that open chromatin regions associated with promoters and regulatory domains of expressed genes are fragmented more frequently by sonication. Therefore, the sequencing library could be biased towards the open chromatin regions as a result of non-uniform sonication (Auerbach et al., 2009). Repeat sequences could also confer aligning difficulties which results in their exclusion from the library of sequenced reads. As a result, a control sample could be sequenced such as input DNA or non-specific IP using a non-specific IgG antibody. By using the input DNA the biases causes by non-uniform fragmentation and PCR amplification could be corrected (Park, 2009).

Aligning the sequenced reads in ChIP-seq datasets should take into account the mismatches resulting from the sequencing errors, single-nucleotide polymorphisms (SNPs), and other potential differences between the sequenced genome and the reference genome. Several algorithms for aligning have been developed which include Eland by Illumina, Mapping and Assembly with Qualities (MAQ) (Li et al., 2008), and Bowtie, which is a very fast aligner based on a file compression algorithm (Langmead et al., 2009). Many mapping pipelines discard

non-unique tags, which are mapped to more than one genomic location (multiple hits), and only report the number of unique alignments to the reference genome; however, stringent mapping pipelines have been recently introduced for aligning the repetitive elements (Day et al., 2010).

After identifying the uniquely aligned DNA fragments, the statistically significantly enriched genomic regions relative to the control can be identified by the 'peak calling' algorithms. Enriched regions of different types can be found, such as sharp, broad, and mixed. Sharp peaks are largely associated with histone modifications or chromatin-binding proteins at regulatory elements and the broad regions are mainly found across repressed or transcribed domains. Earlier peak calling algorithms (Kharchenko et al., 2008, Mikkelsen et al., 2007, Rozowsky et al., 2009) were mainly designed to identify the sharp peaks, whereas the broad peaks were located by merging the adjacent peaks *post hoc*. Overall, an ideal peak calling algorithm should take into account the strand-specific pattern at any binding location, correct for the local variation shown by input DNA, adjust for sequence alignability, and measure the expected ratio of incorrectly discovered sites to those that are found to be significant by the FDR. Validation of a large set of sites by qPCR can be used to measure the accuracy of a peak caller (Park, 2009, Storey and Tibshirani, 2003).

In order to identify the overlap between the identified peaks between multiple samples or to discover the overlap between different protein binding sites and a defined genomic feature, such as a regulatory element, the extracted signals from peak callers are compared. Furthermore, by calculating the number of events that are close enough between multiple datasets, overlapping regions could be identified. This complex downstream analysis has several limitations which include a bias towards the peak-calling step and the empirical threshold values used to distinguish between real signal and noise, and complexity of comparing datasets with different binding patterns, such as sharp (H3K4me3) or broad enrichments (H3K27me3), where peak calling is more difficult. Therefore, to facilitate qualitative and quantitative comparisons between multiple ChIP-seq datasets at a reference set of genomic sites, seqMINER *k*-means clustering tool was developed that requires uniquely aligned tags to compute a density array over a defined window around the reference coordinates. Furthermore, the density array is normalised linearly in each sample to correct for inter-sample variations in sequencing depth.

The normalised values are used for the clustering step and graphically represented through heatmaps (Ye et al., 2011).

In this Chapter, the sequenced tags were aligned to the reference human genome (RefSeq hg18 build) using the Bowtie algorithm. Subsequently, the unique alignments were clustered in seqMINER at the annotated gene promoters in all four cell types for H3K4me3 and H3K27me3 histone modifications. The subgroups of genes associated with distinct histone modification patterns were isolated for further investigations, such as pathway analysis using the PANTHER database.

5.4 Aims of the Chapter

The patterns of histone modifications and their relationship to gene expression profiles will be examined in this Chapter.

1. To establish the genome-wide signatures of H3K4me3 and H3K27me3 histone modifications at the annotated gene promoters in HSCs, HPCs, LSCs and LPCs by low-cell ChIP-seq.

2. To identify the epigenetic mechanisms that regulate HSC and HPC identity pathways in normal and leukaemic cell types.

3. To establish the link between observed expression changes and the underlying epigenetic programming in normal and leukaemic cell types.

4. To investigate the gene expression levels of chromatin modifiers that add or remove H3K4me3 and H3K27me3 to try and account for the differences in histone modification levels between normal and leukaemic cell types.

Results

5.5 Overall strategy

Initially, the conventional ChIP-seq method was optimised for the low-cell ChIP assays. The low-cell ChIP-seq method was used to generate histone modification patterns of HSCs and LSCs, while the conventional ChIP-seq method was used in the case of LPCs. The histone modification patterns of HPCs was obtained from

publicly available dataset (Cui et al., 2009). In order to identify the epigenetic differences between CML cells and their normal counterparts the promoter histone modification signatures in each cell type were identified and characterised. Subsequently, the relationship between the promoter signatures and the pathways that were identified to regulate stem cell maintenance and proliferation was established. Finally, the changes in promoter signatures were investigated in response to changes in gene expression between the cell types.

5.6 Optimisation of low-cell ChIP-seq library preparation

Initially, the low-cell ChIP was employed to investigate the distribution of H3K4me3 and H3K27me3 marks in the HSC1 sample. The ChIP-purified DNA was subsequently coupled to deep sequencing (ChIP-seq) using the Illumina Genome Analyzer platform. The ChIP-seq libraries were generated using the original Illumina method in which library size-selection was performed prior to PCR amplification. After aligning the sequencing reads to the reference human genome, only 9% of the reads could be aligned (Figure 5.1). Conversely, 63% of the sequencing reads were aligned to the mouse genome and the rest (28%) were associated with an unknown source (Figure 5.1). Further analysis of non-mappable reads that were aligning to the mouse genome demonstrated that the contaminating reads were evenly distributed across the mouse genome at a low tag density with no bias for specific gene features. Therefore, the source of contaminating sequences was mouse genomic DNA.

Moreover, from the 9% of sequencing reads that were aligned to the human genome, only two thirds of them (6% of total reads) could be aligned uniquely to one location in the genome (Figure 5.1). The rest of the reads constituted repetitive sequences that could not be uniquely aligned. Furthermore, two thirds of the uniquely aligned reads (4% of total reads) were found to be PCR duplicates which could not be reliably used in downstream analysis (Figure 5.1).

Sample name	Total reads	Reads aligned to human genome	Multiple hits	Unique reads	Duplicate alignments	Total non-mappable reads	Non-mappable reads aligned to mouse genome	Non-mappable reads with unknown source
HSC1 H3K4me3	24,769,487	1,856,630	661,339	1,195,321	796,880	22,912,857	15,619,639	7,293,218
HSC1 H3K27me3	25,742,339	2,756,680	1,079,645	1,677,035	1,118,023	22,985,659	16,217,674	6,767,985
Average	25,255,913	2,306,655	870,492	1,436,178	957,452	22,949,258	15,918,656	7,030,602
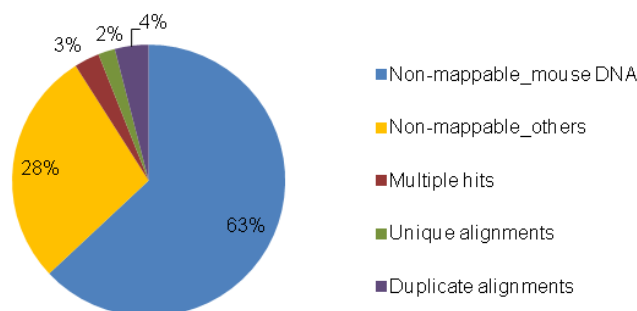
Average total number of reads: ~25 million

Figure 5.1: Contaminating mouse genomic DNA and PCR duplicates affect the quality of low-cell ChIP-seq output.
Top: table shows the number of sequencing reads obtained from low-cell ChIP-seq assays using in which gel size selection was performed prior to PCR amplification. Table also indicates the number of sequencing reads that were aligned to human genome, the number of multiple hits, the number of unique alignments, the number of PCR duplicates, and the number of non-mappable sequencing reads aligning to mouse genomic DNA or with an unknown source. Bottom: the average results from the two low-cell ChIP-assays are illustrated as a pie chart. The average total number of sequencing reads was ~25 million, 63% of which was aligned to the mouse genome and only 9% could be aligned to the human genome. A third of the sequencing reads (3% of total reads) that aligned to the human genome were multiple hits that comprised repetitive sequences. Two thirds of the uniquely aligned reads (4% of total reads) were also derived from PCR duplicates.

The mouse genomic DNA contamination could have been introduced as a result of cross contamination by using a probe sonicator, or by using contaminated ChIP reagents. In order to tackle this issue, a separate probe sonicator was used for the human samples and all the ChIP reagents were renewed and kept separate from mouse genomic DNA samples. Furthermore, in order to ensure that the above measures could successfully eliminate the mouse genomic DNA contamination, qPCR primers were designed at multiple mouse genomic sites which did not show sequence homology with the human genome. The contamination level was monitored at two stages; first, after the purification of ChIP DNA and second, after the preparation of ChIP-seq libraries (Figure 5.2B). After the ChIP-qPCR the percentage ratio of mouse DNA to the human DNA was measured to determine the level of contamination in each sample. For human ChIP amplifications a region of genome enriched with the histone modification of interest was chosen based on the previously reported histone modification enrichments at the TAL1/SCL locus

(Dhami et al., 2010). The sequenced libraries did not exceed 5% mouse genomic contamination before and after library preparations as shown for the typical examples in Figure 5.2B. The templates or the PCR-amplified libraries that had more than 5% contamination were discarded before sequencing.

Furthermore, the presence of a large proportion of PCR duplicates (up to 60%) was previously reported in libraries derived from low-complexity templates (Quail et al., 2008). Low-cell ChIP from 10,000 cells was estimated to generate ~30 pg of DNA (Acevedo et al., 2007). In the original Illumina ChIP-seq library preparation method, the adapter ligated ChIP-DNA were subjected to gel electrophoresis and size-selection at ~200bp. Therefore, by using small amounts of DNA as ChIP-seq library template, the size-selection step purifies even lower amounts of DNA, which is referred to as a low-complexity template. Consequently, during the PCR amplification of ChIP-seq library, the low-complexity template would leave the exponential phase in early cycles as there would be fewer number of template strands available. Therefore, PCR amplification of a low-complexity template would increase the incidence of PCR product duplicates. In order to tackle this issue, the adapter ligated low-cell ChIP template was directly PCR amplified without reducing it complexity further. Subsequently, the size selection was performed at 200-300 bp range (Figure 5.2A), which comprised 92 bp of ligated adapters and 108-208 bp of ChIP DNA insert. This would ensure the elimination of unincorporated adapters or adapter dimers from the sequencing library in addition to maintaining the optimum library size for cluster generation (reviewed in Section 5.1.1).

Through optimising the above steps and careful handling of the low-cell ChIP material during library preparation, the percentage of uniquely aligned sequencing reads to the human genome without PCR duplicates increased from 2% to 35% (Figure 5.2D). However, on average 40% of the sequencing reads could still not be aligned to the human genome due to the presence of contaminating DNA material or sequencing errors. The conventional ChIP-seq assays performed in Chapter 4 showed an average of ~10% non-mappable sequencing reads which could be due to the sequencing errors or the presence of primer dimers. Furthermore, ~10% of the total number of sequencing reads resulted from PCR duplicates (Figure 5.2C). Therefore, in comparison with the conventional ChIP-seq results, the low-cell ChIP-seq datasets demonstrated significantly reduced levels

of PCR duplicates, although the presence of non-mappable reads was still an issue. However, from an average total number of 30 million reads obtained from each low-cell ChIP-seq assay, 10.5 million sequencing reads (35%) could be uniquely aligned to the human genome, which could provide sufficient depth for performing downstream analyses.
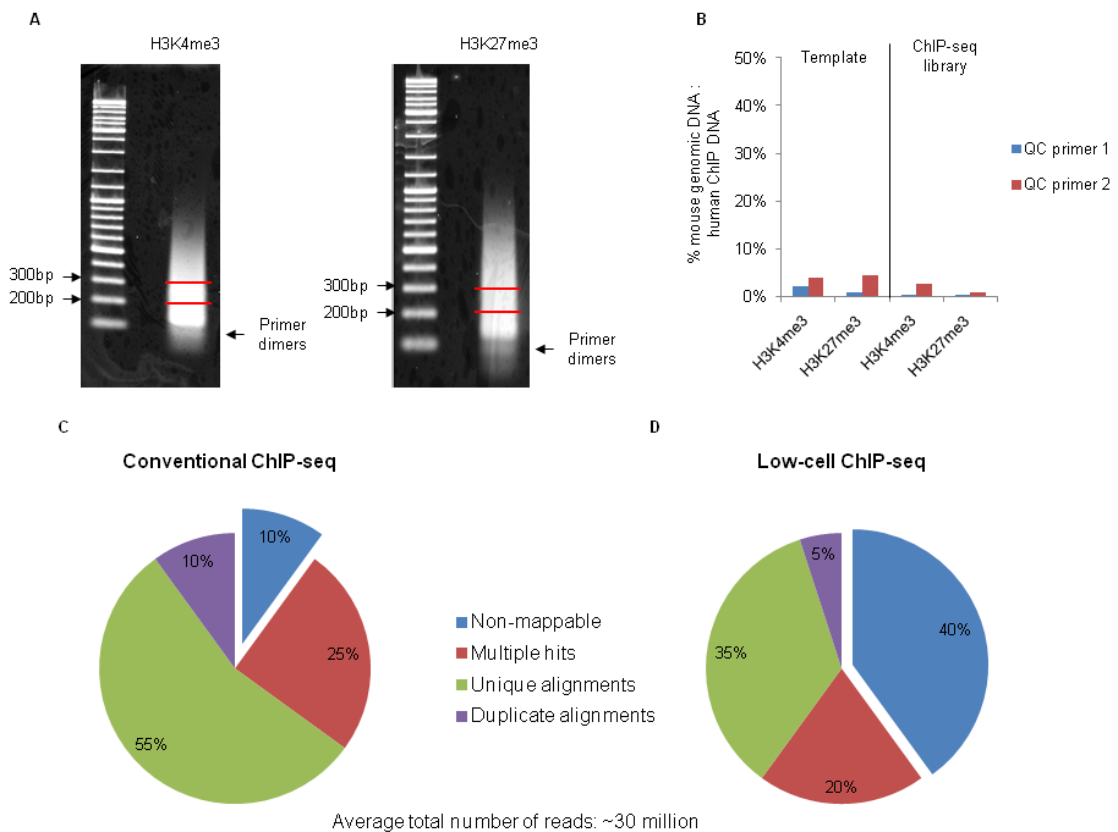
Figure 5.2: The optimised low-cell ChIP-seq method.
(A) Typical examples of gel electrophoresis of PCR-amplified ChIP-seq libraries for two histone modifications H3K4me3 and H3K27me3. The ChIP-seq libraries were found to be in the range of 100 bp – 1 kb with the majority of DNA fragments at the 100-400 bp range. The <100 bp bands represent un-incorporated primers and adapters. Size selection was performed at 200-300 bp, shown by the red lines. First lane in each gel shows the 1 kb DNA ladder. The electrophoresis was performed on 2% 1X TAE buffer agarose gels and visualised by ethidium bromide staining. (B) Typical examples showing the percentage ratio of mouse genomic DNA to the human ChIP DNA as measured by performing qPCR at various mouse genomic regions. The percentage of DNA contamination was below 5% before and after the ChIP-seq library preparation. (C) The percentages of non-mappable reads (blue), multiple alignments (red), unique alignments (green), and PCR duplicates (purple) on average in conventional ChIP-seq, and (D) low-cell ChIP-seq datasets from an average 30 million sequencing reads.

5.7 Genome-wide mapping of H3K4me3 and H3K27me3 in HSCs, HPCs, LSCs, and LPCs

H3K4me3 and H3K27me3 are well-studied in the context of gene promoters and predicting the transcriptional activity of their marked loci (Barski et al, 2007). Furthermore, these histone modifications are associated with bivalent promoters which have an important role during development (Azuara et al., 2006; Bernstein

et al., 2006; Section 1.5.4.6). Therefore, it was important to understand their role in gene expression regulation in the course of leukaemic transformations.

After the optimisation of the low-cell ChIP-seq method, the genome-wide distribution of H3K4me3 and H3K27me3 histone modifications were studied in HSCs and LSCs using ~10,000 cells/IP. The histone modification maps for LPCs were generated by conventional ChIP-seq assays since the cell numbers were not limiting (~100,000 cells/IP). The epigenetic profiles were obtained from the same samples and cell populations that were used to analyse the genome-wide gene expression profiles in Chapter 3. However, the histone modification maps for the HPCs were obtained from the previously published conventional ChIP-seq datasets that were generated from a pool of normal HPCs (CD133$^+$ CD34$^+$ CD38$^+$) (Cui et al., 2009). After aligning the sequencing reads to the human genome, the unique alignments without PCR duplicates (Table 5.1) were used to cluster all the annotated promoters (n=33,626) in the human genome (RefSeq hg18 build) by seqMINER. The promoters were defined as 2.5 kb regions upstream and 2.5 kb regions downstream of the TSSs. The clusters were generated graphically through heatmaps (Figure 5.3).

Sample name	Histone modification	Total reads	Reads aligned to human genome	Multiple hits	Unique alignments without duplicates
HSC1	H3K4me3	63,461,941 †	18,398,835	6,876,946	9,341,848
HSC1	H3K27me3	32,542,162	12,816,958	4,501,010	7,578,757
HSC2	H3K4me3	31,174,635	23,011,157	7,393,544	13,973,655
HSC2	H3K27me3	28,340,002	21,599,394	6,558,474	14,427,179
HSC3	H3K4me3	33,360,258	11,663,794	4,133,965	7,065,038
HSC3	H3K27me3	28,880,717	15,111,977	4,882,554	9,712,432
LSC1	H3K4me3	63,780,018 †	17,611,300	28,725,316	9,923,145
LSC1	H3K27me3	32,776,847	18,111,830	6,136,001	8,285,004
LSC2	H3K4me3	28,030,676	17,252,948	6,089,775	10,506,371
LSC2	H3K27me3	27,814,395	23,191,059	7,235,164	14,901,509
LSC3	H3K4me3	28,702,247	16,162,797	5,451,088	8,829,733
LSC3	H3K27me3	27,622,366	21,398,632	7,057,697	11,489,945
LPC1	H3K4me3	32,188,685	21,332,150	6,465,923	12,465,500
LPC1	H3K27me3	33,039,878	18,289,788	6,384,753	10,828,982
LPC2	H3K4me3	33,549,637	15,083,263	5,870,757	8,284,141
LPC2	H3K27me3	30,461,443	16,212,416	6,285,829	9,038,837
LPC3	H3K4me3	33,366,697	28,638,182	8,820,411	17,947,385
LPC3	H3K27me3	33,059,304	29,942,850	8,340,095	20,032,195
HPC‡	H3K4me3				2,477,162
HPC‡	H3K27me3				9,831,630

Table 5.1: The number of aligned sequencing reads obtained from H3K4me3 and H3K27me3 ChIP-seq assays in HSCs, LSCs, LPCs and HPCs.
The total number of sequencing reads, the number of sequencing reads aligned to the human genome, the number of unique alignments without PCR duplicates and the number of multiple hits are listed for each ChIP-seq assay. HSCs and LSCs were subjected to low-cell ChIP-seq, whereas HPCs and LPCs were subjected to the conventional ChIP-seq. ‡ The ChIP-seq datasets for HPCs were obtained from publicly available database (Cui et al., 2009) and the number of unique alignments without duplicates could be reported. † The results obtained by merging two independent low-cell ChIP-seq datasets.

5.8 defining H3K4me3 and H3K27me3 epigenetic signatures for promoters

The seqMINER clustering identified three classes of promoters based on the H3K4me3 and H3K27me3 signatures. These three classes of promoters included (i) H3K4me3, (ii) bivalent (H3K4me3 and H3K27me3), and (iii) neither histone modifications (Figure 5.3). Subsequently, further clustering of the neither modification category resulted in the separation of two subclasses of H3K27me3^{hi} promoters and promoters that harboured neither modifications. Therefore, in total four classes of promoters were distinguished, including (i) H3K4me3, (ii) bivalent, (iii) H3K27me3^{hi}, and (iv) neither modifications (Figure 5.3). Furthermore, these

clusters were generated by clustering all three bioreplicates for each cell type together as opposed to clustering on a sample by sample basis.
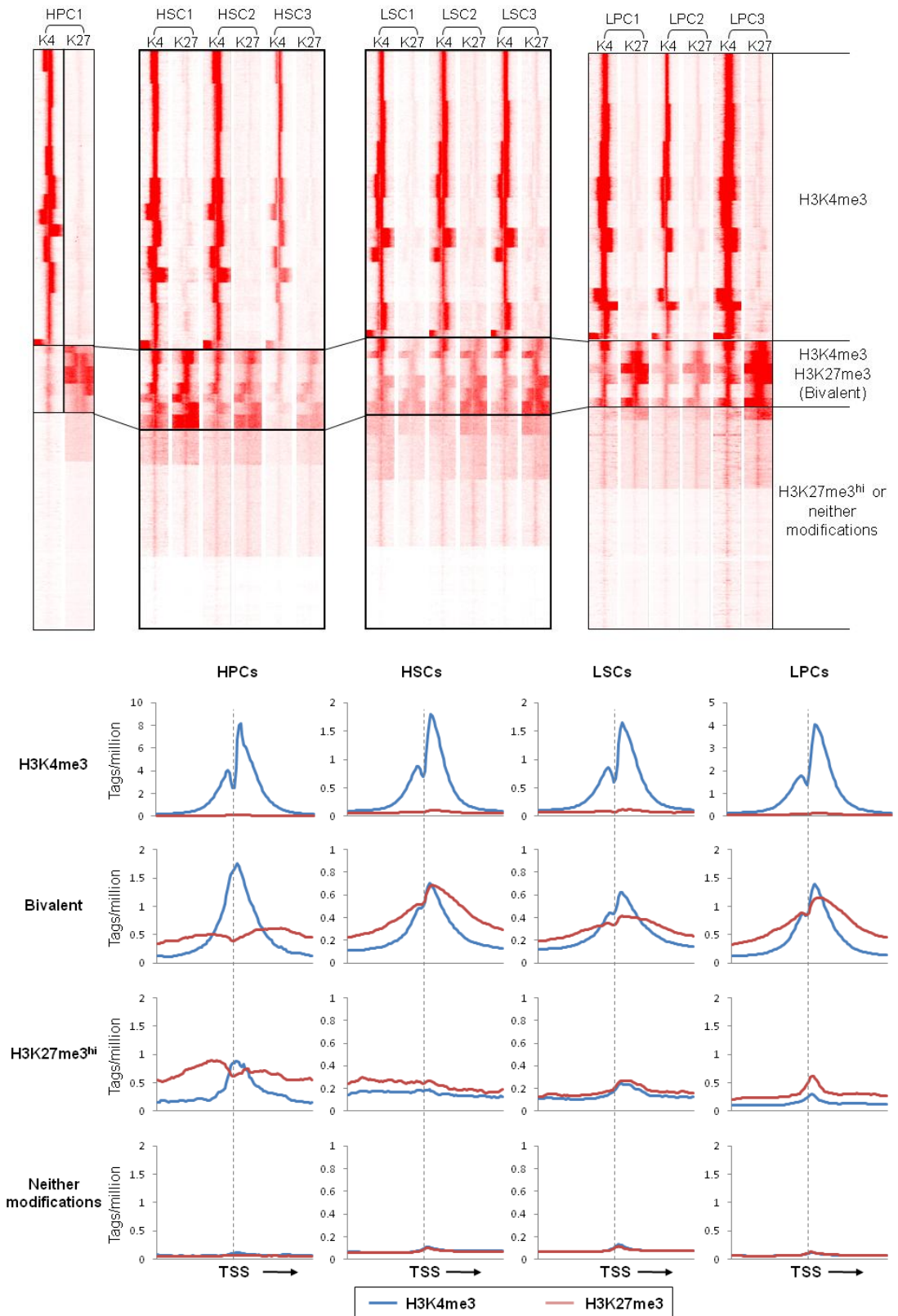
Figure 5.3: Clustering heatmaps and composite plots of H3K4me3 and H3K27me3 distribution at the annotated promoters.

Top: the seqMINER-generated heatmaps demonstrate the distribution of H3K4me3 and H3K27me3 histone modifications at the promoters of the annotated transcripts in the human genome (RefSeq hg18 build) in HPCs, HSCs, LSCs, and LPCs. Each heatmap represents six lanes illustrating H3K4me3 followed by H3K27me3 in three biological replicates (patient samples), except the HPCs for which data was obtained from previously published datasets (Cui et al, 2009) comprising only one pooled sample. Heatmaps could identify three major classes which include H3K4me3, bivalent (H3K4me3 and H3K27me3), and H3K27me3[hi] or neither modifications promoters. Further sub-clustering of the latter resulted in two separate classes of H3K27me3[hi] and neither modifications promoters. Bottom: the composite plots for each major class are shown for each cell type based on the normalised averaged number of tags across the 5 kb region centred at the TSSs for H3K4me3 (blue line) and H3K27me3 (red line). The x-axes show the average genomic coordinates where the orientation of transcription is shown by an arrow and the position of the TSSs is marked by vertical dotted lines. The y-axes show the number of tags per million at each genomic location.

The meta-gene profiles, which are referred to as composite plots in this Chapter, were created for each class of promoters that were defined by clustering. The composite plots were generated by allocating 100 bins across the defined promoters – 2.5 kb upstream and 2.5 kb downstream of TSSs – and computing the normalised averaged number of sequencing reads, which are also referred to as tags, per bin in all the genes represented in a promoter class. The final composite plots showed an average of three bioreplicates in each cell type except for the HPCs which represented only one pair of datasets (Figure 5.3). These composite plots, therefore, represent an average distribution of histone modifications across an average promoter in a promoter class. The H3K4me3 promoters in all cell types were associated with significantly high enrichment levels of H3K4me3 right after the TSS, which could be located by a noticeable dip as a result of nucleosomal depletion (Bernstein et al., 2004, Lee et al., 2004). The difference between the tag density at the TSSs of H3K4me3 promoters in HPCs and LPCs with HSCs and LSCs is a result of different ChIP efficiencies. This phenomenon was demonstrated in Chapter 4 which showed higher enrichment levels for H3K4me3 in conventional ChIP-seq assays compared to the low-cell ChIP-seq assays. Therefore, the differences are observed here is not due to the biological differences between the progenitor samples and their stem cell compartment, but mainly due to the differences between low-cell and conventional ChIP-seq assays.

The bivalent class was distinguished by the presence of appreciable levels of both H3K4me3 and H3K27me3 marks at the promoters. The enrichment levels of H3K4me3 were much lower at the bivalent promoters but featuring a nucleosomal depletion dip at the TSS immediately followed by a downstream peak similar to the H3K4me3 promoters. The H3K27me3 enrichment pattern at bivalent promoters in

HSCs, LSCs, and LPCs highly resembled that of H3K4me3, with a dip at the TSS followed by a downstream peak, but featured a broader distribution in HPCs as noticed by a wider dip at the TSS and the absence of downstream enrichment peak. The H3K27me3^{hi} promoters showed the most varied patterns between the four cell types, but all displayed higher levels of H3K27me3 than H3K4me3 forming peaks at the TSSs with further extension of enrichment blocks over the gene bodies, which distinguished this class of promoters from the bivalent class. Neither modifications category showed barely detectable levels for both histone modifications which formed very small peaks of enrichment over the TSSs. This class of promoters, primarily devoid of H3K4me3 or H3K27me3, is associated with DNA methylation (Meissner et al., 2008). These four classes of promoters were also previously reported in genome-wide histone modification analyses of normal haemopoietic cells (Cui et al., 2009, Weishaupt et al., 2010). Therefore, the results presented above were in agreement with the previously reported histone modification signatures at the gene promoters.

Since the HPC profiles were obtained from previously published datasets (Cui et al., 2009), a comparison between the number of bivalent genes identified by seqMINER in this section and the previously reported SICER analysis was performed (Figure 5.4). The seqMINER clustering identified 2,766 bivalent promoters whereas the SICER results indicated 2,602 promoters. 1,660 promoters were found in the overlap which comprised of approximately 65% of the previously reported promoters by SICER, whereas 942 promoters were only recognised as bivalent by SICER and instead 1,106 bivalent promoters were only identified as bivalent by seqMINER. The discrepancy could be explained by the SICER approach in which the H3K4me3 and H3K27me3 profiles were analysed independent of each other. Therefore, a bias towards the peak calling step could be expected as the broad H3K27me3 domains could be more difficult to distinguish as enrichment peaks in a given window (Zang et al., 2009). Conversely, seqMINER generated the clusters for both H3K4me3 and H3K27me3 marks simultaneously (Ye et al., 2011). Furthermore, different studies use different empirical threshold values to distinguish between real signal and noise.
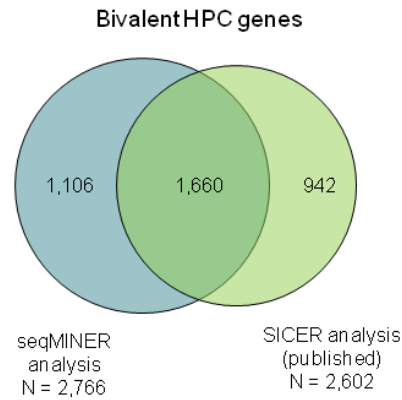
Figure 5.4: The comparison between the number of bivalent promoters identified by seqMINER with the published SICER analysis in HPCs.
seqMINER analysis identified 2,766 bivalent promoters in HPCs (blue circle), whereas the reported SICER analysis (green circle) identified 2,602 bivalent promoters in the same dataset. Both analyses agreed on 1,660 promoters. However, seqMINER identified 1,106 promoters that were not recognised by SICER and 942 promoters were only identified as bivalent by SICER.

5.9 The association of different promoter class with gene expression levels

It has been previously shown that each of these promoter classes is associated with different levels of transcription of their relevant gene loci (Mikkelsen et al., 2007). In order to confirm similar association between the promoter classes identified by seqMINER and the expression levels of their relevant loci, the average RMA (gene expression) values for the genes in each promoter category were extracted for each cell type from the Affymetrix datasets. The spread of RMA levels in each promoter class was reported as box plots (Figure 5.5).
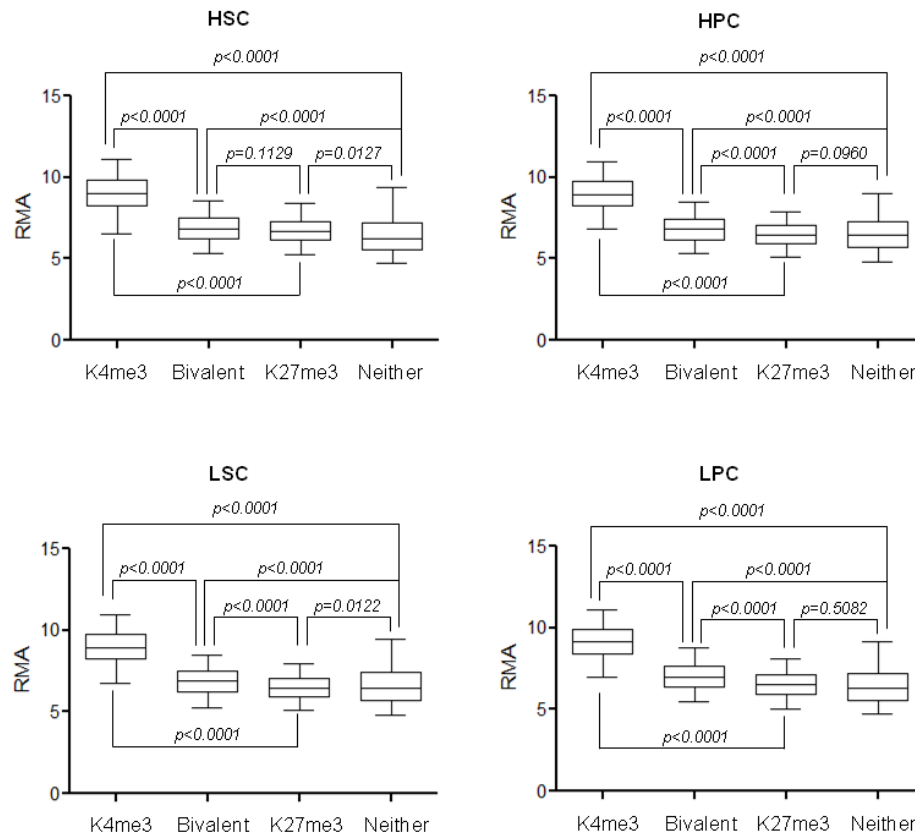
Figure 5.5: The relationship between each epigenetic category and the gene expression in each cell type.
Box plots represent the spread of the gene expression (RMA) values in each epigenetic category in each cell type. The whiskers represent 5-95 percentile with the medians shown as horizontal lines. The y-axes show the RMA values (gene expression) and the x-axes show different epigenetic categories. The p-values were calculated using unpaired, two-tailed t-test with 95% confidence intervals comparing the average RMA values between two compartments and $p < 0.05$ was considered to be significant.

The box plots showed that the genes in H3K4me3 compartment were associated with a significantly higher average of expression values ($p < 0.0001$) than the other three compartments in all cell types (Figure 5.5). This was in agreement with previous reports which showed a positive correlation between the H3K4me3 enrichments at the promoters and the transcriptional activity of the associated gene (Barski et al., 2007, Kim et al., 2005). On the other hand, bivalent promoters were associated with low transcriptional activity in four cell types (Figure 5.5) despite H3K4me3 enrichments, indicating that the repressive impact of PcG proteins at these promoters is dominant over the activity of TrxG proteins. Furthermore, the H3K27me3[hi] promoters, which are also regulated by PcG proteins, were associated with low gene expression levels (Figure 5.5). These promoters could be categorised as bivalent promoters with a higher H3K27me3 enrichments than H3K4me3, as observed in their composite plots (Figure 5.3). The similarities in the range of gene expression levels in both bivalent and

H3K27me3^{hi} promoters could suggest similar transcriptional regulation programming mediated by the PcG proteins. The relationship between bivalent and H3K27me3^{hi} promoters and low or repressed transcriptional activities was in agreement with previous reports (Boyer et al., 2006, Lee et al., 2006, Mikkelsen et al., 2007). The neither modification promoter category was also associated with repressed loci. However, the absence of PcG binding to these promoters could indicate a broader range of gene expression levels, as was observed with the spread of the box plots (Figure 5.5). However, previous reports indicated that up to 80% of these promoters can be associated with DNA methylation in ES cells (Meissner et al., 2008), which could explain for their repressed gene expression status. Overall, the gene expression levels associated with each promoter class in each of the four cell types were in agreement with previous reports, suggesting that the clusters were generated accurately.

5.10 Validation of ChIP-seq data by ChIP-qPCR

Prior to performing further analysis on the ChIP-seq datasets, the observed histone modification signatures were validated using ChIP-qPCR approach, as described in Chapter 2. A panel of 67 promoters representing different promoter classes were selected. The gene promoters were selected from three pathways that were of special interest to this research group, which include TGF-β signalling pathway, PI3K signalling pathway and the intracellular signalling downstream of BCR-ABL1 oncoprotein. The validation panel was comprised of 49 putative H3K4me3 promoters, 13 putative bivalent promoters, and five putative H3K27me3^{hi} or neither-modifications promoters in HSCs (Figure 5.6) as defined by seqMINER clustering. The qPCR results were represented as the log$_2$ fold enrichment levels of H3K4me3 or H3K27me3 enrichments at the tested promoters relative to the median of H3K4me3 and H3K27me3 enrichments of negative control regions, which were selected as intergenic regions within the TAL1/SCL locus (Dhami et al., 2010). Figure 5.6 demonstrates the average of the log$_2$ fold enrichments for each tested promoter as measured for all three HSC bioreplicates. The criteria used for passing the validation step were as follows: H3K4me3 promoters to be associated with significantly positive H3K4me3 enrichment and significantly negative enrichment of H3K27me3, bivalent promoters to be associated with significantly positive enrichment for both H3K4me3 and H3K27me3, and the promoters with H3K27me3^{hi} or neither modifications with only

significantly positive H3K27me3 enrichment or significantly negative enrichment for both marks, respectively. A total of 9 promoters failed to validate by ChIP-qPCR (marked by asterisks in Figure 5.6) and therefore, the validation rate of the ChIP-seq dataset was 87% suggesting a high confidence in sequencing and clustering results.
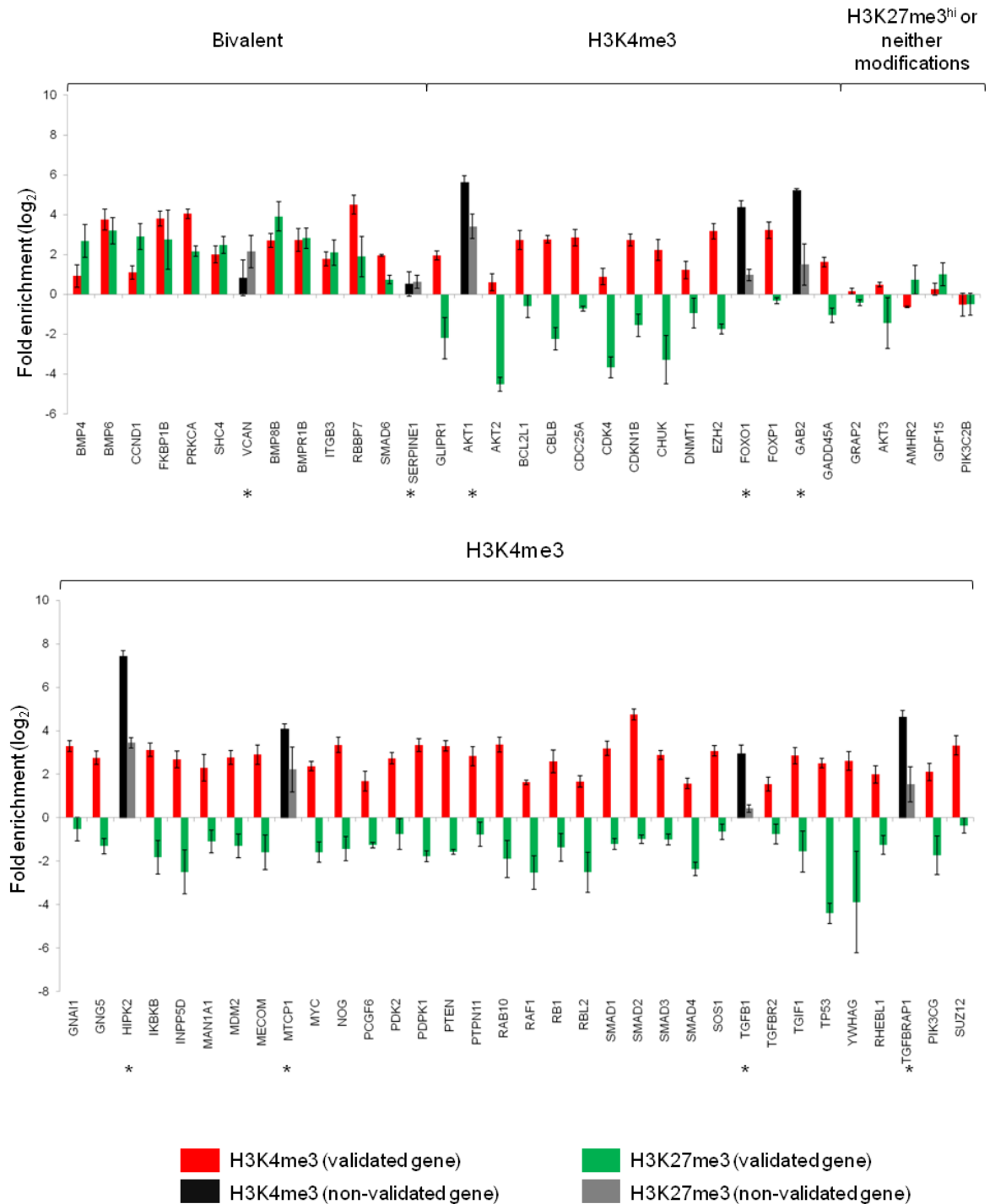
Figure 5.6: Validation of ChIP-seq datasets by ChIP-qPCR in HSCs.
A panel of 67 gene promoters from H3K4me3, bivalent, and H3K27me3 or neither-modifications categories were selected for validation by qPCR in HSCs. The height of the bars (y-axis) shows the log_2 fold enrichment levels relative to the median of negative control enrichments averaged from

three individual samples. The positive log$_2$ fold enrichments indicate the presence of an enrichment and the negative values indicate their absence. The error bars show the standard error of mean. The H3K4me3 and H3K27me3 enrichments which passed the validation are illustrated as red and green bars, respectively, whereas the H3K4me3 and H3K27me3 enrichments which could not be validated are displayed as black and grey bars, respectively. The promoters which did not pass the validation step are marked with asterisks (n=9).

5.11 CML cells show elevated levels of bivalent promoters

As discussed in Section 5.7, the H3K27me3^{hi} category of promoters is a subset of the bivalent class with higher H3K27me3 levels than H3K4me3 levels. Therefore, The genes identified in these two categories in all four cell types were collectively referred to as bivalent/H3K27me3^{hi} class of promoters. Previous reports have indicated that the pluripotent cells, such as ES cells, are associated with a large number of bivalent genes, which are significantly reduced in the course of differentiation (Mikkelsen et al., 2007). A large subset of bivalent promoters have also been reported in HSCs, whereas the more differentiated cell types, such as MEPs and T cells, are associated with fewer bivalent promoters (Cui et al., 2009, Weishaupt et al., 2010). Therefore, less mature cell types are associated with a large number of bivalent promoters. Consequently, the number of transcripts and genes associated with each promoter class in each cell type was investigated (Figure 5.7).
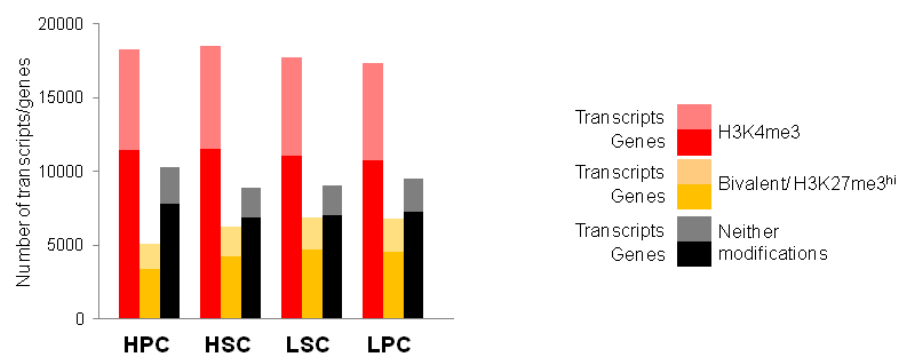
Cell type	H3K4me3		Bivalent/H3K27me3^{hi}		Neither modifications	
	Transcripts	Genes	Transcripts	Genes	Transcripts	Genes
HSC	18,477	11,496	6,248	4,249	8,901	6,848
HPC	18,274	11,408	5,050	3,373	10,302	7,829
LSC	17,720	11,034	6,847	4,725	9,059	7,053
LPC	17,337	10,771	6,788	4,533	9,501	7,282

Figure 5.7: Number of transcripts or genes in different promoter categories in each cell type. The number of transcripts or genes associated with H3K4me3 (red), bivalent/H3K27me3^{hi} (yellow), and neither histone modifications (black) are shown in each cell type as a bar chart (top) and a table (below).

The number of bivalent/H3K27me3^{hi} promoters showed a 20% drop in HPCs (n=5,050) relative to HSCs (n=6,248). However, the number of bivalent/H3K27me3^{hi} promoters was elevated by approximately 10% in LSCs (n=6,847) and LPCs (n=6,788) relative to HSCs. The observation in HPCs was in agreement with the expected reduction of bivalent/H3K27me3^{hi} promoters in the course of differentiation. Furthermore, the observation of large number of bivalent/H3K27me3^{hi} promoters in LSCs could be expected due to their stem cell phenotype. However, the elevated levels of bivalent/H3K27me3^{hi} promoters were not in agreement with the progenitor status of LSCs. Consequently, the fate of many bivalent promoters were still undecided in LPCs which may confer a less mature phenotype to LPCs relative to their normal counterparts. Thus, it is plausible to propose profound differences in epigenetic programming between LPCs and HPCs.

5.12 Bivalent promoters are significantly enriched for the stem cell pathways

In order to understand the role of each promoter class in the regulation of biological pathways, the genes that were associated with each promoter category in each cell type were subjected to pathway analysis using PANTHER database. The pathway analysis was performed by providing a reference list that contained all the annotated genes in the RefSeq hg18 database (n=21,733). This reference list was divided into genes groups that constitute different biological pathways. Subsequently, the list of genes in each promoter class was provided as an input list, which was divided into similar categories of biological pathways. The input list was then compared against the reference list and a binomial test was applied to statistically determine how significantly a group of genes was over- or under-represented in the input list relative to the reference (Mi et al., 2010). The pathways that were significantly over-represented in a promoter class were assigned +1 values and the pathways that were significantly under-represented were assigned -1 values. The pathways that were not significantly over- or under-represented in a promoter class in a cell types were assigned 0 values. Furthermore, these values were used in a *k*-means clustering approach in order to group subset of pathways based on similarities in the epigenetic signatures of their promoters (Figure 5.8).
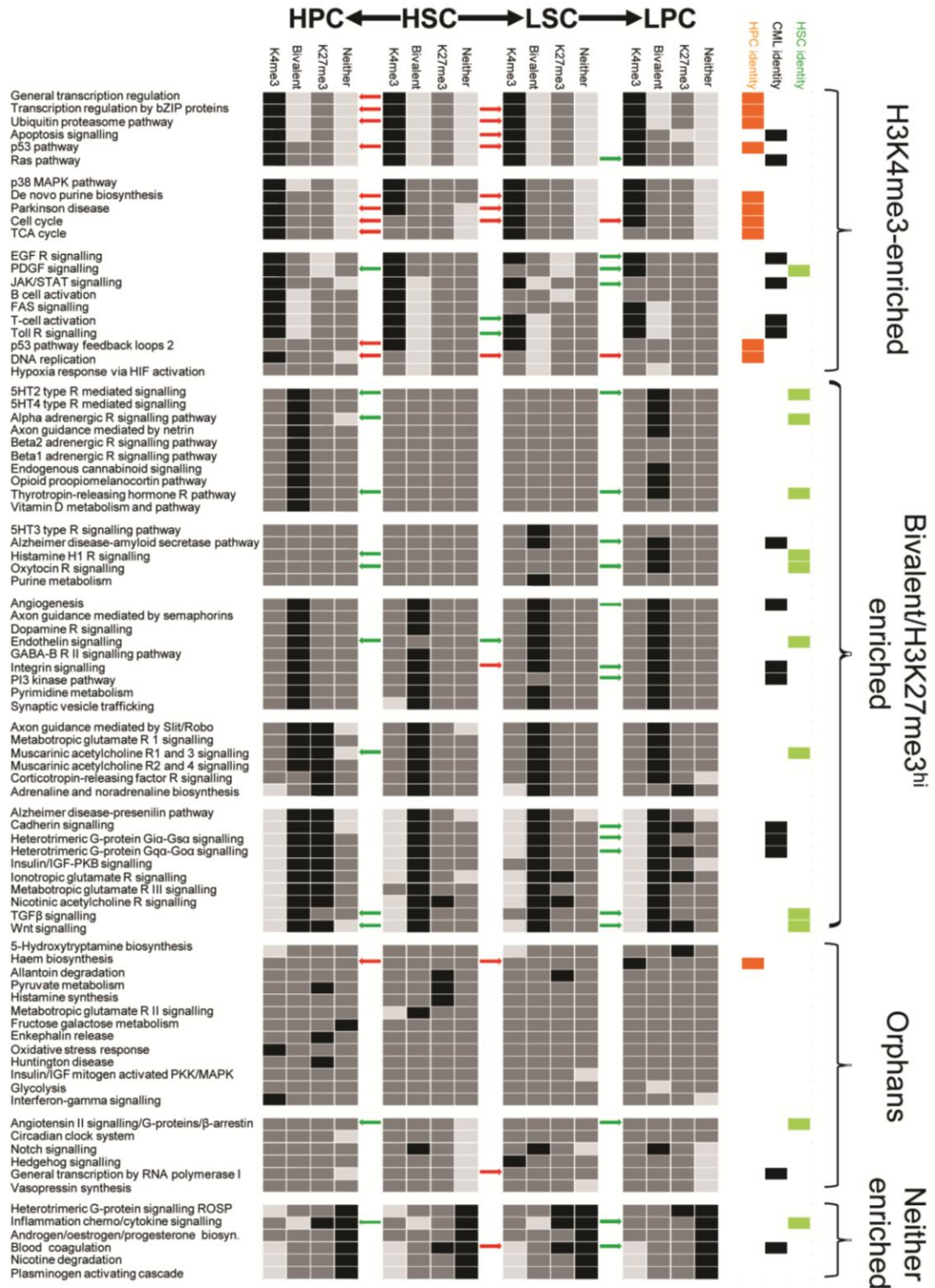
Figure 5.8: Clustering of the pathways significantly associated with each promoter class in the four cell types.
Pathways that were found to be significantly under- or over-represented in at least one promoter class in a cell type were clustered using a *k*-means clustering approach. Four major clusters of pathways were identified as labelled on the right. The name of each pathway is represented on the left. A black box indicates over-represented, grey box indicates expected representation and light grey box indicates under-represented in a particular promoter class. The pathways that were shown to be significantly downregulated or upregulated at the level of expression between the cell types (Chapter 3) were represented by green and red arrows, respectively, which indicate the orientation of the expression change. The pathways that were shown to be associated with HPC identity (orange boxes), CML identity (black boxes), and HSC identity (green boxes) (as defined in Chapter 3) are also indicated.

The clustering of significantly under- or over-represented pathways in each promoter class in four cell types identified four classes of pathways based on similarities in the epigenetic signatures of their promoters. The identified classes included H3K4me3-enriched, bivalent/H3K27me3^{hi}-enriched, neither modifications-enriched, and the orphans. The classes were labelled according to the pathways over-representation in a particular promoter class in the four cell types.

The H3K4me3-enriched class was characterised by the pathways that were pre-dominantly over-represented in the H3K4me3 compartment in the four cell and were under-represented in the bivalent/H3K27me3^{hi} or neither modifications categories dependent on the cell type. Intriguingly, 10 out of 11 HPC identity pathways which were associated with proliferative capacities, as identified in Chapter 3, were significantly enriched with this class of promoters. These included transcriptional regulation by bZIP proteins, the ubiquitin proteasome pathway, the p53 pathway and its regulatory feedback loops, the cell cycle, the TCA cycle, and DNA replication. Furthermore, six out of 15 CML identity pathways were also found to be enriched in this category, including apoptosis signalling, RAS pathway, EGF receptor signalling, JAK/STAT signalling, Toll-like receptor signalling, and T-cell activation mechanism. The pathways classified in this category were largely associated with upregulation between cell types, as indicated by red arrows in Figure 5.8. Therefore, it can be suggested that the H3K4me3 promoters have a predisposition for upregulation.

The bivalent/H3K27me3^{hi}-enriched category was comprised of the pathways that were only over-represented in the bivalent/H3K27me3^{hi} promoter class in a cell type or becoming significantly over-represented during the transition to the progenitor states (HPCs or LPCs). Interestingly, nine out of 12 HSC identity pathways, identified in Chapter 3, were found in this category, which included Wnt and TGF-β signalling as well as α-adrenergic receptor, serotonin 5HT2 receptors, acetylcholine muscarinic receptors 1 and 3, histamine H1 receptor, oxytocin receptor, TRH receptor and endothelin signalling pathways. Furthermore, seven out of 15 CML identity pathways were also found in this category, including angiogenesis, integrin signalling, cadherin signalling, heterotrimeric Gα-protein signalling, and PI3K pathway. The pathways identified in these two categories were largely associated with downregulation events between cell types, as

indicated by green arrows in Figure 5.8. Therefore, it can be suggested that the bivalent/H3K27me3^{hi} promoters have a predisposition for downregulation.

The neither modifications-enriched pathways were defined as the pathways mainly over-represented in the neither modifications promoter class, such as the chemokine/cytokine-mediated inflammation signalling, which was linked to the HSC identity, and the blood coagulation, which was linked to the CML identity, pathways. The orphans, on the other hand, included the pathways which did not share a common epigenetic signature between the four cell types. The orphans included several metabolic pathways as well as notch signalling, hedgehog signalling, and angiotensin II receptor signalling pathways.

The above observation indicated that the genes with bivalent promoters were significantly enriched for the HSC identity pathways, which were proposed to have a role in the maintenance of stem cell features (Chapter 3). However, 40 pathways were identified in the bivalent/H3K27me3^{hi} categories (Figure 5.8) from which only 16 pathways were demonstrated to be associated with the HSC or CML identities at the level of expression. Therefore, it is plausible to suggest that the other pathways that were not associated with a significant change of expression were also capable of using similar transcriptional regulation mechanisms during normal differentiation and leukaemic transformations and, consequently, might have a role in the maintenance of stem cell phenotype in both HSCs and LSCs.

5.13 Gene expression changes of bivalent genes and bivalent-enriched pathways are skewed towards upregulation in LSCs.

In order to elucidate the role of epigenetic programming in transcriptional regulation of different epigenetic classes of biological pathways (Figure 5.8), the significant genome-wide gene expression changes associated with each promoter class were studied in the following transitions: (i) HSCs to HPCs, (ii) HSCs to LSCs, (iii) LSCs to LPCs, and (iv) HSCs to LPCs (Figure 5.9).
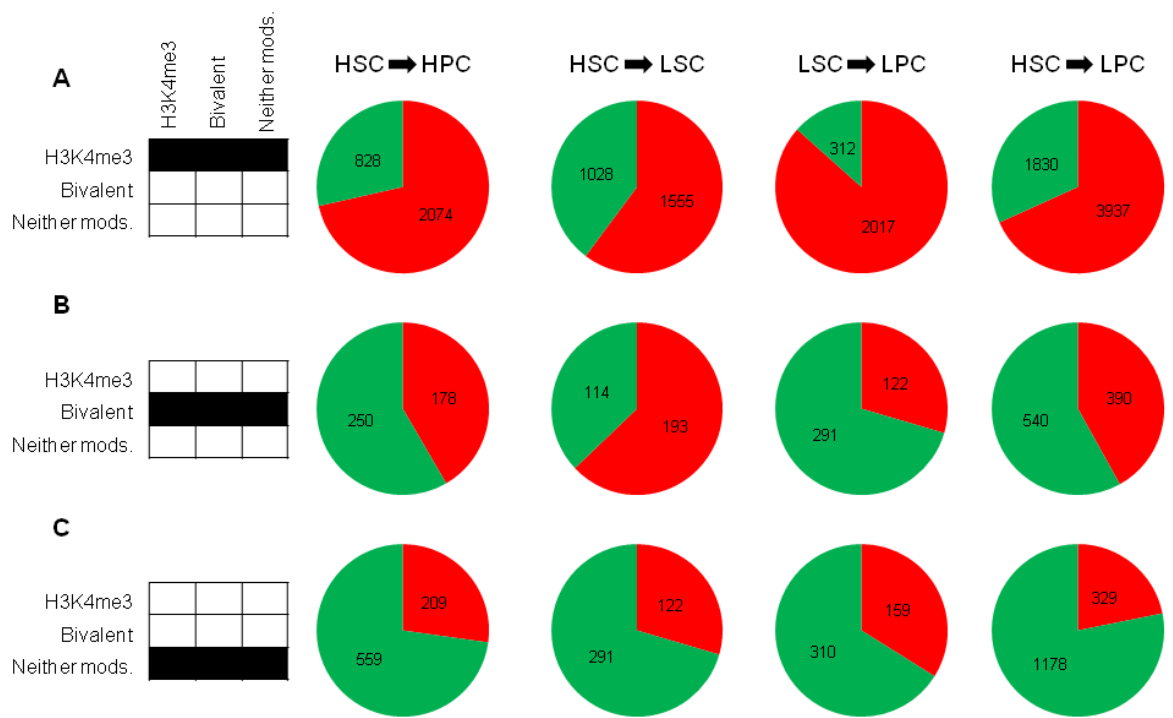
Figure 5.9: The association between each promoter class and genome-wide gene expression changes in the four studied transitions.
The gene expression changes associated with each promoter class during each transition are illustrated as pie charts. The status of the genes in the cell of origin in each transition is shown on the left with their potential outcomes in the new cell type. The number of significantly upregulated (red) and downregulated (green) genes are also shown. (A) Genes that were associated with H3K4me3 promoters in the cell of origin; (B) genes that were associated with bivalent/H3K27me3^{hi} promoters; and (C) genes that were associated with neither modifications.

The genes that were associated with H3K4me3 compartment in the cell of origin in each of the above transitions were associated with upregulation of gene expression more profoundly than the downregulation events (Figure 5.9A). However, the activating capacity of H3K4me3 promoters were reduced to 60% in the HSC-to-LSC transition in comparison to the normal differentiation (70%). Furthermore, LPCs regained the activating capacity similar to HPCs. This was in agreement with our earlier observation at the level of pathways, which showed that that the H3K4me3 promoters have a predisposition for upregulation (Section 5.10). Therefore, these findings further strengthen the requirement for the HPC identity pathways, which are associated with proliferation, to be enriched for the H3K4me3 promoters in order to be upregulated upon leaving the HSC compartment. Conversely, the gene promoters that were associated with neither modifications in the cell of origin in each of the above transitions favoured significant downregulation of gene expression over upregulation events (Figure 5.9C).

Furthermore, bivalent/H3K27me3^{hi} genes were associated with significant downregulation of gene expression more profoundly than the upregulation events in the HSC-to-HPC, LSC-to-LPC, and HSC-to-LPC transitions (Figure 5.9B). This observation was in agreement with the results at the pathway level which indicated that bivalent promoters have a predisposition for downregulation (Section 5.10). However, the HSC-to-LSC transition was linked to a skewed tendency for bivalent genes to be upregulated. The importance of this finding was in connection with the significant association between the bivalent promoters and the HSC identity pathways, which may have a role in the maintenance of stem cell features. Therefore, the tendency for downregulation in bivalent promoters during HSC-to-HPC, LSC-to-LPC, and HSC-to-LPC was in agreement with the significant downregulation of the HSC identity pathways in above transitions. However, the bivalent promoters are abnormally programmed during HSC-to-LSC transition which could have an impact in the regulation of the bivalent/H3K27me3^{hi} pathways in these cells. In order to further investigate the transcriptional regulation of the bivalent/H3K27me3^{hi} pathways in each of the above transitions, the number of significantly upregulated and downregulated genes in these pathways were analysed (Figure 5.10).
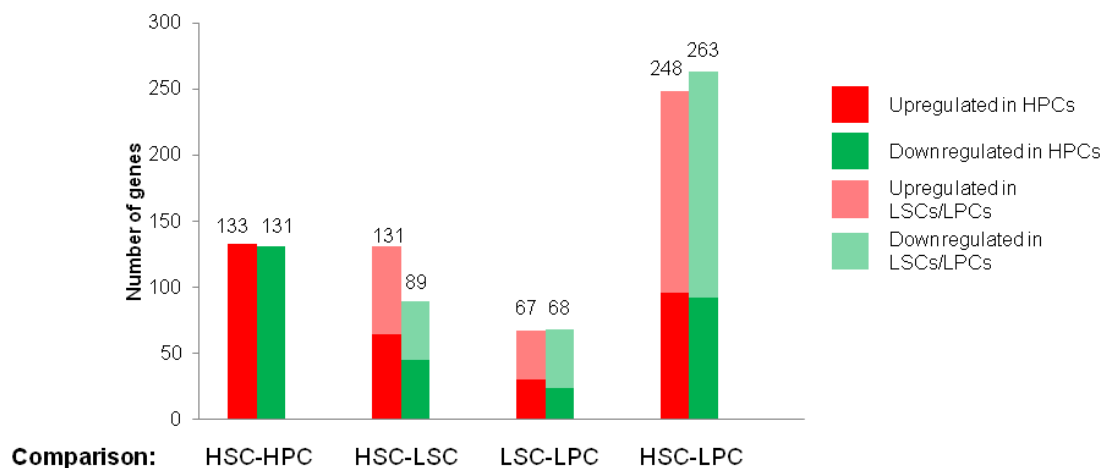
Figure 5.10: The analysis of significant expression changes for the genes of bivalent/H3K27me3^{hi} pathways in the four studied transitions.
The number of significantly (p<0.05) upregulated (red bars) and downregulated genes in the bivalent/H3K27me3^{hi} pathways in each of the studied transitions. The overlap between the significant changes in the HSC-to-HPC transition and the other three leukaemic transformations is shown as darker overlaid red and green bars. The total number of gene expression differences in each transition is also indicated on top of each bar.

The significant expression changes of the genes in the bivalent/H3K27me3^{hi} pathways during the HSC-to-HPC transition were balanced between upregulation

and downregulation events. However, the gene expression changes in these pathways were skewed towards upregulation in the HSC-to-LSC transition. This was in agreement with the earlier observations that indicated a tendency towards upregulation for the bivalent/H3K27me3^{hi} genes in this transition. Therefore, it is plausible to suggest a different epigenetic programming for the bivalent/H3K27me3^{hi} genes and pathways that is associated with reduced repressive potential and increased activation or upregulation capacities.

LPCs, on the other hand, demonstrated a tendency towards downregulation of the genes in bivalent/H3K27me3^{hi} pathways, suggesting that the epigenetic programming is more efficient in downregulating genes in this cell type.

5.14 Changes in H3K4me3 and H3K27me3 levels regulate the transcriptional activity of bivalent promoters in CML cells

The fate of the majority of genes associated with bivalent/H3K27me3^{hi} promoters was shown to be repression in the course of normal differentiation (Figure 5.9). However, during the HSC-to-LSC transition these promoters behaved differently with regards to the transcriptional regulation of their controlled loci (Figure 5.9). Therefore, it was important to examine whether these promoters were linked to a different epigenetic programming in LSCs and LPCs relative to their normal counterparts. Consequently, the levels of H3K4me3 and H3K27me3 histone modifications in 1 kb upstream and 1 kb downstream of TSSs of all bivalent genes in the four cell types were investigated. This was achieved by computing the total number of sequencing reads for both H3K4me3 and H3K27me3 marks at all the annotated promoters across a 2 kb window centred at their TSSs. The obtained values for each promoter were normalised against the total number of uniquely aligned reads without PCR duplicates for each of the H3K4me3 and H3K27me3 datasets in individual bioreplicates. Subsequently, the normalised H3K4me3 values were z-scored across the H3K4me3 and bivalent/H3K27me3^{hi} promoters for which this enrichment could be associated with a normal distribution. Furthermore, the normalised H3K27me3 values were z-scored for the bivalent/H3K27me3^{hi} promoters for which this enrichment was found to be normally distributed. Z-scoring was performed to normalise against differences in ChIP-seq efficiency between bioreplicates and between cell types. The results

were presented as the median of z-scored values of all the bioreplicates for each cell type, except HPCs that were associated with only one dataset.

In order to establish the differences in the levels of H3K4me3 and H3K27me3 at the bivalent promoters between LSCs/LPCs and their normal counterparts, the levels of these histone marks were initially examined at the promoters of a control gene set (n=200). This control gene set was defined as the bivalent genes that were not associated with significant differential expression between the four cell types. Furthermore, these genes were chosen based on their matched expression levels in HSCs with the significantly differentially expressed bivalent genes (Figure 5.11A). Therefore, the relationship between cell types with regards to the H3K4me3 and H3K27me3 levels at the control promoters was used as the baseline.

In order to establish the changes in the levels of H3K4me3 and H3K27me3 at the bivalent promoters during HPC commitment and leukaemic transformations, two subsets of genes were investigated. First, the bivalent genes that were significantly downregulated in HPCs and LSCs/LPCs relative to HSCs (n=182) and; second, the bivalent genes that were significantly upregulated in HPCs and LSC/LPCs relative to HSCs (n=138). The first subset of genes, which were downregulated upon leaving the HSC compartment, showed a significant increase in the levels of H3K27me3 repressive mark at the LSC promoters (Figure 5.11B). However, the H3K27me3 levels in HPCs remained significantly unchanged relative to HSCs. Similarly, the levels of H3K27me3 remained significantly unchanged between LSCs and LPCs. On the contrary, the changes of H3K4me3 levels at these promoters were not significantly different from the baseline changes (Figure 5.11B). Therefore, the significantly downregulated bivalent genes in LSCs/LPCs use a different epigenetic regulation from HPCs which is mediated by increasing the levels of H3K27me3 repressive marks in LSCs.

The analysis of second subset of genes that were upregulated in HPCs and LSCs/LPCs showed that the changes of H3K4me3 and H3K27me3 levels were not different from the baseline changes in both HPCs and LSCs (Figure 5.11C). However, LPCs demonstrated a significant reduction in the levels of H3K27me3 levels relative to HSCs, in the absence of a significant change of H3K4me3 levels. Therefore, the significantly upregulated bivalent genes in LSCs/LPCs use a

different epigenetic regulation from HPCs which is mediated by reducing the levels of H3K27me3 repressive marks in LPCs.

Overall, the bivalent genes that are differentially expressed upon leaving HSC compartment are associated with gains and losses of H3K27me3 in LSCs and LPCs, which are not observed in HPCs.
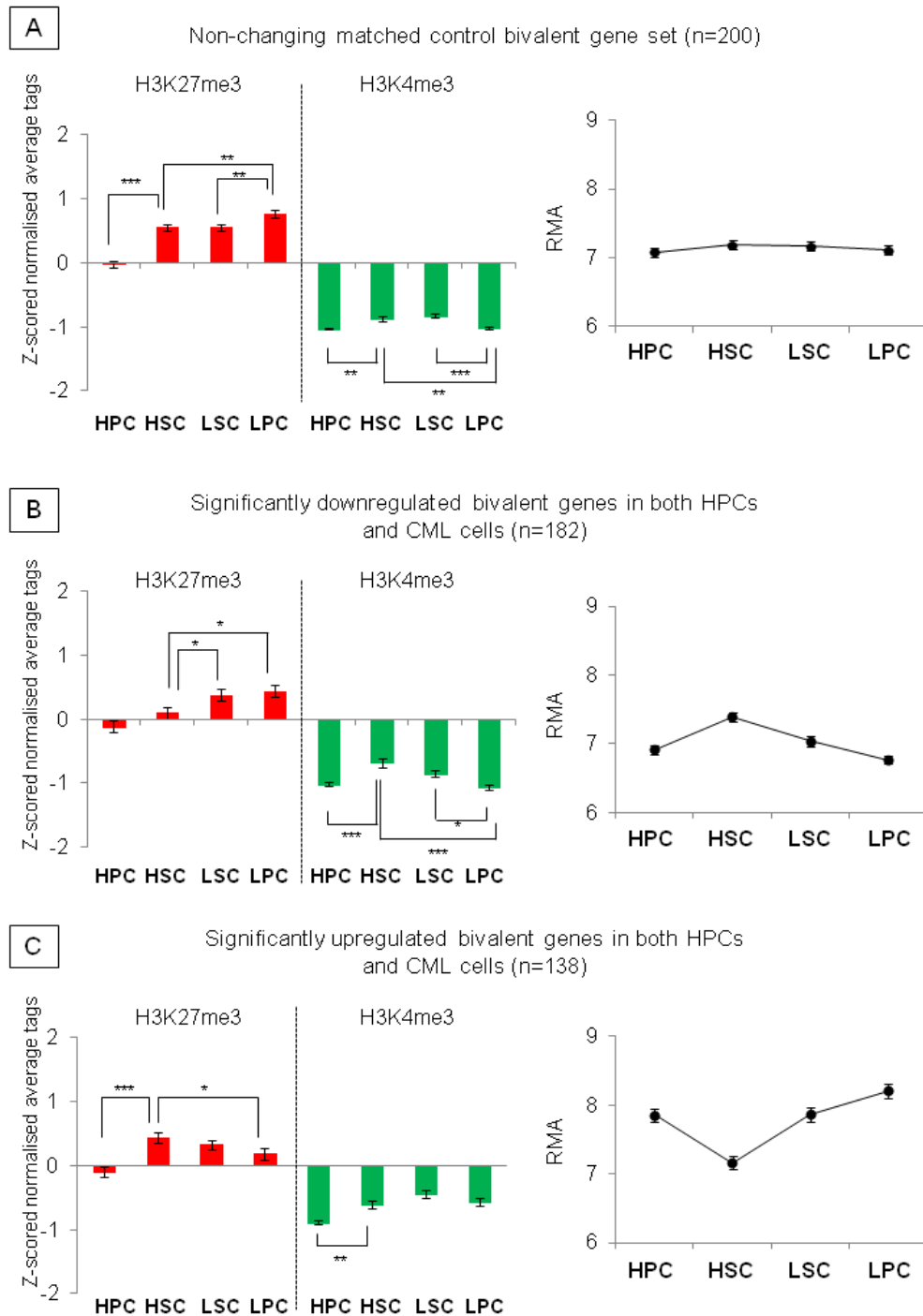
Figure 5.11: Regulation of H3K4me3 and H3K27me3 levels at the promoters of bivalent genes in the four cell types.

The average total number of sequenced tags for H3K4me3 and H3K27me3 marks over a 2 kb region centred at the TSS of all bivalent genes in the four cell types were computed and z-scored. The histograms demonstrate the z-scored levels of H3K4me3 (green) and H3K27me3 (red) at the promoters of (A) non-changing matched control bivalent gene set, (B) significantly downregulated bivalent genes in common between HPCs and LSCs/LPCs, and (C)) significantly upregulated bivalent genes in common between HPCs and LSCs/LPCs. The average expression levels of the genes investigated in each category were also expressed as RMA values. The error bars indicate standard error of mean. * $p<0.05$; ** $p<0.01$; *** $p<0.0001$. P-values were calculated by unpaired two-tailed t-test.

The above results indicated a different usage of H3K4me3 and H3K27me3 histone marks at the promoters of bivalent genes in LSCs/LPCs relative to their normal counterparts. However, these results did not indicate at what stage during leukaemic transformations the abnormal epigenetic programming was initiated. Therefore, the differentially expressed bivalent genes were examined during the HSC-to-LSC transition, referred to as type I changes, and during the LSC-to-LPC transition, referred to as type II changes (Figure 5.12A).

The type I significantly downregulated genes were associated with a significant increase in the levels of the repressive H3K27me3 marks and a significant reduction in the levels of the activating H3K4me3 marks (Figure 5.12B). Furthermore, the type I significantly upregulated genes showed a significant increase in the levels of H3K4me3, in the absence of a significant change in the levels of H3K27me3 (5.13C). Therefore, the type I differentially expressed genes were associated with the changes of both H3K4me3 and H3K27me3 marks in LSCs. Conversely, type II significantly downregulated and upregulated genes were not associated with a significant change of H3K4me3 and H3K27me3 levels between LSCs and LPCs (5.13D-E) relative to the observed baseline changes (5.12A).

Thus, the above results indicated that the abnormal epigenetic programming was initiated during the HSC-to-LSC transition. Furthermore, differential expression of bivalent genes in LSCs/LPCs was associated with the changes of both H3K4me3 and H3K27me3 histone marks, suggesting an abnormal mechanism of action for both PcG and TrxG complexes in LSCs/LPCs. Therefore, the expression levels of these protein complexes were investigated in the four cell types (Section 5.14).
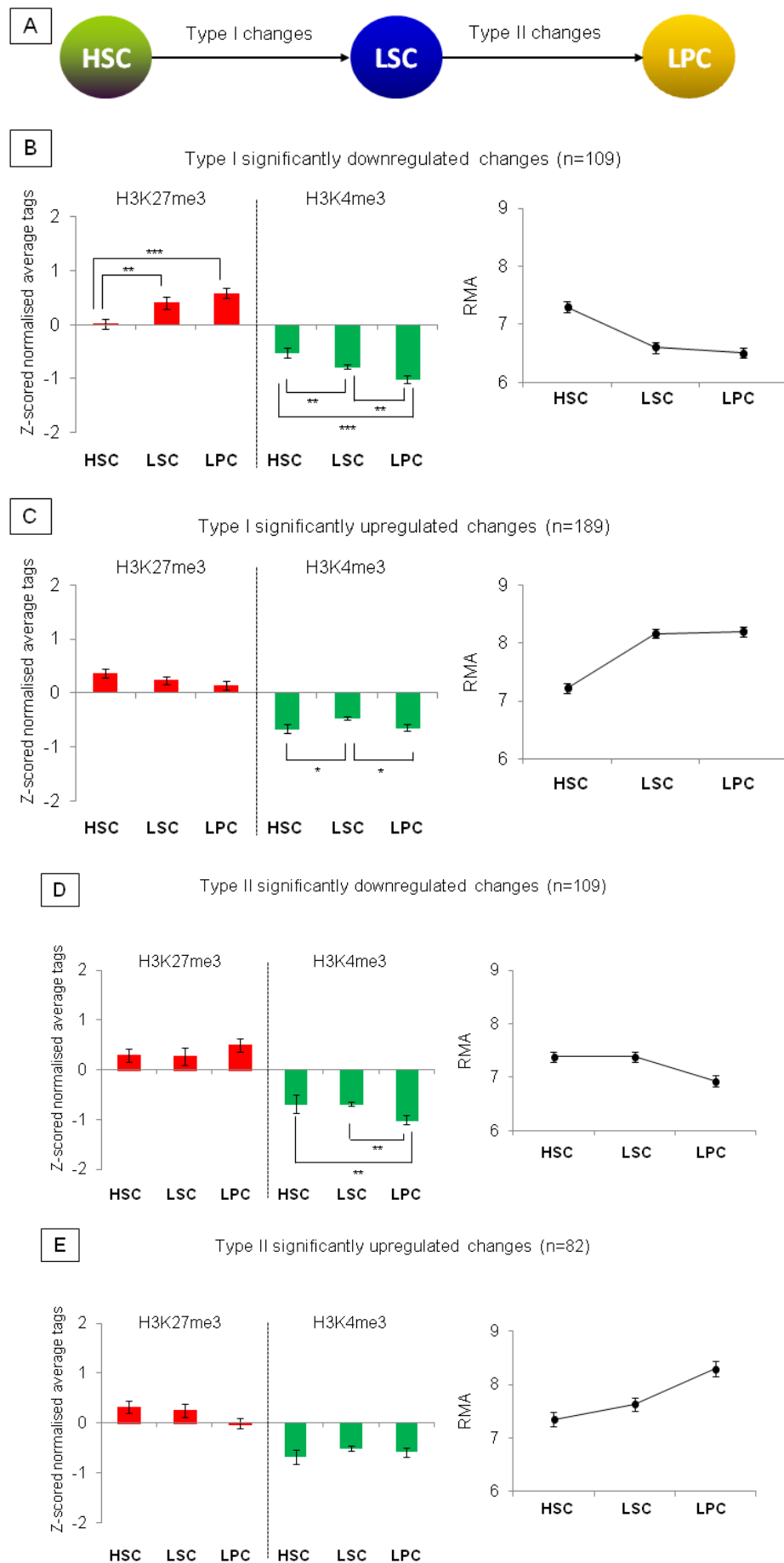
Figure 5.12: Regulation of H3K4me3 and H3K27me3 levels at the promoters of bivalent genes differentially expressed in CML cells.

(A) The differentially expressed bivalent genes in LSCs/LPCs were examined during HSC-to-LSC transition (type I) and LSC-to-LPC transition (type II). (B-E) The histograms demonstrate the z-scored levels of H3K4me3 (green) and H3K27me3 (red) at the promoters of (B) type I significantly downregulated changes, (C) type I significantly upregulated changes, (D) type II significantly downregulated changes, and (E) type II significantly upregulated changes. The average expression levels of the genes investigated in each category were also expressed as RMA values. The error bars indicate standard error of mean. * p<0.05; ** p<0.01; *** p<0.0001. P-values were calculated by unpaired two-tailed t-test.

5.15 Gain and loss of bivalency are associated with significant differential expression in CML cells

In the previous Section it was established that the gains and losses of H3K27me3 at the bivalent/H3K27me3^{hi} promoters were associated with significant differential expression in LSCs/LPCs, whereas this mechanism was not observed in during normal differentiation. Consequently, these findings suggested that the resolution of bivalent promoters to H3K4me3 by losing H3K27me3 marks or conversion of H3K4me3 promoters to bivalent by gaining H3K27me3 marks could have a significant impact on transcriptional regulation in CML. This was further investigated by demonstrating the significant gene expression changes (ΔRMA values) that were associated with the above epigenetic changes during four transitions, including (i) HSC-to-HPC, (ii) HSC-to-LSC, (iii) LSC-to-LPC, and (iv) HSC-to-LPC (Figure 5.13).
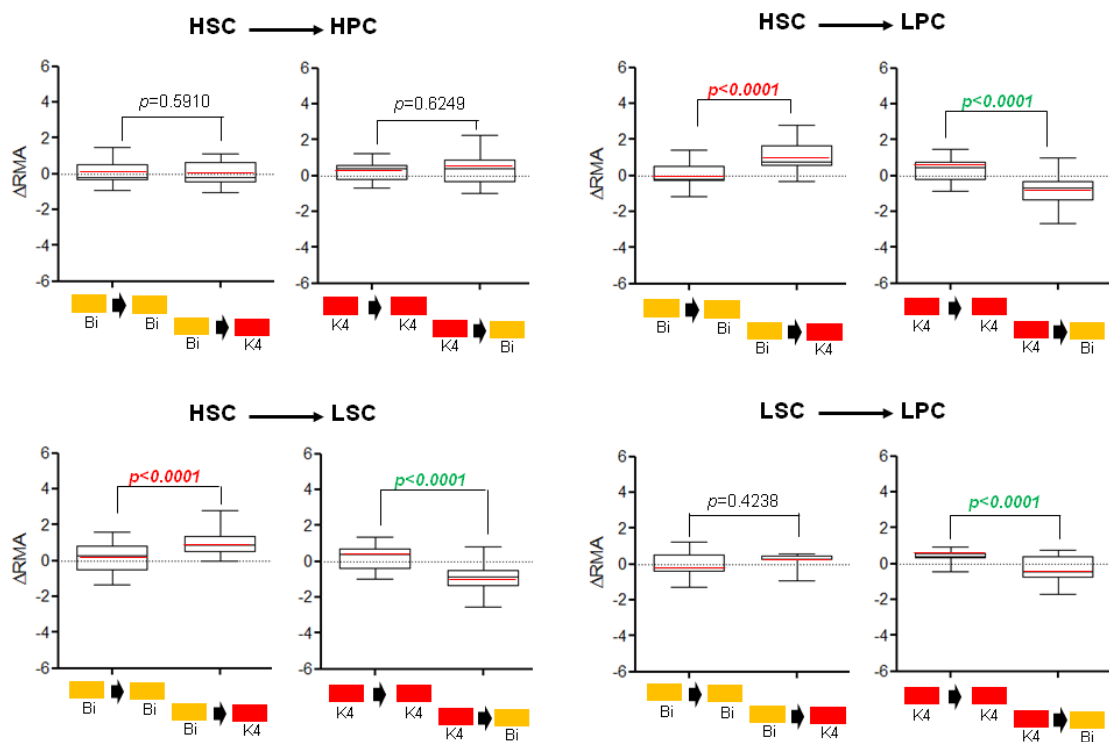
Figure 5.13: The effects of gain and loss of bivalency on transcriptional regulation during normal differentiation and leukaemic transformations.

Box plots represent the spread of the gene expression changes (ΔRMA) as a result of the gain and loss of bivalency in the four transitions. The whiskers represent 5-95 percentile with the medians shown as horizontal black lines and the average values shown as horizontal red lines. The y-axes show the ΔRMA values (gene expression differences) and the x-axes show the type of epigenetic change. The p-values were calculated using unpaired, two-tailed t-test comparing the average ΔRMA values between the classes of epigenetic changes. The $p<0.05$ was considered to be significant and highlighted red for a significant upregulation or green for a significant downregulation.

The resolution of bivalency to H3K4me3 and the gain of bivalency from the H3K4me3 compartment were not associated with a significant differential expression in the HSC-to-HPC transition (Figure 5.13). However, the bivalency resolution to H3K4me3 was associated with a significant upregulation of gene expression during HSC-to-LSC and HSC-to-LPC transitions. Furthermore, the gain of bivalency from H3K4me3 compartment was associated with a significant downregulation during all leukaemic transformations, including HSC-to-LSC, LSC-to-LPC, and HSC-to-LPC transitions.

Therefore, the above observations were in agreement with the unique association between the gains and losses of H3K27me3 in transcriptional regulation in LSCs and LPCs that were not observed with their normal counterparts. Thus, a profound role for the PcG proteins can be suggested in CML cells.

5.16 The chromatin modifiers are misregulated in leukaemic cells.

Various lines of evidence presented in previous Sections demonstrated that the bivalent/H3K27me3^{hi} promoters were differentially regulated in LSCs/LPCs. Consequently, it was important to investigate the abnormalities in the function of chromatin modifiers that regulate bivalent/H3K27me3^{hi} promoters in CML cells and their normal counterparts. H3K4me3 and H3K27me3 marks are added to the histone tails by TrxG and PcG PRC2 complexes, respectively. Furthermore, the removal of these histone modifications is mediated by KDM proteins (Section 1.5.4.5).

The abnormalities in the function of these protein complexes could be mediated by the abnormal expression of various members of the above protein complexes. Therefore, their gene expression levels (RMA values) were extracted from the Affymetrix datasets. Furthermore, differential expression analyses were performed between the four cell types for all possible comparisons. The gene expression

differences between all four cell types and the gene expression levels for individual genes in these complexes were plotted (Figures 5.14 – 5.16).

5.16.1   *PcG PRC2 members are misregulated in CML.*

The bivalent/H3K27me3^{hi} promoters in LSCs showed a loss of repressive power, as the majority of gene expression changes were associated with upregulation. Nevertheless, the bivalent/H3K27me3^{hi} promoters in LPCs demonstrated the regaining of this repressive power. Furthermore, differential expression of bivalent genes in CML cells was associated with significant gains and losses of H3K27me3 levels. Similarly the gain and loss of bivalency in CML cells were also demonstrated significant differential expression. The above observations suggested that the PcG PRC2 complex could be misregulated in CML cells.

Interestingly, the two catalytic subunits of PcG PRC2 complex, EZH1 and EZH2, were found to be significantly misregulated in both LSCs and LPCs (Figure 5.14). EZH1 was significantly downregulated in LSCs and LPCs relative to both HSCs and HPCs, whereas EZH2 was significantly upregulated in LSCs and LPCs relative to HSCs and HPCs, respectively. EZH1 levels were significantly higher than EZH2 in HSCs, as indicated by their RMA levels. This was in agreement with previous reports indicating EZH1 as the dominant catalytic subunit of PRC2 HSCs (Mochizuki-Kashio et al., 2011). In the course of normal differentiation, significant downregulation of EZH1 and upregulation of EZH2, resulted in similar mRNA levels for both proteins in HPCs. However, both LSCs and LPCs demonstrated significantly higher levels of EZH2 than EZH1. Although the EZH1 levels remained significantly unchanged between LSCs and LPCs, the EZH2 levels were significantly elevated in LPCs than LSCs.

Therefore, the observed loss of repressive capacity of the bivalent/H3K27me3^{hi} promoters in LSCs is likely to be associated with the significant downregulation of EZH1, whereas the regaining of repressive capacity in these promoters could be linked to the higher levels of EZH2.

In addition to the catalytic subunits, another PRC2 subunit, SUZ12, which is important in activating the catalytic subunits (Pasini et al., 2004), was also abnormally upregulated in LSCs and LPCs (Figure 5.14). RBBP7 and PHF19,

which are associated with chromatin recruitment of PRC2 (Nekrasov et al., 2007, Song et al., 2008), were also upregulated in LSCs. Furthermore, RBBP4 and RBBP7 were upregulated abnormally in LPCs, whereas MTF2, which is another recruitment factor, was significantly downregulated in LPCs. JARID2, PHF19, and PHF1 were also significantly differentially expressed in LPCs relative to HSCs, but their levels were not significantly different from HPCs and therefore, could be mainly associated with the proliferative capacities of progenitor cells. Overall, the abnormal regulation of non-catalytic subunits of PRC2 could have an additional effect on the function of PRC2 complex and its recruitment to the target loci.

Figure 5.14: Gene expression levels of the PcG PRC2 complex members in the four cell types.
The gene expression levels were extracted from the Affymetrix datasets for each cell type and the significant expression changes were calculated using the HSC expression levels as the calibrator. Left: the gene expression differences are shown schematically between the four cell types. horizontal lines indicate the significance levels (p<0.05) and the error bars represent the 95% confidence intervals. Right: the raw RMA values are represented for the genes that showed significant differential expression in at least one pair-wise comparison. The error bars indicate the 95% confidence intervals.

5.16.2   TrxG complex members are misregulated in CML.

In addition to the changes of H3K27me3 levels, H3K4me3 levels were also significantly altered at the promoters of differentially expressed bivalent genes in

CML cells. Therefore, the expression levels of TrxG complex members were investigated in the four cell types (Figure 5.15). MLL and ASH1L were the catalytic subunits that were significantly abnormally downregulated in LSCs. Furthermore, MLL2 and SETD1A were the catalytic subunits that were significantly abnormally downregulated and upregulated in LPCs, respectively. However, LPCs also showed significant downregulation of MLL, MLL3, and ASH1L relative to HSCs similar to HPCs (Figure 5.15), which could be associated with the proliferative capacities of progenitor cells.

The effect of MLL and ASH1L reduction in LSCs could imply a reduction in H3K4me3 levels which was observed in the case of significantly downregulated bivalent genes in LSCs (Figure 5.12B). Furthermore, significantly lower levels of H3K4me3 at the promoters of the control bivalent gene set in LPCs, which was also observed in the case of HPCs (Figure 5.11A), could be due to the significant downregulation of MLL, MLL2, MLL3, and ASH1L catalytic subunits in LPCs.

In addition to the catalytic subunits of TrxG complex, several accessory proteins, which are involved in recruiting and activating the complex, were also misregulated in leukaemic cells (Figure 5.15). PAXIP1 and WDR5 were significantly abnormally upregulated and NCOA6 was significantly abnormally downregulated in LSCs. Furthermore, ASH2L, CXXC1, DPY30, HCFC1, PAXIP1, RBBP5, and WDR5 were significantly abnormally upregulated in LPCs. The upregulation of TrxG accessory proteins in CML cells could potentially increase the assembly, catalytic activity, and recruitment of TrxG complex to certain target loci.
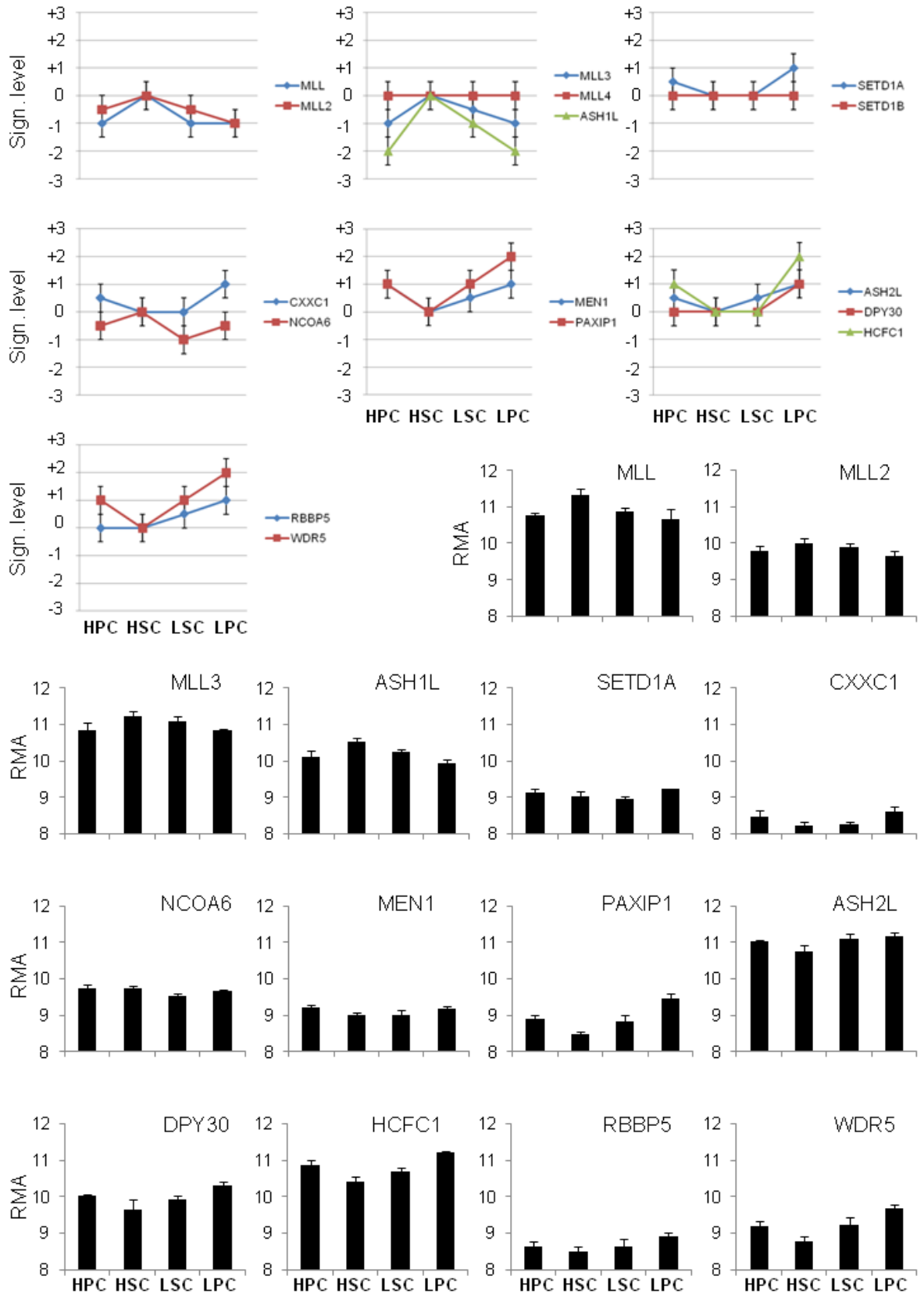
Figure 5.15: Gene expression levels of the TrxG complex members in the four cell types.
The gene expression levels were extracted from the Affymetrix datasets for each cell type and the significant expression changes were calculated using the HSC expression levels as the calibrator. Top: the gene expression differences are shown schematically between the four cell types. horizontal lines indicate the significance levels (p<0.05) and the error bars represent the 95% confidence intervals. Bottom: the raw RMA values are represented for the genes that showed significant differential expression in at least one pair-wise comparison. The error bars indicate the 95% confidence intervals.

5.16.3   H3K4 demethylases are misregulated in CML.

As mentioned earlier, changes in the levels of H3K4me3 and H3K27me3 is a net outcome of the chromatin modifiers that are involved in adding and removing these histone marks. TrxG and PcG PRC2 complexes are involved with incorporating histone methylation, whereas KDM proteins are required for their removal (reviewed in Section 1.5.4.5). Consequently, the gene expression levels of the KDMs were established in the four cell types. Interestingly, the expression levels of H3K27 demethylases, KDM6A and KDM6B, were not significantly different between the four cell types (Figure 5.16). This could indicate that the losses of H3K27me3 levels at the promoters of bivalent/H3K27me3^{hi} promoters in CML cells were not directly linked to abnormalities in the expression levels of H3K27 KDMs.

On the other hand, H3K4 demethylases were shown to be misregulated in CML cells (Figure 5.16). KDM5B was the only H3K4 demethylase that was significantly abnormally downregulated in LSCs. Furthermore, KDM5A and KDM5B were also abnormally downregulated in LPCs. The reduced expression of H3K4 demethylases could result in increased levels of H3K4me3. However, LSCs and LPCs did not show an impairment in removing H3K4me3 marks at the promoters of bivalent/H3K27me3^{hi} genes (Figure 5.12). Moreover, KDM1A was the only H3K4 demethylase that was significantly upregulated in LPCs. The increased levels of KDM1A in LPCs could potentially neutralise the effect of KDM5A and KDM5B downregulation. Overall, the effect of abnormalities in the levels of H3K4 KDMs on the epigenetic regulation of bivalent/H3K27me3^{hi} promoters could be part of an integrated network of abnormalities involving both PcG PRC2 and TrxG complex proteins.
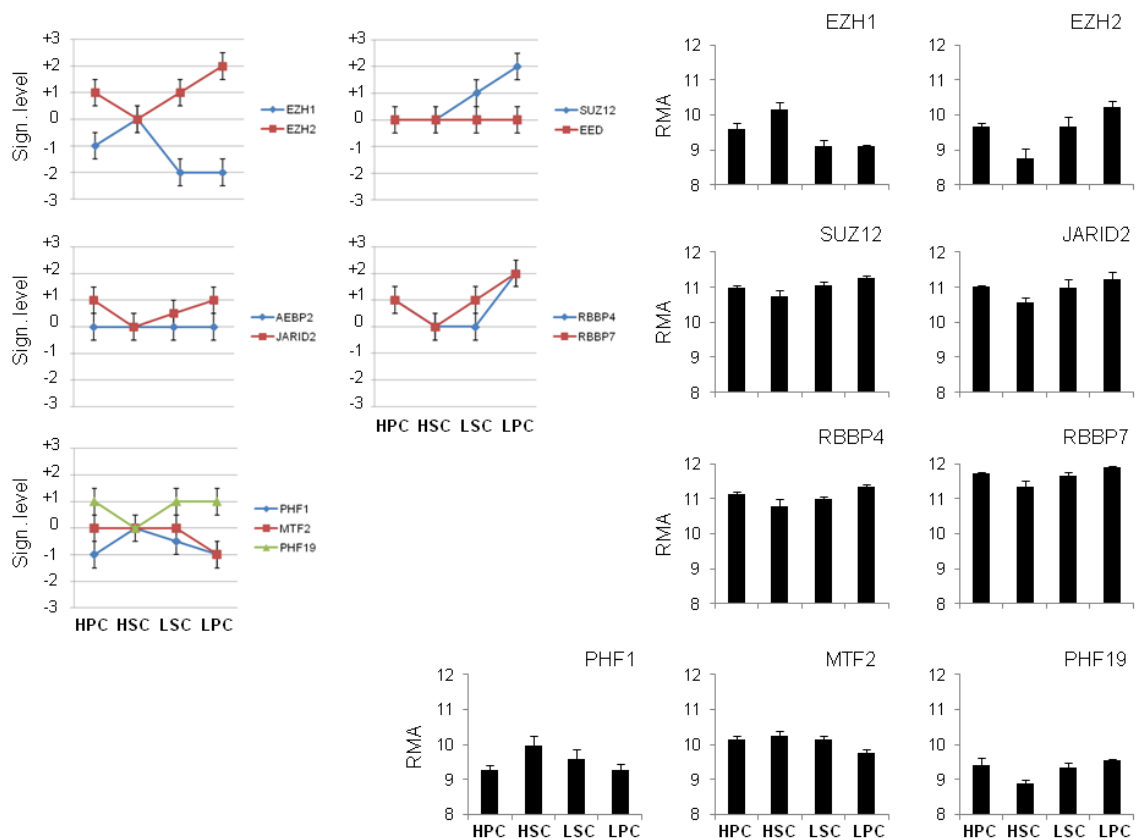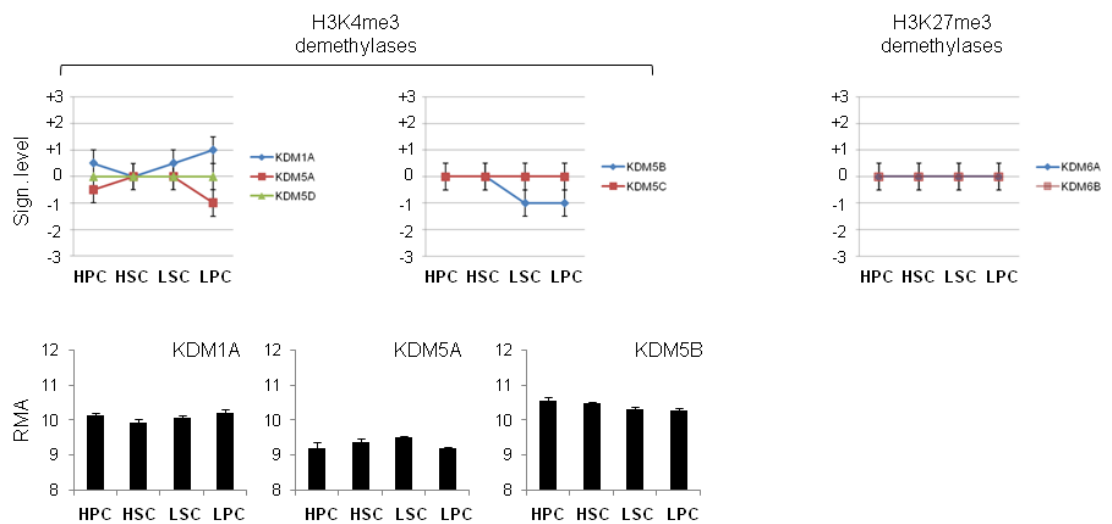
Figure 5.16: Gene expression levels of the KDM proteins in the four cell types.
The gene expression levels were extracted from the Affymetrix datasets for each cell type and the significant expression changes were calculated using the HSC expression levels as the calibrator. Top: the gene expression differences are shown schematically between the four cell types. horizontal lines indicate the significance levels (p<0.05) and the error bars represent the 95% confidence intervals. Bottom: the raw RMA values are represented for the genes that showed significant differential expression in at least one pair-wise comparison. The error bars indicate the 95% confidence intervals.

Overall, several subunits of the PcG PRC2 and TrxG complexes as well as H3K4 demethylases were misregulated in CML cells. These abnormalities could be linked with the abnormal regulation of bivalent/H3K27me3^{hi} promoters in CML cells. However, the significant changes in mRNA levels are not sufficient on their own to provide mechanistic links between the abnormal regulation of these promoters and the abnormalities in the function of chromatin modifiers. Therefore, further analysis is required to establish the changes in the protein levels of misregulated chromatin modifiers in addition to elucidating their target specificity by ChIP assays in CML cells and their normal counterparts.

5.17 Discussion

The work described in this Chapter characterised the histone modification signatures of a cancer stem cell for the first time using the developed low-cell ChIP method in combination with deep sequencing. The analysis of global H3K4me3 and H3K27me3 distributions in HSCs, HPCs, LSCs, and LPCs resulted in the identification of three major promoter classes based on their epigenetic signatures, including H3K4me3, bivalent/H3K27me3^{hi}, and neither modifications. The H3K4me3 promoters were significantly enriched for the HPC identity pathways, while the bivalent/H3K27me3^{hi} promoters were significantly enriched for the HSC

identity pathways. Furthermore, the epigenetic mechanisms associated with the regulation of bivalent/H3K27me3^{hi} promoters in CML cells and their normal counterparts were investigated. The major findings of this Chapter are discussed below.

5.17.1   LSCs and LPCs showed elevated number of bivalent/H3K27me3^{hi} promoters

The clustering of all annotated promoters in the human genome identified three major classes of promoters based on the H3K4me3 and H3K27me3 enrichments in all four cell types, including H3K4me3, bivalent/H3K27me3^{hi}, and neither modifications. The genes identified in the H3K4me3 promoter class were demonstrated to be expressed at significantly higher levels with a more dynamic range than the other categories. The bivalent/H3K27me3^{hi} and neither modifications categories were associated with low expressed and repressed genes, while the bivalent/H3K27me3^{hi} promoters showed the least dynamic range of expression. In the course of normal differentiation, HPCs demonstrated 20% reduction in the number of bivalent/H3K27me3^{hi} promoters relative to HSCs. However, the number of bivalent/H3K27me3^{hi} promoters was increased in LSCs and LPCs.

Bivalent promoters were initially reported in association with the developmentally regulated genes in ES cells and were found to be expressed at very low levels (Azuara et al., 2006, Bernstein et al., 2006, Guenther et al., 2007, Stock et al., 2007). Bivalent promoters were demonstrated to be kept at a poised state which can provide plasticity for transcriptional regulation in the course of development and differentiation (Mikkelsen et al., 2007). The same study also established that very few novel bivalent promoters were formed in the course of normal differentiation, which was in agreement with the observations in this chapter. The pluripotency factors whose function was no longer required in differentiated cells were demonstrated to be associated with the progressive resolution of their bivalent promoters to H3K27me3 (Cui et al., 2009, Mikkelsen et al., 2007, Mohn et al., 2008). Therefore, observing more bivalent bivalent/H3K27me3^{hi} promoters in CML cells could imply that LSCs and particularly LPCs maintain a less mature phenotype relative to HPCs by keeping a large set of developmentally regulated genes poised for longer.

Previous reports also supported the identification of other promoter classes, for instance H3K4me3 histone modifications were found as sharp peaks over the active promoters (Barski et al., 2007, Guenther et al., 2007, Kim et al., 2005). Furthermore, the promoters that were not associated with neither H3K4me3 or H3K27me3 histone modifications were reported to be located in the CpG poor regions which were largely found in association with DNA methylation and were therefore transcriptionally silenced (Fouse et al., 2008, Meissner et al., 2008).

5.17.2 Bivalent/H3K27me3^{hi} genes are enriched for the HSC identity pathways

A cluster of 40 biological pathways were identified which were significantly associated with bivalent/H3K27me3^{hi} promoters in the four cell types. Intriguingly, 9 HSC identity and 7 CML identity pathways were found in this category, which were all significantly downregulated at the expression level upon commitment to HPCs and/or LPCs. Furthermore, the HPC identity pathways were significantly associated with the H3K4me3 promoters in the four cell types. These pathways were significantly upregulated upon leaving the HSC compartment.

The analysis of genome-wide significant differential expression associated with each class of promoters, demonstrated that 70% of H3K4me3 promoters were associated with the upregulation of expression during normal differentiation. Although the upregulation events was favoured to downregulation during HSC-to-LSC transition, the proportion of upregulated genes were reduced to 60%, suggesting a decrease in the activating capacity of these promoters in LSCs. However, this activating capacity was restored in LPCs. Consequently, the HPC identity pathways, which are linked to proliferation, are required to be associated with H3K4me3 promoters in order to be upregulated upon leaving the HSC compartment.

Conversely, the bivalent/H3K27me3^{hi} promoters were mainly associated with downregulation of gene expression during normal differentiation. This was in agreement with the observed link between these promoters and the HSC identity pathways that potentially regulate stem cell maintenance. However, the repressive capacity of these promoters was shifted towards activation and upregulation during HSC-to-LSC transitions. LPCs demonstrated a regain of repressive capacity for these promoters. Therefore, the above observation suggested that

there are fundamental differences in the epigenetic regulation of bivalent promoters in LSCs.

Intriguingly, the gene expression changes of the bivalent/H3K27me3^{hi}-enriched pathways were also skewed towards activation and upregulation in the HSC-to-LSC transition. However, the number of downregulated genes were higher than the upregulated events in LPCs. Therefore, these results indirectly suggest a different epigenetic programming in LSCs and LPCs in the regulation of HSC identity pathways from their normal counterparts.

5.17.3   The Polycomb target genes are regulated differently in CML cells.

The analysis of the bivalent/H3K27me3^{hi} genes that were significantly differentially expressed in both HPCs and CML cells upon leaving the HSC compartment, revealed that the changes of H3K27me3 levels at these promoters are associated with significant up- or downregulation events more profoundly in CML cells. Significantly downregulated bivalent/H3K27me3^{hi} genes demonstrated a significant increase in the H3K27me3 levels at their promoters in LSCs. Additionally, the H3K4me3 levels were significantly reduced at the promoters of these genes in LSCs. The results demonstrated a crosstalk between the two histone modifications at the promoters of the repressed bivalent/H3K27me3^{hi} genes during HSC-to-LSC transition. Intriguingly, the levels of H3K4me3 and H3K27me3 marks at the promoters of differentially expressed bivalent/H3K27me3^{hi} genes in the LSC-to-LPC transition were not significantly altered, suggesting other mechanisms of gene regulation are potentially involved in this transition. Thus, it is plausible to suggest that the different epigenetic programming at the bivalent/H3K27me3^{hi} promoters in CML is initiated during the HSC-to-LSC transition.

Furthermore, it was demonstrated that resolution of bivalency to H3K4me3 and the conversion of H3K4me3 promoters to bivalent during leukaemic transformations were linked to significant upregulation and downregulation of gene expression, respectively. This phenomenon was not observed in the case of normal differentiation. A recent study that investigated the mechanism of gene repression during normal development showed that the repressed genes in differentiated cells were associated with the expansion of H3K27me3 marks over to the gene bodies

(Hawkins et al., 2010). This study also demonstrated the repressed promoters were positively correlated with the gain of H3K9me3 marks at their promoters, suggesting two different mechanisms involved in suppression of developmentally regulated promoters during normal differentiation.

However, the gains and losses of H3K27me3 at the promoters are linked with significant changes of gene expression in CML cells. Thus, the Polycomb PRC2 complex in CML cells have a more profound role in transcriptional regulation at the gene promoters than their normal counterparts.

5.17.4   Several members of PcG PRC2, TrxG and KDM complexes are misregulated in leukaemic cells.

The results presented in this Chapter indicated a unique epigenetic programming of bivalent/H3K27me3^{hi} genes in CML cells which involved the usage of H3K4me3 and H3K27me3 levels at their promoters to regulate their transcriptional activity. Therefore, the differences in the epigenetic programming of these two histone modifications between HPCs and CML cells could indicate a different mechanism of function for the protein complexes responsible for the addition and removal of these histone modifications, which include PcG PRC2 and TrxG complexes as well as KDM proteins.

5.17.4.1   Abnormalities in the levels of PcG PRC2 members

Analysis of the gene expression levels of the PcG PRC2 members in the four cell types, demonstrated profound differences in the levels of two catalytic subunits, EZH1 and EZH2, between normal and leukaemic cell types. Both LSCs and LPCs displayed significant downregulation of EZH1, which was the more expressed catalytic subunit in HSCs, and significant upregulation of EZH2. The levels of EZH1 and EZH2 were significantly imbalanced in both LSCs and LPCs, whereas HPCs showed similar expression levels for both proteins. The loss of EZH1 in LSCs could explain the loss of repressive capacity in bivalent/H3K27me3^{hi} promoters. However, significant upregulation of EZH2 in LPCs relative to LSCs could provide evidence for the regaining of repressive capacity in the bivalent/H3K27me3^{hi} promoters. Furthermore, several other PRC2 subunits were also upregulated in LSCs and LPCs that could be important in targeting and assembly of the catalytic subunits at the Polycomb target. Overall, the

abnormalities at the gene expression levels of PRC2 members may explain the epigenetic reprogramming in CML cells that exploit the H3K27me3 levels at the gene promoters more efficiently in order to transcriptionally silence or activate genes.

Previous reports indicated that PRC2-deficient ES cells were impaired in their differentiation and lineage commitment and therefore, the role of PRC2 in fate determination processes was to inactivate the pluripotency-specific factors (Chamberlain et al., 2008, Pasini et al., 2007, Shen et al., 2008). The observed role of EZH2 in ES cells could explain its upregulation in the course of normal differentiation, but whether leukaemic cells also exploit the PRC2 complex in the same fashion is not known. However, in a recent study, the deletion of EZH2 was linked to the differentiation of AML cells, indicating an oncogenic role for EZH2 in AML (Tanaka et al., 2012). The EZH2 double knock-out AML mice exhibited increased apoptosis but maintained their proliferative activity. However, the AML LSCs were found to be significantly affected and showed impaired function in serial transplantation assays. This study showed that the EZH2 activity resulted in the maintenance of LSCs through blocking differentiation mechanisms. Therefore, it was suggested that LSCs were more dependent on EZH2 function than the normal HSCs (Tanaka et al., 2012). Recent studies also provided evidence for the role of bivalent promoters during cancer in which these promoters showed a higher tendency for becoming abnormally hypermethylated (Rodriguez et al., 2008, Widschwendter et al., 2007). It was suggested that bivalent promoters in adult stem cells were more likely to gain DNA methylation which could result in the formation of self-renewing tumour cells that exhibit reduced differentiation capacities (Widschwendter et al., 2007).

Based on the observations in this chapter and the role of EZH2 during normal differentiation and AML formation, a model of action for EZH2 can be proposed in CML. LSCs could be more dependent on the EZH2 function that their normal counterparts as a result of significant downregulation of EZH1. However, the exit from stem cell compartment in CML is potentially mediated by the increased activity of EZH2 that could in turn repress bivalent/H3K27me3[hi] promoters many of which associated with the HSC identity. Although the suppression of the HSC identity pathways in LSCs is potentially a key event in initiating the development of LPCs, as discussed in Chapter 3, the elevated levels of bivalent/H3K27me3[hi]

genes in LPCs could be indicative of their less differentiated phenotype than their normal counterparts. However, the link between EZH2 upregulation and the suppression of differentiation-related genes could not be directly verified in this Chapter. Thus, it is plausible to hypothesise a model in which the abnormal epigenetic programme in LSCs is initiated by the loss of EZH1, but the exit from stem cell compartment in LPCs is mediated by the increase in the EZH2 levels.

5.17.4.2   Abnormalities in the levels of TrxG members

Analysis of the TrxG complex also identified several misregulated members in CML cells. MLL and ASH1L catalytic subunits were found to be significantly downregulated in LSCs. MLL protein has been reported to have an important role in the resolution of bivalency to H3K4me3 promoters by recruiting H3K27 demethylases (Agger et al., 2007, Lee et al., 2007). One study demonstrated an increase in H3K27me3 levels at the bivalent promoters in MLL knock-out neural stem cells (Lim et al., 2009). Therefore, the downregulation of MLL in LSCs could potentially explain the significant increase in H3K27me3 levels and decrease in H3K4me3 levels at the promoters of repressed bivalent/H3K27me3[hi] genes.

Furthermore, ASH1L was also reported to indirectly counteract PRC2 function as a result of H3K36 methylation (Yuan et al., 2011). Therefore, the significant downregulation of ASH1L in LSCs could result in an increase in the levels of H3K27me3 in certain loci.

In addition to the downregulation of catalytic subunits of TrxG complex, several accessory subunits were significantly abnormally upregulated in leukaemic cells. It was reported that the stability, the catalytic activity and the recruitment of the H3K4 methyltransferases were largely dependent on their core subunits (Dou et al., 2006). Therefore, the abnormalities in the levels of these proteins in CML cells could result in abnormalities in target specification and the activity of catalytic subunits in a complex manner. Consequently, in order to better elucidate the impact of above abnormalities, the binding sites of the misregulated subunits should be investigated *in vivo* in CML cells.

5.17.4.3   Abnormalities in the levels of H3K4 demethylases

The regulation of H3K4me3 in CML cells could be more complicated as several H3K4 demethylases were found to be misregulated in LSCs and LPCs. KDM5A and KDM5B which were significantly downregulated in leukaemic cells were previously reported to be associated with bivalent promoters. KDM5A, in particular, was shown to be recruited by PRC2 complex to the target genes and causing bivalency resolution to H3K27me3 (Pasini et al., 2008). Furthermore, KDM5B was also identified at the TSSs of bivalent genes and its depletion resulted in a global increase of H3K4me3 (Schmitz et al., 2011). KDM5B was found to share many PRC2 targets, although a direct interaction between PRC2 and KDM5B was not detected. Thus, the reduction in the levels of KDM5A and KDM5B in leukaemic cells could result in an increase in the levels of H3K4me3 at bivalent promoters. Although a global increase in the levels of H3K4me3 was not observed at the bivalent/H3K27me3^{hi} promoters in LSCs and LPCs, the upregulated bivalent/H3K27me3^{hi} genes in LSCs demonstrated increased levels of H3K4me3. Therefore, these promoters could be the specific target of these KDMs. Furthermore, the loss of repressive power at the bivalent/H3K27me3^{hi} promoters in LSCs and their shift towards gene activation could also be induced by the reduced expression of KDM5B in LSCs.

Nevertheless, KDM1A, which is known to be a key player in stem cell maintenance by controlling the H3K4me3 levels at bivalent promoters (Adamo et al., 2011), was significantly upregulated in LPCs. Therefore, it is plausible to suggest that LPCs compensate for the depletion of KDM5A and KDM5B by upregulating the expression of KDM1A. Consequently, this could also contribute to the regaining of repressive capacity at the bivalent/H3K27me3^{hi} promoters in LPCs.

5.18 Conclusion

The work presented in this Chapter demonstrated histone modification signatures of a cancer stem cell for the first time. It also provided crucial evidence that the HSC identity pathways, which potentially regulate stem cell maintenance, were significantly associated with the bivalent/H3K27me3^{hi} genes. Furthermore, the bivalent/H3K27me3^{hi} promoters in LSCs showed reduced repressive capacities which were restored in LPCs. The results indicated abnormal epigenetic

programming in CML cells, which was initiated during the HSC-to-LSC transition. The differential expression of bivalent/H3K27me3^{hi} genes in CML cells was associated with significant changes of H3K27me3 and H3K4me3 levels at their promoters. This mechanism was profoundly different from their normal counterparts. The significant reduction in EZH1 levels in LSCs was suggested to be the underlying mechanism for the loss of repressive capacity in bivalent/H3K27me3^{hi} genes, whereas the significant increase in EZH2 levels could explain the regaining of this repressive capacity. Furthermore, abnormalities in the levels of TrxG complex members and H3K4 demethylases could have further impact in the regulation of bivalent/H3K27me3^{hi} promoters and the significant changes in the levels of H3K4me3 marks. Overall, there is a global epigenetic reprogramming in CML cells which have a significant impact on transcriptional regulation in these cells which can be linked to the differences observed in terms of global regulation of gene expression changes in Chapter 3.

6. Neurotransmitter signalling pathways maintain stem cell features of CML stem cells

6.1 Abstract

In Chapter 3 of this thesis, genome-wide gene expression analysis implicated several novel neurotransmitter pathways in the maintenance of HSCs. These pathways exhibited similar expression levels in LSCs when compared to HSCs – suggesting that they also played a role in the stem cell phenotype of LSCs. This Chapter investigates the roles that neurotransmitters may have in maintaining the LSC phenotype. Acetylcholine, norepinephrine, serotonin, histamine, and glutamate were added to cultures of CML stem cells and their effects on quiescence and the ability to maintain a primitive stem cell phenotype in replating assays were measured. Neurotransmitter signalling pathways also show a large overlap with heterotrimeric G protein core signalling pathways – and the ability of neurotransmitters to modulate these pathways was also tested. The *in vitro* functional assays performed in this Chapter demonstrated that LSCs respond to these neuroactive molecules through cell cycle arrest as well as increasing their self-renewal capacity as observed in replating assays. These observations were found to be independent of apoptosis or toxicity-induced cell cycle arrest. Furthermore, the activity of neurotransmitter signalling pathways was demonstrated to be associated with the activation of genes of heterotrimeric G protein signalling pathways and the upregulation of cell cycle inhibitors such as ID1 and p21/CDKN1A which are important in the maintenance of the quiescent stem cell phenotype.

6.2 Introduction

Neuropeptides such as acetylcholine, catecholamines, serotonin, glutamate, histamine, endothelin, and angiotensin II stimulate various cellular functions through activating their respective G protein-coupled receptors GPCRs. In addition to neurotransmission, the GPCRs are associated with numerous other physiological functions such as immune response, hormone release, cardiac and smooth muscle contraction, and blood pressure. They are also linked to many human diseases and are reported to be direct or indirect targets of 50-60% of all current drugs (Pierce et al., 2002). Misregulation and malfunction of GPCRs have

been found during cancer progression and metastasis. The tumour cells were shown to abnormally exploit GPCRs to evade immune surveillance, proliferate autonomously, invade the surrounding tissues and metastasise to other organs, and enhance their uptake of oxygen and nutrients. The abnormal overexpression of GPCRs and/or activating mutations of G proteins are the major events causing the overt activation of GPCR signalling in tumours which could be further increased by the release of ligands produced by tumours themselves or their stromal microenvironment. Therefore, abnormal GPCR function in cancer provides opportunities for developing novel chemotherapeutic targets (Dorsam and Gutkind, 2007).

The GPCRs are a transmembrane protein family comprised of approximately 1,000 members which all share a conserved 7-transmembrane domain structure. The GPCRs undergo a conformational change in their transmembrane and intracellular domains upon binding to their ligands which results in interaction with heterotrimeric G proteins. This change of conformation converts the GPCRs into guanine nucleotide exchange factors (GEFs) for the α subunits of heterotrimeric G proteins, promoting guanosine triphosphate (GTP) binding by facilitating the release of guanosine diphophate (GDP) and therefore, activating G proteins. The activated α subunit in conjunction with β and γ subunits will bind various downstream effectors to mediate a variety of cellular responses. The α subunits of G proteins are found in multiple subtypes such as $G\alpha_{i/o}$, $G\alpha_s$, $G\alpha_q$, and $G\alpha_{12/13}$. The downstream effectors also vary dependent on the Gα subtype, including adenylyl cyclase, phospholipase Cβ (PLCβ), and Rho-GEF. The effectors directly or via second messengers such as cyclic adenosine monophosphate (AMP), diacylglycerol (DAG), and inositol-1,4,5-triphosphate ($Ins(1,4,5)P_3$) regulate downstream effectors, including protein kinase A (PKA) and protein kinase C (PKC). After dissociation of heterotrimeric G proteins, the βγ subunits can modulate other downstream effectors such as PLCβ and ion channels (Ritter and Hall, 2009) (Figure 6.1). Table 6.1 demonstrates the association between different GPCRs and α subunit families and their reported mechanisms of action.
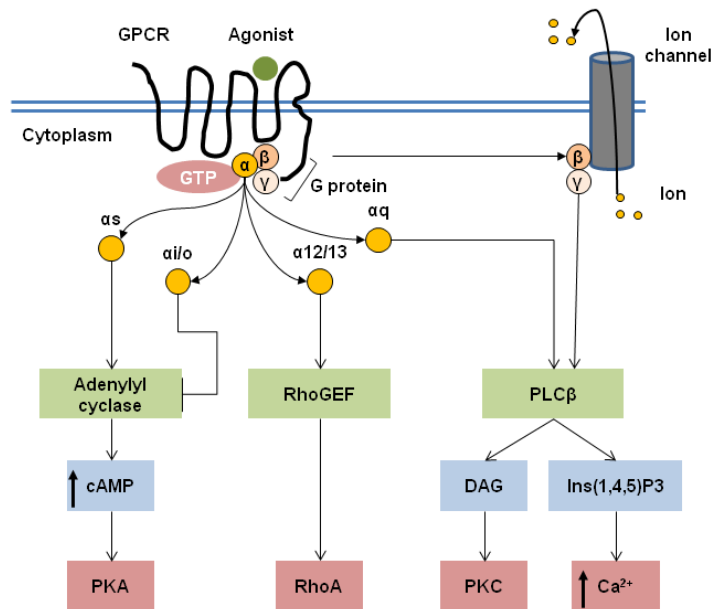
Figure 6.1: The mechanism of heterotrimeric G protein activation and downstream effectors by the ligand-bound GPCRs.
Schematic diagram depicts the mechanism of heterotrimeric G protein activation upon binding the agonist to the 7-transmembrane GPCRs. The GPCR activation results in the substitution of GDP with GTP in the α subunit of heterotrimeric G proteins. Different subtypes of α subunit (yellow circles) are linked with the activation or inhibition of different downstream effectors (green boxes) which in turn activate another set of downstream effectors (red boxes) directly or via the second messenger system (blue boxes). The dissociated βγ subunits, upon agonist binding, are also capable of activating downstream effectors through PLCβ or activating the ion channels in the cell membrane.

α subunit	Signal transduction	Receptors	Genes
αi αo	Inhibition of adenylyl cyclase Activation of K^+ channels Inhibition of Ca^{2+} channels	α2-adrenergic Acetylcholine muscarinic 2 and 4 Serotonin 5HT1 and 5HT5 Glutamate metabotropic group II and III Histamine H3 and H4 Dopamine D2, D3, D4 GABA-B Opioid Chemokine (CXCR4)	GNAI1 GNAI2 GNAI3 GNAO1
αs	Activation of adenylyl cyclase	β-adrenergic Serotonin 5HT4, 5HT6, 5HT7 Histamine H2 Dopamine D1 Corticotropin-releasing hormone	GNAS
αq α11 α14 α15 α16	Activation of phospholipase C	α1-adrenergic Acetylcholine muscarinic 1, 3, 5 Serotonin 5HT2 Glutamate metabotropic group I Histamine H1 Angiotensin II Oxytocin Thyrotropin-releasing hormone	GNAQ GNA11 GNA14 GNA15
α12 α13	Activation of Rho GTPases		GNA12 GNA13
αolf	Activation of adenylyl cyclase	Olfactory	GNAL
αt	Activation of phosphodiesterase 6	Rhodopsin	GNAT1 GNAT2

Table 6.1: The α subunit families and their mechanism of action.
The four major α subunit families $α_{i/o}$, $α_s$, $α_q$, $α_{12/13}$ are listed along with less frequent α subunit families associated with olfactory receptors ($α_{olf}$) and rhodopsin ($α_t$). The mechanism of action of each family and their associated GPCRs are also listed. The last column shows the gene names of the α subunits in each family (Ritter and Hall, 2009).

The ligand-bound GPCRs can also become phosphorylated as a result of interaction with GPCR kinases (GRKs). Upon phosphorylation, GPCRs limit their interaction with G proteins and increase their association with the scaffold proteins known as arrestins which inhibit G protein-mediated GPCR signalling (Reiter and Lefkowitz, 2006). Activation of arrestins could enhance the endocytosis of active GPCRs by clathrin-coated pits to regulate the desensitisation and resensitisation of GPCRs (Hanyaloglu and von Zastrow, 2008) (Figure 6.2). The internalised arrestin-bound GPCRs could activate distinct signalling pathways via arrestin proteins such as MAPK and ERK pathways. Recent reports indicate that different agonists have preferences for arrestin-mediated GPCR signalling over the G protein-mediated signalling pathways, or vice versa (Urban et al., 2007).
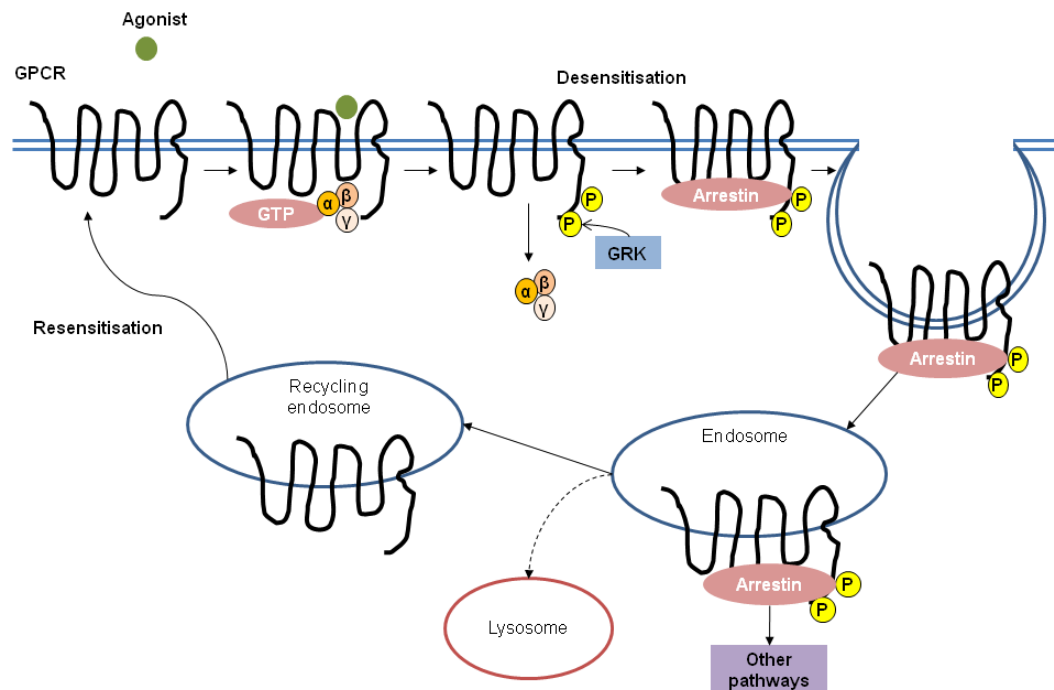
Figure 6.2: Arrestin-mediated desensitisation and resensitisation of activated GPCRs.
Activated GPCRs can be phosphorylated by GRKs that in turn favour arrestins over heterotrimeric G proteins in a process known as desensitisation. The association of arrestins results in internalisation of GPCRs via clathrin-coated pits and endocytosis of GPCRs. The internalised GPCRs can be recycled and re-presented to the cellular membrane in a process known as resensitisation or targeted to lysosomes for degradation. Furthermore, the arrestin-bound internalised GPCRs can activate several other intracellular pathways such as MAPK or ERK via arrestin proteins.

The link between abnormal activation of neurotransmitter receptors and tumour progression was established in early studies, including serotonin and acetylcholine GPCRs (Gutkind et al., 1991, Julius et al., 1989). Furthermore, the abnormal activation of endothelin receptors and angiotensin II receptors was identified in prostate cancer and found to be associated with high levels of the ligands in

plasma (Daaka, 2004). The antagonists of these receptors caused growth arrest in prostate tumour cells both *in vitro* and *in vivo* (Daaka, 2004). The activation of neuropeptide GPCRs in some cases results in the activation of PKC via $G\alpha_q$ subunit (Heasley, 2001). However, further investigations showed that neuropeptide signalling can trigger activation of ERK, MAPK, and Rho GTPases (Marinissen et al., 2001). Therefore, activation of neuropeptide receptors via autocrine and/or paracrine routes stimulates an interconnected network of intracellular pathways that lead to the survival and/or proliferation of tumour cells (Dorsam and Gutkind, 2007).

Despite the suggested role of neuropeptide signalling via heterotrimeric G proteins in cancer progression, their function in the maintenance of cancer stem cells has not been studied to date. The work presented in previous chapters, however, provided compelling *in vitro* evidence regarding involvement of several neuropeptides in the maintenance of LSCs, at both the epigenetic and gene expression levels; and these neuropeptides are also likely to be associated with HSC maintenance. Therefore, the work presented in this Chapter was to further investigate the role of neurotransmitter signalling pathways in stem-cell maintenance in the context of LSCs in the presence of stromal cells that partially mimic the bone marrow microenvironment *in vitro*.

6.3 Aims of the Chapter

1. To determine whether neuropeptides are capable of promoting stem cell features of LSCs, including quiescence and self-renewal.

2. To begin to understand the relationships between neuropeptides, stem cell features of LSCs and intracellular signalling pathways linked via heterotrimeric G-proteins.

Results

6.4 Overall strategy

Flow-cytometry isolated primary LSCs ($CD34^+ CD38^-$) were first stained with CFSE and then cocultured *in vitro* in the presence of mouse OP9-GFP stromal cells.

Mouse OP9 cells were used as the stromal support as they have previously been shown to successfully induce differentiation of mouse and human ES cells into haemopoietic cells, including $CD34^+$ cells (Kodama et al., 1994, Vodyanik et al., 2005). OP9 cells were shown to promote haemopoiesis by specifically promoting the survival and proliferation of primitive and clonogenic $CD34^+$ cells, but not the mature $CD34^-$ population (Ji et al., 2008); therefore, providing a unique *in vitro* microenvironmental support that promotes haemopoiesis in a similar hierarchical manner to the bone marrow niche.

Acetylcholine, norepinephrine, serotonin, L-glutamate and histamine neuropeptides were added to the coculture in independent assays on day 0, alongside untreated control. After 72 hrs, the GFP^- CML cells were isolated from the GFP^+ stromal cells by FACS in each condition, and the progression of cells through divisions (CFSE) was investigated. Furthermore, after 72 hrs the $CD34^+$ $CD38^-$ CML cells were isolated and cultured in semi-solid methylcellulose-based medium to analyse the colony forming potential (CFC assay) of the treated and untreated cells as indicators of haemopoietic cell growth and differentiation. Subsequently, replating assays were performed using the colonies obtained from the CFC assays to investigate the self-renewal capacities of CML cells in the presence and absence of each neurotransmitter. The $CD34^+$ $CD38^-$ CML cells which demonstrated an enhanced self-renewal and/or quiescence in the presence of each neurotransmitter were analysed further at the level of gene expression to identify the activity of downstream heterotrimeric G protein signalling pathway and various cell cycle regulators (Figure 6.3).
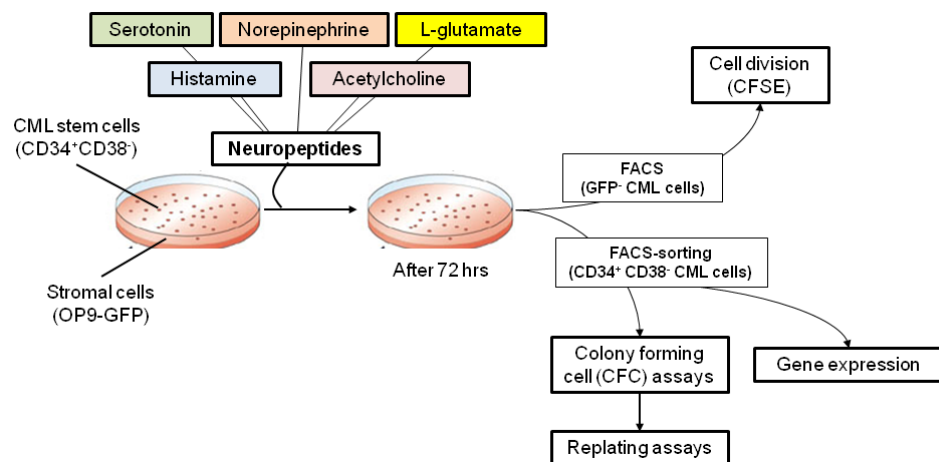
Figure 6.3: The overall strategy in functional validation of novel neurotransmitter signalling pathways in the context of CML stem cells.

The schematic diagram shows the major steps in the functional validation of several neuropeptides in the maintenance of LSC features such as quiescence and self-renewal, as well as their effect on cell growth and proliferation after 72 hrs of treatment in the presence of agonists. Gene expression signatures of several downstream effectors were investigated to validate the intracellular activity of neurotransmitter signalling pathways.

6.5 Heterotrimeric Gα protein signalling pathway shows HSC-like epigenetic signature in LSCs

The gene expression analysis in Chapter 3 revealed that the expression levels of 12 HSC identity pathways did not change significantly between HSCs and LSCs, suggesting potential roles for these pathways in the maintenance of stem cell features in LSCs. Furthermore, it was also important to identify the pathways that showed an HSC-like epigenetic signature in LSCs, which in turn may have a role in the maintenance of stem cell features of LSCs. Therefore, the genes that showed similar epigenetic signature in both HSCs and LSCs, but were associated with a different epigenetic signature in HPCs were identified (Figure 6.4A). Intriguingly, the bivalent/H3K27me3^{hi} genes demonstrated the largest number of genes that were associated with an HSC-like signature in LSCs. Subsequently, pathway analysis was performed for the genes with HSC-like epigenetic signature using PANTHER database, similar to the pathways analyses performed for different promoter classes in Chapter 5. This analysis identified 18 biological pathways that were significantly ($p<0.05$) associated with the genes that showed HSC-like epigenetic signature in LSCs (Figure 6.4B). The top two most significantly over-represented pathways in this list were the heterotrimeric G protein-mediated signalling via $G\alpha_i$, $G\alpha_s$, $G\alpha_q$ and $G\alpha_o$ subunits. Additionally, Wnt signalling, cadherin signalling, glutamate signalling via ionotropic and metabotropic receptors, acetylcholine signalling via muscarinic and nicotinic receptors, endothelin signalling, and GABA-B receptor signalling were also found to be over-represented in this list. This analysis identified several new signalling pathways mainly associated with neurotransmitters that could be potentially involved in the maintenance of stem cell features in LSCs.

Furthermore, since the heterotrimeric Gα protein signalling pathways were the most over-represented pathway in the above analysis, the overlap between the genes in these pathways and other signalling pathways were investigated (Figure 6.4C). As was expected, many neurotransmitter signalling pathways showed a large overlap with the heterotrimeric Gα protein signalling, suggesting that the Gα

protein signalling serves as downstream effector of these neurotransmitters. Seven neurotransmitter HSC identity pathways that were identified at the level of gene expression (Chapter 3) were shown to have a large overlap with the heterotrimeric Gα protein signalling pathways, including serotonin signalling via 5HT2 receptors, histamine signalling via H1 receptor, α-adrenergic receptors signalling, acetylcholine signalling via muscarinic receptors 1 and 3, angiotensin II, oxytocin, and TRH signalling pathways (Figure 6.4C).

Overall, the above results demonstrated a noticeable overlap between the pathways that were identified at the level of gene expression and epigenetics which could be associated with the maintenance of stem cell features in LSCs. Additionally, through epigenetic analysis several novel neurotransmitter signalling pathways were identified, such as glutamate signalling pathways, which may also be important in the maintenance of stem cell like features in LSCs. Therefore, in order to verify whether this comprehensive study, which combined gene expression and epigenetic analyses, could identify novel signalling pathways that control LSC maintenance, the role of serotonin, acetylcholine, catecholamines (α-adrenergic receptors), histamine, and glutamate signalling pathways were investigated *in vitro* (Section 6.5).
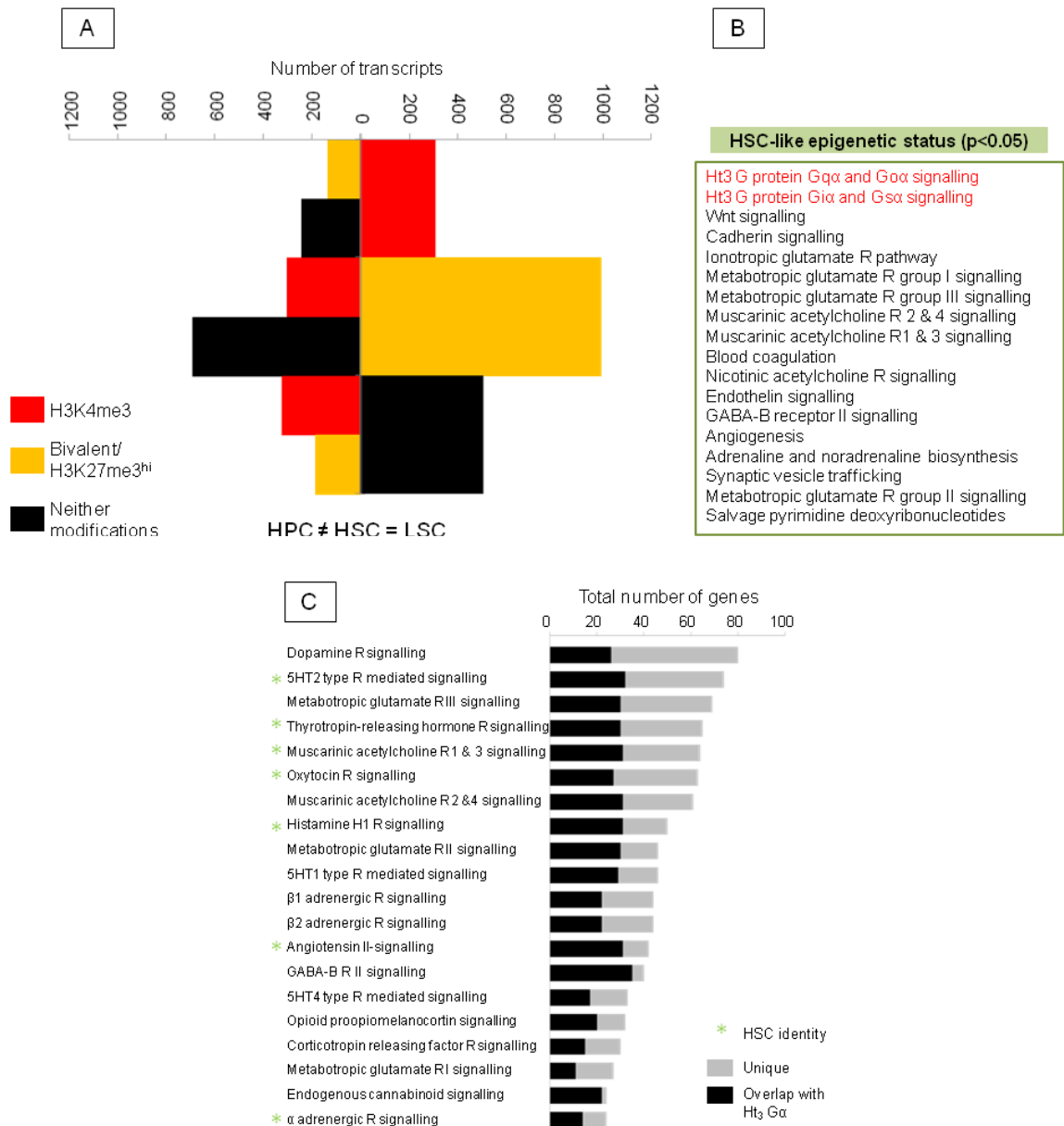
Figure 6.4: Pathways that showed HSC-like epigenetic signature in LSCs.
(A) The histogram shows the number of transcripts (y-axis) which have a similar epigenetic signature in both HSCs and LSCs (right), but have a different epigenetic signature in HPCs (left). The fates of the genes that did not show similar epigenetic signatures in HPCs are also indicated. The bivalent/H3K27me3^{hi} genes demonstrated highest number of transcripts with HSC-like epigenetic signature in LSCs, which were resolved to neither modifications or H3K4me3 in HPCs. (B) The pathways that were significantly over-represented for the genes associated with the HSC-like epigenetic signature in LSCs. The heterotrimeric (Ht3) Gα protein signalling pathways were the top two hits (highlighted in red). (C) The pathways that showed a large degree of overlap (black bars) with the Ht3 Gα protein signalling pathways. The HSC identity pathway are marked with asterisks. The total number of genes in each pathway is also represented (y-axis).

6.6 Investigating the role of neurotransmitters in maintenance of stem cell features in LSCs

The role of neurotransmitters in the maintenance of LSCs was investigated *in vitro*, as described in Chapter 2. Briefly, confluent cultures of GFP-tagged OP9 mouse stromal cells were established in collagen-coated plates and the flow cytometry-

isolated primary LSCs (CD34$^+$ CD38$^-$ population) from three different patient samples (bioreplicates) were cocultured in the presence or absence of each neuropeptide agonist for 72 hrs. The CD34$^+$ GFP$^-$ CML cells were subsequently isolated from the GFP$^+$ OP9 cells by flow cytometry (Figure 6.6) and the effect of neuropeptides on cell division progression, colony forming potential, and self-renewal capacities upon replating assays were assessed in comparison to the no-drug control (NDC). Furthermore, CD34$^+$ CD38$^-$ CML cells were isolated in each condition to investigate the gene expression levels of the heterotrimeric Gα protein signalling pathway as further confirmation of the activity of neurotransmitter signalling and their effect on cell cycle regulation.
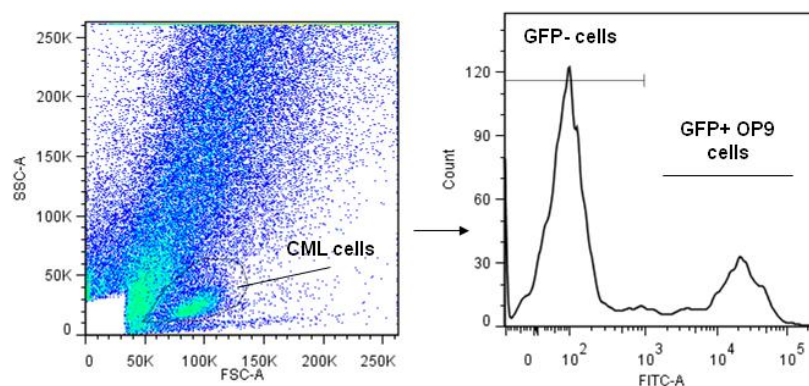
Figure 6.5: The flow-cytometry isolation of CML cells from the OP9-GFP stromal cells after 72 hrs of coculture.
The live CML cells were initially separated from the OP9 stromal cells based on their smaller size (left). Subsequently, the contaminating GFP-tagged OP9 cells were filtered out by excluding the GFP$^+$ (FITC$^+$) cells. The GFP$^-$ CML cells were then used for the downstream analysis.

The concentration of each neuropeptide agonist was chosen based on previously reported results using primary cells, although some of the agonists were not reported in the context of haemopoietic cells and therefore, the choice of concentration was made based on the best physiological response in other tissues. To induce muscarinic acetylcholine receptors, the acetylcholine agonist was applied at 100nM (Al-Zi'abi et al., 2009). For induction of α-adrenergic receptors, norepinephrine was used at 1μM (Spiegel et al., 2007). 5HT2 type receptors were stimulated by applying serotonin at 5μM (Ebrahimkhani et al., 2011). L-glutamate was used as an agonist which triggers both ionotropic and metabotropic type receptors at 10μM (Porras and Stutzin, 2011). Histamine, which is capable of activating all histamine receptors, was administered at 100μM (Szukiewicz et al., 2009).

6.6.1 The effect of neurotransmitters on LSC quiescence

The effects of neurotransmitters on the maintenance of quiescence and the regulation of cell division were measured by using CFSE intracellular fluorescence dye, as described in Chapter 2, and evaluating the number of cell divisions that the cells underwent after 72 hrs in comparison with the no drug control (NDC) (Figure 6.7).
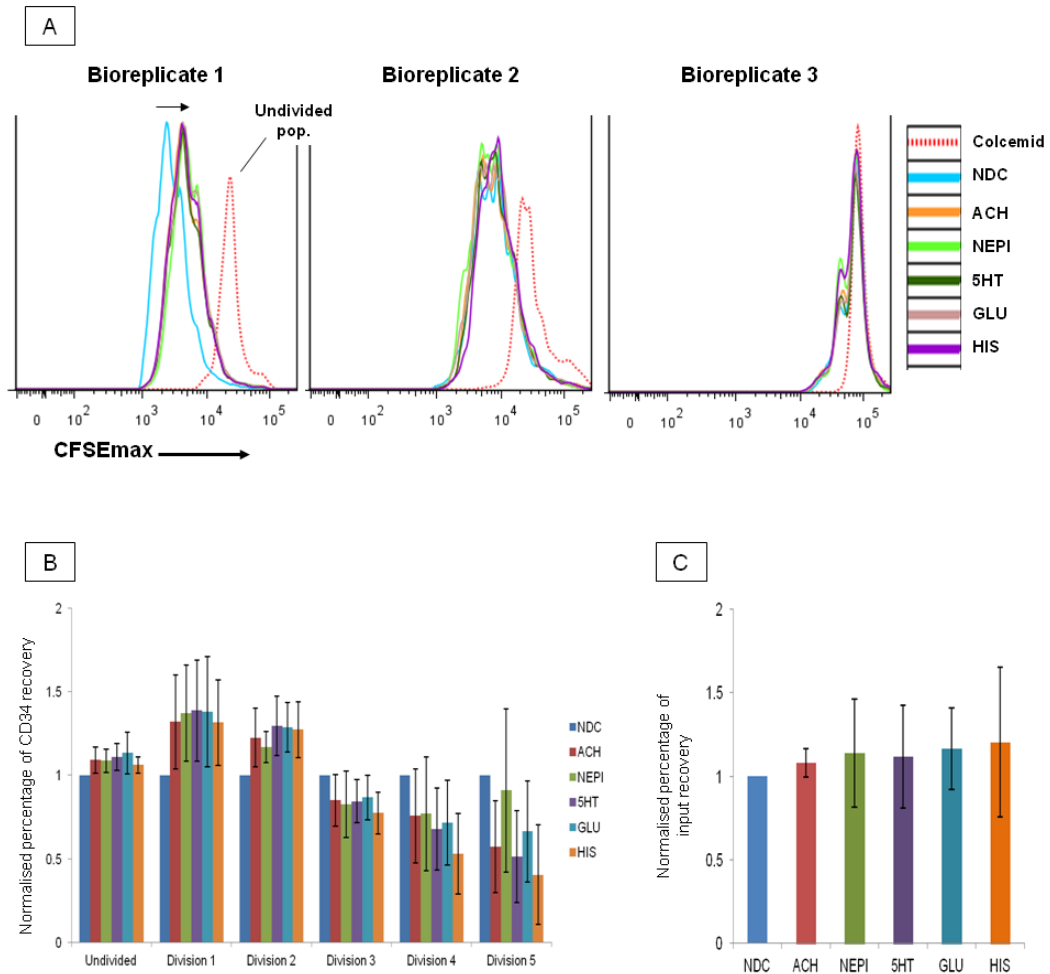
Figure 6.6: Analysis of cell division progression (CFSE) in the presence of neurotransmitters.
(A) Histograms representing the intensity of CFSE after 72 hrs in the presence or absence of agonists in three different CML samples. The dotted red line represents the intensity of CFSE in the colcemid-treated sample which indicates the position of the undivided populations. The no-agonist control (NDC) is shown in blue. The arrow indicates the shift in the fluorescence intensity (x-axis) of LSCs upon neurotransmitter treatments relative to NDC. (B) The percentage recoveries of CD34$^+$ cells in undivided and each cell division were normalised against the percentage of CD34$^+$ recovery in the NDC condition. (C) The total percentage recovery of the CD34$^+$ CD38$^-$ input cells was normalised against the total percentage recovery of input in the NDC condition. The data represent the average of three bioreplicates and the error bars indicate the standard error of mean. The non-overlapping error bars from each condition with the NDC value were assigned significant. ACH, acetylcholine; NEPI, norepinephrine; 5HT, serotonin; GLU, L-glutamate; HIS, histamine.

The fluorescence intensity of undivided populations was determined by the intensity of the colcemid-treated CFSE$^+$ LSCs which were blocked in their cell division progression by colcemid (Figure 6.7A). Intriguingly, the effect of neurotransmitters on the progression of cell division varied between CML samples due to the potential heterogeneity between LSC populations. The first bioreplicate showed a considerable shift towards the undivided population in the presence of neurotransmitters, whereas the other two bioreplicates only showed small shifts in the presence of agonists, possibly due to their less proliferative characteristics as could be observed by their distance from the undivided population.

When the results of all three bioreplicates were analysed together, the percentage of CD34$^+$ cells in the undivided population was only marginally higher than NDC, in the presence of ACH, NEPI, 5HT, and HIS (Figure 6.7B). The percentage of CD34$^+$ cells was noticeably higher in the first and second divisions of all neurotransmitter conditions as compared to NDC. Furthermore, 5HT, GLU, and HIS showed noticeable reduction of cells in the fourth and fifth divisions, in addition to ACH treatment which also showed a noticeable reduction of cells in the fifth division as compared to NDC. Any effect of NEPI treatment did not noticeably affect the percentage of CD34$^+$ cells in the third, fourth, and fifth divisions. The observed differences in the presence of agonists were not due to noticeable increases or decreases in the total percentage of recovered input cells (CD34$^+$ CD38$^-$) with respect to the NDC condition (Figure 6.7C); thus, excluding toxicity of the agonists at the administered concentrations as an explanation for the noticeable changes detected. The above observations suggested that the neurotransmitter treatments resulted in reduced cell division progression and induced quiescence as the percentage of CD34$^+$ cells was increased in earlier divisions and dropped at later divisions, particularly in the case of ACH, 5HT, GLU, and HIS. NEPI treatment, on the other hand, only increased the percentage of cells in the undivided population and the first two divisions without noticeably affecting later divisions.

6.6.2 The effect of neurotransmitters on colony forming potential and self-renewal

CFC assays were established after 72 hrs of treatment with agonists or NDC on a semi-solid methylcellulose medium. The total number of colonies generated after

10 days was counted under a microscope and the effect of each condition on increasing or decreasing colony forming potentials of LSCs was evaluated relative to NDC (Figure 6.9A). Furthermore, 50 colonies were plucked at random and replated onto a semi-solid methylcellulose medium to investigate their self-renewal capacities. The total number of colonies and their type in each condition were analysed under a microscope after 10 days and compared to NDC (Figure 6.9B-D).

NEPI was the only neuropeptide that significantly increased the total number of colonies in the CFC assays relative to NDC (Figure 6.9A). On the contrary, HIS was the only neurotransmitter that significantly reduced the colony forming potential of LSCs. Any changes under the other agonists did not achieve significance. ACH, NEPI, and 5HT enhanced the self-renewal capacities of LSCs as measured by the significantly increased number of colonies during replating assays (Figure 6.9B). NEPI was the only neuropeptide to significantly increase both granulocytic (CFU-GM) and erythroid (CFU/BFU-E) types of colonies, whereas ACH and 5HT treatments resulted only in the significant increase of erythroid colonies (Figure 6.9C-D). GLU treatment also caused a marginally significant increase in the number of erythroid colonies post-replating. Furthermore, the types of observed colonies in the ACH, 5HT, and NEPI treatments were bulkier and larger, which were indicative of more primitive types of colonies, than the types of colonies observed in the presence of GLU, HIS, or NDC (Figure 6.9E). The above results indicated a potential role in the self-renewal maintenance of LSCs for ACH, 5HT, and NEPI.
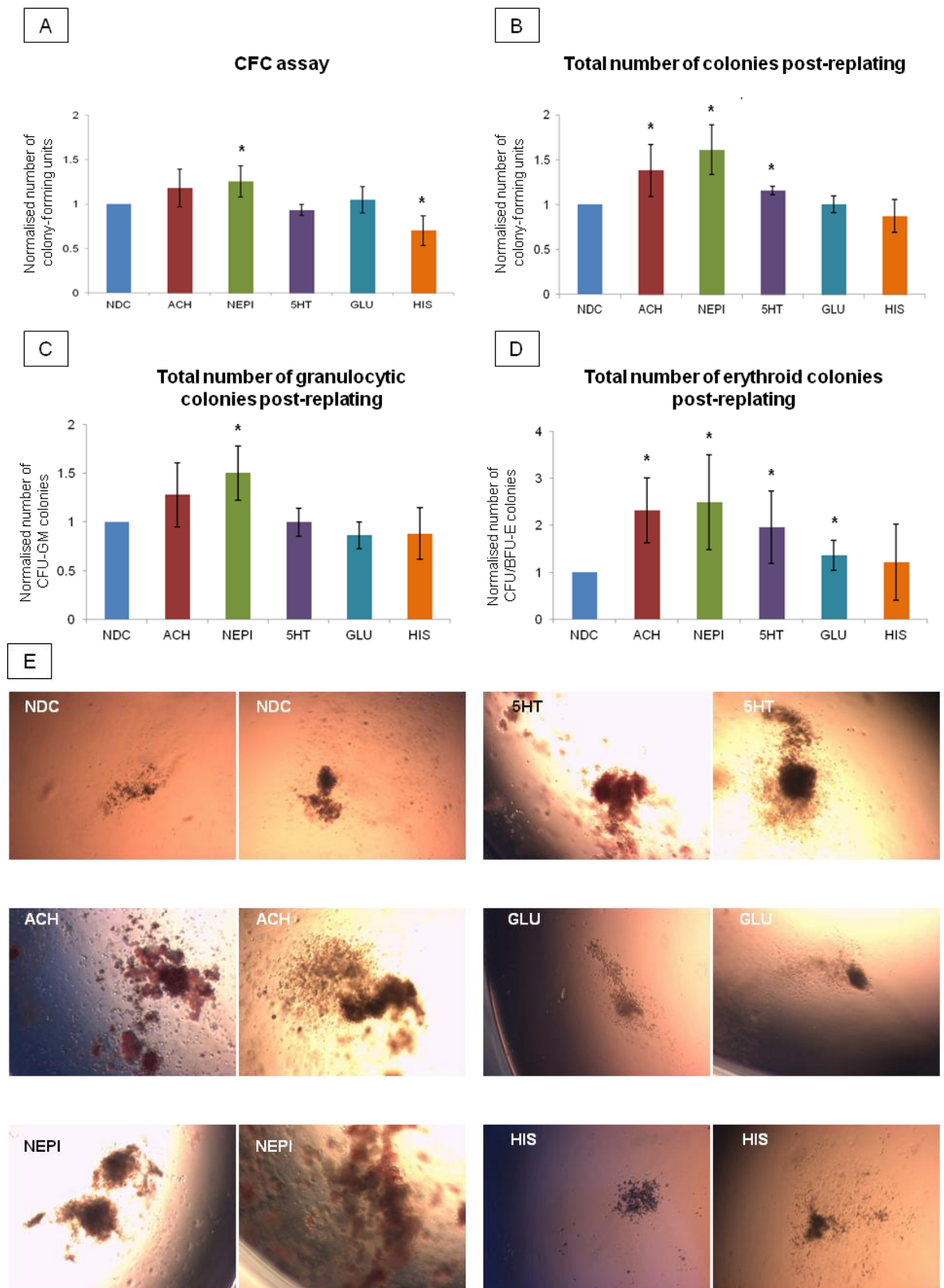
Figure 6.7: Investigating the colony-forming potential and self-renewal capacities of LSCs upon treatment with neurotransmitters.

(A) Averaged normalised total number of colonies generated in the CFC assays under each condition relative to NDC, which were obtained from three independent assays. The average total number of colonies in each condition was: NDC (n=144), ACH (n=181), NEPI (n=212), 5HT (n=155), GLU (n=147), HIS (n=134) (B) Averaged normalised total number of colonies post-replating from three independent assays. The average total number of colonies in each condition was: NDC (n=42), ACH (n=52), NEPI (n=60), 5HT (n=47), GLU (n=40), HIS (n=42). (C) Averaged normalised total number of granulocytic colonies (CFU-GM) post-replating from three independent

assays. The average number of CFU-GM colonies in each condition was: NDC (n=26), ACH (n=32), NEPI (n=39), 5HT (n=27), GLU (n=22), HIS (n=27). (D) Averaged normalised total number of erythroid colonies (CFU/BFU-E) post-replating from three independent assays. The average number of CFU/BFU-E colonies in each condition was: NDC (n=14), ACH (n=19), NEPI (n=17), 5HT (n=19), GLU (n=16), HIS (n=13). The error bars indicate standard error of mean. The error bars which did not overlap with the NDC value were assigned significant and marked with asterisks. (E) The post-replating colonies in each condition as observed under microscope (4X magnification). Two colonies are shown for each condition from two independent assays. The larger and bulkier that can be observed in the case of ACH, NEPI, and 5HT contained both CFU-GM and CFU-E colonies (CFU-GEMM), whereas the NDC, GLU, and HIS colonies were mainly associated with the smaller type of colonies of either CFU-GM or CFU-E.

6.6.3 The effect of neurotransmitters on the activation of heterotrimeric G protein signalling pathway

The effect of neurotransmitter treatment on the activation of heterotrimeric G protein-mediated signalling pathway as well as their unique downstream effectors was investigated by TaqMan RT-PCR performed on the Fluidigm gene expression platform, as described in Chapter 2. ACH, NEPI and 5HT neurotransmitters noticeably enhanced the self-renewal capacity of LSCs. Consequently, in order to further elucidate the mechanism of self-renewal and quiescence induction by these neurotransmitters the differential expression of the genes in the heterotrimeric G protein signalling pathway and several cell cycle regulators were analysed in $CD34^+$ $CD38^-$ CML cells treated with ACH, NEPI or 5HT and compared to NDC (Figure 6.10).
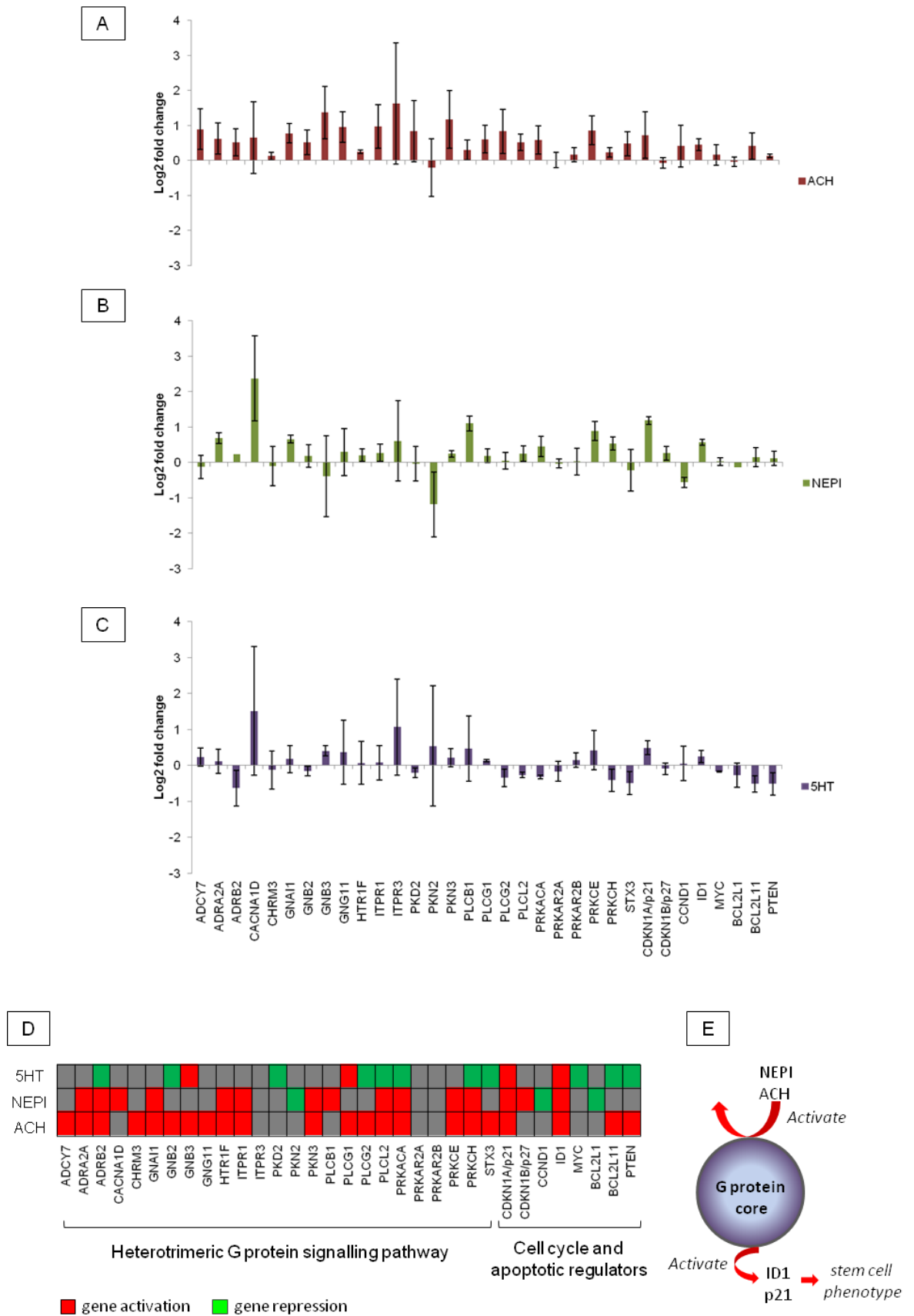
Figure 6.8: Gene expression analysis of the core heterotrimeric G protein signalling pathway in response to neurotransmitter treatments.
This histograms demonstrate the $\log_2$ fold change of expression for the genes in the heterotrimeric G protein signalling pathway and cell cycle and apoptotic regulators in the presence of (A) ACH, (B) NEPI, and (C) 5HT relative to NDC. The error bars indicate the standard error of mean. The

error bars which were not overlapping with the x-axis were considered as significant. (D) The schematic representation of the activated (red) and repressed (green) genes in each condition. The grey boxes indicate no significant change relative to NDC. (E) The proposed model of heterotrimeric G protein core activation by neurotransmitters which resulted in the upregulation of ID1 and p21/CDKN1A genes in LSCs.

The gene expression analysis of core G protein signalling pathway (Figure 6.10) showed that ACH treatment induced the expression of 18 genes out of 25 tested genes in this pathway (72%), including the muscarinic cholinergic receptor 3 (CHRM3). Furthermore, NEPI induced the expression of 13 genes (52%) in the G protein signalling pathway, including alpha-2a adrenergic (ADRA2A) and beta-2 adrenergic (ADRB2) receptors. ACH and NEPI induced the expression of several PKC subunits such as PRKCE and PRKCH as well as the cAMP-dependent protein kinase catalytic subunit alpha (PRKACA), which is downstream of cAMP second messenger. However, NEPI treatment showed the specific activation of PLCβ (PLCB1) while ACH treatment specifically activated PLCγ genes, PLCG1 and PLCG2, which indicated a preferential usage of the downstream effectors by these neuropeptides. On the contrary, the 5HT treatment resulted largely in the repression of G protein core pathway by silencing eight genes (32%) as opposed to activating only two genes. PRKACA, PRKCH, and PLCG2 showed significant downregulation upon 5HT treatment, whereas the other PLC and PKC genes did not change significantly, which suggested a mechanism of action for 5HT independent of PKA and PKC activation.

ACH, NEPI, and 5HT treatment resulted in the upregulation of p21/CDKN1A and ID1, which are important in the maintenance of self-renewal and quiescence in HSCs (Cheng et al., 2000, Perry et al., 2007). Therefore, it was plausible to deduce a common mechanism of action for these neurotransmitters in conferring stem cell-like phenotype to LSCs. Additionally, NEPI upregulated p27/CDKN1B, another cell cycle inhibitor important in HSC maintenance (Dao et al., 1998), and downregulated CCND1, which promotes the cell cycle progression.

The above observations led to a model (Figure 6.10E) in which ACH and NEPI activate the core heterotrimeric G proteins which in turn result in the activation of p21/CDKN1A and ID1 genes that maintain the stem cell phenotype of LSCs, as shown by their induced quiescence and self-renewal capacities. However, while this model supports the data obtained from the expression analysis, the mechanism by which this occurs could not be elucidated.

6.7 Discussion

Several novel neurotransmitter pathways with potential role in the maintenance of stem cell features in LSC were identified in this study at the level of gene expression and epigenetics. These neurotransmitters included ACH, NEPI, 5HT, HIS, and GLU. In this Chapter the role of these neurotransmitters in the regulation of LSCs self-renewal and quiescence was investigated *in vitro* in the presence of stromal cells. It was demonstrated that these neurotransmitters reduced the cell division progression rate of LSCs and subtly increased the number of cells in G_0 compartment in comparison to untreated condition. Furthermore, ACH, NEPI and 5HT treatments increased the self-renewal capacities of LSCs as observed in replating assays. The gene expression analyses of LSCs treated with ACH and NEPI showed significant upregulation of several genes in the heterotrimeric Gα protein signalling pathway as well as cell cycle regulators p21/CDKN1A and ID1.

6.7.1 Neurotransmitters appear to reduce the cell division progression rate of LSCs

The above results, therefore, suggest a potential role for the studied neurotransmitters in reducing cell division progression rate. Ideally with the heterogeneity observed in terms of response further samples should be analysed and an attempt made to uncover any features of individual samples or patients that may make them more or less susceptible to neurotransmitter signalling.

Although the above results indicated that CML cells could progress, albeit at a slower rate, through cell division even in the presence of neurotransmitters, the self-renewing or differentiating nature of these divisions was not investigated here but would be very interesting to explore. Numb, a protein that can be inherited asymmetrically during differentiation, was used in a recent CML study and it was demonstrated that the low expression levels of Numb were associated with the more primitive cell types, whereas its upregulation was associated with differentiation mechanisms (Ito et al., 2010). To further investigate the influence of neurotransmitters on the ratio of self-renewing to differentiating cell divisions, the presence of membrane-bound Numb could be investigated during agonist treatments by immunofluorescence assays and compared to the NDC condition. However, due to the limited number of LSCs per condition and time constraints

immunofluorescence staining was not performed in parallel with the other downstream assays in this body of work.

6.7.2 ACH, NEPI, and 5HT appear to increase the self-renewal capacities of LSCs

ACH, NEPI, and 5HT agonists caused significant increases in the total number of colonies post replating which indicated their potential role in the maintenance of LSCs self-renewal. NEPI was the only neuropeptide to significantly increase both granulocytic (CFU-GM) and erythroid (CFU/BFU-E) types of colonies after replating, whereas ACH and 5HT significantly induced differentiation towards erythroid colonies.

Previous studies provided evidence for the role of adrenergic signalling pathways in haemopoiesis. In a study by Fonesca et al, an increased level of erythropoiesis was observed upon treatment of BM MNCs *in vitro* with NEPI at 100 nM, whereas higher concentrations ($\geq$1 mM) completely blocked erythropoiesis (Fonseca et al., 2005). Furthermore, stimulation of adrenergic signalling pathway by NEPI in human HSCs induced BM repopulation via activating the canonical Wnt signalling pathway (Spiegel et al., 2007). Overall, the role of neurotransmitters in the maintenance of stem cell features in HSCs have not been fully elucidated and the work presented in this Chapter was the first reported account of neurotransmitters in enhancing self-renewal of LSCs.

It should also be noted that the regulation of haemopoiesis by neurotransmitters potentially involves a complex interconnected network of direct and indirect mechanisms as the neurotransmitter receptors are identified on both haemopoietic cells and stromal cells in the niche. For example, α1a-adrenergic receptors were found on the surface of BM stromal cells, while α2-adrenergic receptors were found on the surface of both MNCs and stromal cells (Pereira et al., 2003). Therefore, the presence of neurotransmitter receptors could be investigated on the surface of OP9 cells in the future.

6.7.3 ACH and NEPI activate the core G protein signalling pathway in LSCs

The expression analysis of the genes in the heterotrimeric G protein signalling pathway upon neurotransmitter treatments revealed the activation of the G protein signalling pathway in response to ACH and NEPI, whereas 5HT treatment induced repression of the genes in this pathway. ACH and NEPI activated several PKA, PKC, and PLC genes, therefore, suggesting an activation mechanism via $G_{\alpha s}$ and $G_{\alpha q}$ subunits. α- and β- adrenergic receptors as well as muscarinic receptors 1 and 3 mediate their function via $G_{\alpha s}$ and $G_{\alpha q}$ subunits (Table 6.1). There was also evidence for upregulation of α- and β- adrenergic receptors in NEPI treatment and muscarinic receptor 3 in ACH treatment, which could suggest a role for these receptors in the maintenance of LSCs. 5HT treatment, on the contrary, reduced the expression of PKA, PKC, and PLC genes, except PLCG2, which could be linked to the activation of $G_{\alpha i/o}$ subunits as a result of stimulation of 5HT1 and/or 5HT5 receptors (Table 6.1). ACH, NEPI, and 5HT agonists induced the expression of p21/CDKN1A and ID1 genes that are important quiescence and self-renewal markers (Cheng et al., 2000, Perry et al., 2007). The gene expression analysis could, therefore, explain the observed CFSE and replating assays results in which ACH NEPI, and 5HT induced quiescence and self-renewal of LSCs.

6.7.4 Further functional validation of neurotransmitter involvement in LSC maintenance

The results presented in this Chapter were obtained in *ex vivo* experiments which only partially mimicked the bone marrow niche. However, further *in vivo* functional validation is required where the effect of physiologically secreted neurotransmitters could be reverted by the administration of antagonists to the above neurotransmitter pathways. Furthermore, the biological impact of neurotransmitters is mediated in a dose- and gradient-dependent manner which could oscillate in response to various stimuli such as stress, which could not be fully simulated *in vitro* (Kalinkovich et al., 2009). However, only NEPI was used at concentrations that were validated previously in the context of haemopoietic cells, whereas the other agonists were used at concentrations that showed a physiological response in other human tissues. Therefore, some of the observed effects could be a result of over- or under-exposure to the agonists which could be

overcome in the future by testing different concentrations and/or replenishing the agonists at different intervals.

As mentioned earlier, various neurotransmitter receptors were also identified on the surface of stromal cells which could be stimulated upon treatment and result in the release of various factors that could indirectly regulate the stem cell features of LSCs. Therefore, to further understand the effect of stromal cells in the context of neurotransmitter stimulation, genome-wide gene expression changes of OP9 cells could be investigated in the presence and absence of agonists. Furthermore, in order to elucidate the genes and the biological pathways that become activated in response to the neurotransmitter stimulus, genome-wide gene expression analysis of the treated LSCs should be performed.

Although the gene expression and epigenetics analyses identified neurotransmitter signalling via specific receptors, the functional validation was performed with the generic physiological ligand that could stimulate various receptors and potentially lead to the activation of different intracellular pathways. Nevertheless, after the initial round of functional validation by the generic agonists, more specific biochemical compounds could be used to trigger the specific receptors. A recent study investigated the role of neurotransmitter receptors in brain cancer stem cells by testing 1,267 chemical compounds that could specifically block or activate these receptors (Diamandis et al., 2007). This study demonstrated that dopamine and acetylcholine antagonists as well as an opioid agonist have profound impact on inducing apoptosis in brain cancer stem cells.

In this Chapter, the effect of neurotransmitters on the major features of stem cell maintenance, such as quiescence, self-renewal, haemopoietic growth and proliferation were examined, whereas the other aspects of stem cell life cycle including motility, migration, and repopulation capacities upon engraftment were not evaluated. The migration potentials of LSCs upon neuropeptide treatments could be investigated by transwell migration assays (Aiuti et al., 1997) in which cells are placed in a small chambers above permeable membranes and the media containing the agonists are placed below the membranes. Subsequently the number of cells that pass through the membranes in an overnight incubation period are counted as a measure of induced motility. A recent study demonstrated that NEPI treatment can induce HSCs motility in response to both G-CSF and GM-

CSF cytokines (Spiegel et al., 2007). Finally, the enhance self-renewal capacity and repopulation potential of LSCs upon treatment with neurotransmitters could be investigated *in vivo* by using NOD/*scid* mice.

7. Summary and Future work

7.1 Summary

Understanding the underlying mechanisms conferring a stem cell-like phenotype to LSCs is one of the major questions to be addressed in leukaemia biology in order to determine novel strategies required for more effective therapeutics that can eradicate the disease. The study of LSC biology is of significant importance in CML as the tumour initiating population is insensitive to the current TKIs due to the function of BCR-ABL1-independent mechanisms, which could involve epigenetic aberrations that give rise to a persistent leukaemic phenotype. Nevertheless, an enhanced understanding of CML biology will not be possible without unravelling the essential mechanisms involved in the maintenance of HSCs and their lineage commitment. Therefore, the aim of the work presented in this thesis was to obtain a global view of the epigenetic programming that is responsible for the gene expression signatures of LSCs and the expression changes required for the exit from the stem cell compartment. The analysis of genome-wide histone modification patterns required the development of a robust and reproducible ChIP-seq method for profiling small primary stem cell populations.

The work presented in this thesis showed the potential role of HSC identity pathways in regulating the stem cell features of LSCs, whose subsequent downregulation triggers the progression to the more proliferative states. These pathways are significantly enriched for bivalent/H3K27me3^{hi} promoters which were found to be abnormally regulated at the epigenetic level in CML cells which could be due to the abnormal function of the chromatin modifying machinery. Nevertheless, the underlying mechanisms resulting in the misregulation of chromatin modifiers could not be established. These results provide a unique window of opportunity to epigenetically target LSCs and also to identify new targets in some of the newly identified neurotransmitter signalling pathways that enhanced the stem cell features of LSCs *in vitro*.

7.2 Future work

7.2.1 Identifying the EZH1/2 and KDMs targets

To better understand the mechanism of EZH2 suppression in normal and leukaemic cells as well as the role of EZH1-EZH2 imbalance in the context of LSCs, their gene promoter targets should be investigated in the four cell types to elucidate: (i) the genes and the pathways that are specifically targeted by EZH1 or EZH2 and whether there is an overlap between the targets, (ii) whether the loss of EZH1 in LSCs can be compensated by EZH2 and how the expression levels of EZH1 targets are affected, (iii) the genes that are targeted by EZH2 upon exit from the stem cell compartment in LPCs, (iv) the similarities and differences between EZH2 targets in LPCs and HPCs and whether EZH2 functions differently in the context of leukaemic cells compared to normal cells. It was previously reported in embryonic carcinoma cells that EZH1 and EZH2 share many targets in addition to their unique targets (Margueron et al., 2008). The analysis of EZH1/2 targets can be achieved by ChIP-seq, although the low-cell ChIP-seq assays for non-histone proteins that bind DNA less frequently require further improvements, such as antibody titrations and adjusting the fixation step.

On the other hand, abnormalities in the levels of H3K4 demethylases were suggested to have an impact on the loss of repressive capacity of bivalent/H3K27me3^{hi} promoters in LSCs (Chapter 5). Therefore, in order to further understand the role of H3K4 demethylases in abnormal regulation of bivalent/H3K27me3^{hi} promoters in LSCs, the identification of their target genes is crucial. Therefore, their global enrichment patterns should be established in HPCs to identify their binding sites and to compare their profiles in LSCs and LPCs to understand the effect of their downregulation in the leukaemic cells. Furthermore, although KDM1A is linked to H3K4me1/2 marks (Mosammaparast and Shi, 2010) it was also found to be enriched at the bivalent promoters and its loss resulted in abnormalities in the H3K4me3 levels (Adamo et al., 2011). Therefore, its upregulation in LPCs could suggest a potential compensating mechanism in the event of KDM5A and KDM5B downregulation. In order to distinguish the potential links between KDM1A upregulation and KDM5A/5B downregulation, the global binding sites of KDM1A in LPCs and HPCs should be investigated to identify the overlaps with KDM5A and KDM5B targets, and to establish a preferential use for

KDM1A in LPCs that may not be the case in their normal counterparts. ChIP-seq assays should be used for identification of the KDM targets, although the low-cell ChIP-seq assays require further improvements as mentioned above.

Furthermore, in order to distinguish between EZH2-dependent and –independent H3K4 demethylation, the overlap between KDMs and EZH2 targets should be investigated from the ChIP-seq analyses. Co-immunoprecipitation assays should be used to determine a direct association between the H3K4 KDMs and EZH2-PRC2 complex to confirm the EZH2-dependent route, in both normal and leukaemic cell types, as LSCs and LPCs could have adopted different mechanisms in KDMs recruitment. KDM5A was the only H3K4 KDM that is shown to be directly recruited by the EZH2-PRC2 complex in mES cells (Pasini et al., 2008), which may or may not be the case in haemopoietic cells.

7.2.2 The underlying events responsible for the deregulation of PRC2 complex

The abnormalities in the levels of PRC2 complex members could be the result of BCR-ABL1-dependent or –independent factors, or a combination of both. The downregulation of EZH1 and upregulation of EZH2 could represent events preceding the 9;22 translocation in HSCs, or could be mediated by BCR-ABL1 indirectly, as PRC2 members are regulated at the transcriptional level. In order to test the first hypothesis, one copy of the EZH1 gene should be conditionally knocked out in murine HSCs. If a CML-like phenotype was observed, deregulation of the PRC2 complex is independent of BCR-ABL1 and potentially precedes the 9;22 translocation, otherwise the abnormalities are most likely driven by BCR-ABL1 expression. Furthermore, the same model could be used to investigate expression levels of EZH2 and other PRC2 members, such as SUZ12, to elucidate whether the upregulation of EZH2 in LSCs is part of a feedback mechanism that is initiated upon downregulation of EZH1, or whether the two events are regulated independently. To test the BCR-ABL1-dependent mechanism, the expression levels of PRC2 complex members should be analysed upon BCR-ABL1 knockdown by lentiviral shRNA transduction in LSCs. The downregulation of EZH2 and/or upregulation of EZH1 would then indicate a BCR-ABL1-dependent mechanism.

The expression of PRC2 members, other than EZH1, was shown to be upregulated in many cancers by the upregulation of E2F transcription factors (Bracken et al., 2003). E2F1 levels are reported to be significantly upregulated in CML cells (Mohty et al., 2007, Stewart et al., 1995) and therefore, the upregulation of EZH2 could be mediated by the increased levels of E2F1 in a BCR-ABL1-dependent manner. This hypothesis could be further investigated by the lentiviral shRNA knockdown of E2F1 in CD34$^+$ CML cells and by observing the effect on EZH2 and other PRC2 members. Nevertheless, murine models with E2F1, E2F2 and E2F3 triple knockouts may represent better *in vivo* models due to partial redundancies amongst these proteins (Chong et al., 2009).

Another recent study proposed a link between EZH2 upregulation in cancers and the downregulation of microRNA-101 (MIR101) (Varambally et al., 2008). To test whether the abnormal upregulation of EZH2 in CML is mediated by the same mechanism, the expression levels of MIR101 could be established in leukaemic cells relative to normal cells and in the case of a negative correlation with EZH2 levels, the reintroduction of MIR101 into LSCs or LPCs may result in the downregulation of EZH2 and/or upregulation of EZH1.

7.2.3 Targeting the EZH2-dependent LSCs

A significant drop of EZH1 and an insufficient upregulation of EZH2 in LSCs indicated an imbalance between the two catalytic subunits in LSCs, which was suggested to be responsible for the abnormal epigenetic regulation of bivalent/H3K27me3^{hi} promoters in these cells. Therefore, EZH2 is critically required in LSCs in order to maintain the levels of H3K27me3 in the absence of EZH1 at the promoters of bivalent/H3K27me3^{hi} genes which were significantly associated with the HSC identity pathways. Conversely, EZH1 was shown to be the highly expressed catalytic subunit of PRC2 complex in HSCs as EZH2 was expressed at relatively low levels, which in turn provides a unique therapeutic window to block EZH2 in LSCs. 3-Deazaneplanocin A (DZNep) is an anti-EZH2 agent that is currently under investigation as an anti-cancer therapeutic (Tan et al., 2007).

A recent study demonstrated interesting effects of DZNep on AML cells as it induced apoptosis in primary CD34$^+$ cells, alone or in combination with an HDAC

inhibitor, Panobinostat, although the more primitive CD34$^+$ CD38$^-$ population was insensitive to DZNep alone. However, the pro-apoptotic effect of DZNep was not observed with normal CD34$^+$ cells (Fiskus et al., 2009). Although this study failed to demonstrate the expression levels of EZH2 in AML cells relative to normal, it suggested drug specificity towards AML versus normal progenitor cells, which was associated with a global decrease in the levels of EZH2 and H3K27me3 and an increase in the levels of CDKIs p16, p21, and p27. Therefore, treating CML LSCs or LPCs with DZNep could have four potential outcomes: (i) induced apoptosis in more proliferative cells as shown in AML (Fiskus et al., 2009); (ii) increased activity of the stem cell programme as EZH2 is required for the exit from a non-dividing state by suppressing the pluripotency factors as previously reported (Margueron et al., 2008, Shen et al., 2008); (iii) induced differentiation as recently reported in the case of AML where EZH2 overexpression was associated with suppression of differentiation mechanisms (Tanaka et al., 2012); or (iv) have no effect due to the redundancy between EZH1 and EZH2 and potential compensation by EZH1 upon EZH2 inhibition. However, the latter is less likely to happen since the EZH1 levels was observed to be at low levels in LSCs and the inhibition of EZH2 is unlikely to result in adequate compensation by EZH1. The effect of DZNep on LSCs can be examined *in vitro* by evaluating cell division progression, apoptosis, colony forming potentials, and self-renewal capacities upon replating in the presence and absence of drug, in parallel to studies on normal stem cell controls. Furthermore, the effect of DZNep treatment on the global epigenome changes could be monitored by ChIP-seq analysis of H3K4me3 and H3K27me3 marks in combination with genome-wide gene expression analysis.

7.2.4 Using epigenetic biomarkers in diagnosis, prognosis, and therapeutic response of CML

The identification of abnormalities in several members of PcG, TrxG, and KDM families of chromatin modifiers could be used as epigenetic biomarkers in CML diagnosis, prognosis and patient response to TKI therapy. The reported abnormalities were only based on the observations in three chronic phase patients using Affymetrix technology. In order to employ such biomarkers in the clinical context the expression of these genes should be analysed in more chronic phase patients upon diagnosis as well as in patients in advanced phases of the disease, such as blast crisis, using more sensitive RT-PCR-based approaches. Moreover,

changes in the expression levels of these chromatin modifiers could be evaluated during TKI therapy and after achieving remission to monitor the response to TKIs and potentially as markers of disease recurrence.

The overall levels and the global enrichment patterns of H3K27me3 could also be used as epigenetic biomarkers, since CML cells showed significant changes in the levels of H3K27me3 at the promoters to alter transcriptional activity of controlled loci. The global levels of histone modifications have recently been reported as prognostic tools in several cancers (Seligson et al., 2009) and in particular the loss of H3K27me3 marks is associated with poor prognosis in pancreatic, breast, and ovarian cancers (Wei et al., 2008).

7.2.5 Further validation of the role of neurotransmitters in LSCs maintenance

The results presented in Chapter 6 confirmed the role of ACH, NEPI and 5HT in the maintenance of LSCs in the presence of stromal cells *in vitro*. Although the overall expression levels of these pathways were not significantly different between HSCs and LSCs, the expression levels of some of the receptors in these pathways could be significantly misregulated in LSCs, which can be identified from the Affymetrix analyses and further validation by RT-PCR-based methods. Moreover, the presence of these receptors at the cell surface of LSCs can be validated by flow cytometry using indirect immunofluorescence staining with primary antibodies against specific receptors followed by the secondary staining with rhodamine-conjugated antibodies (Kronenwett et al., 2005, Steidl et al., 2004). By comparing the immunofluorescence levels of HSCs and LSCs, the receptors which are presented at higher levels on the cell surface of LSCs can be distinguished to design specific antagonists which block the *in vivo* activation of these receptors and potentially inhibit maintenance of stem cell features by neurotransmitter signalling pathways.

The effects of these neurotransmitters are dose- and gradient-dependent that can only be regulated physiologically. Furthermore, the secretion of neurotransmitters in the niche can also indirectly control the stem cell features of LSCs by controlling the secretion of various cytokines and extracellular signals by the cells in the BM niche (Pereira et al., 2003). Therefore, the role of neurotransmitters in LSC maintenance can only be fully elucidated *in vivo*. The receptors that are uniquely

upregulated in LSCs could be knocked down by the lentiviral shRNA transduction system followed by engraftment of LSCs into NOD-*scid* mice to measure their effect on LSC survival, proliferation, and repopulation capacities in comparison to a control population. The receptors that show a significant change of phenotype upon knockdown could be used to develop antagonists which in turn provide new therapeutic approaches.

7.3 Final thoughts

The work presented in this thesis describes the chromatin landscape of cancer stem cells for the first time, which was only made possible by the development of a ChIP-seq method that can be applied to small primary cell populations. The observations contained here can open new therapeutic windows using epigenetic targets as well as the neurotransmitter signalling pathways, which are shown for the first time to be important in the context of leukaemia biology. The low-cell ChIP-seq method developed in this thesis can be used to study the distribution of other histone modifications to unravel other regulatory elements, such as enhancers and potentially transcription factor binding sites in LSCs. Finally, the outcomes of this thesis will contribute significantly towards developing new therapeutics that eliminate cancer stem cells and eventually improving the life expectancy of cancer sufferers.

Appendix 1

List of the TaqMan probes used to design the TaqMan Micro Fluidic cards
(Chapter 3)

Gene Symbol	Assay ID*	Gene Symbol	Assay ID*	Gene Symbol	Assay ID*
ABL1	Hs01104728_m1	FOXO1	Hs01054576_m1	PIK3C2B	Hs00153248_m1
ACVR2A	Hs00155658_m1	FOXO3	Hs00818121_m1	PIK3CB	Hs00927728_m1
ACVR2B	Hs00609603_m1	FOXP1	Hs00212860_m1	PIK3CG	Hs00277090_m1
AKT1	Hs00178289_m1	GAB2	Hs00202306_m1	PIK3R1	Hs00933163_m1
AKT2	Hs01086102_m1	GADD45A	Hs00169255_m1	PRKCA	Hs00925195_m1
AKT3	Hs00178533_m1	GAPDH	Hs02758991_g1	PTEN	Hs02621230_s1
AMHR2	Hs00179718_m1	GDF11	Hs00195156_m1	PTPN11	Hs00761486_s1
BCL2L1	Hs00236329_m1	GDF15	Hs00171132_m1	RAB10	Hs00211643_m1
BCR	Hs00244716_m1	GDF3	Hs00220998_m1	RAF1	Hs00234119_m1
BMP4	Hs00370078_m1	GLIPR1	Hs01564143_m1	RASSF6	Hs00698249_m1
BMP6	Hs01099594_m1	GNA12	Hs02863396_m1	RB1	Hs01078066_m1
BMP8B	Hs01629120_s1	GNAI1	Hs01053355_m1	RBBP7	Hs00171476_m1
BMPR1B	Hs00176144_m1	GNB2	Hs00366502_g1	RBL2	Hs00180562_m1
CBLB	Hs00180288_m1	GNG5	Hs00359134_g1	RHEBL1	Hs00376020_m1
CBX2	Hs01034268_m1	GRAP2	Hs00191325_m1	RPLP0	Hs99999902_m1
CBX7	Hs00545603_m1	HIPK2	Hs00179759_m1	SERPINE1	Hs01126604_m1
CCND1	Hs00765553_m1	IKBKB	Hs00395088_m1	SHC4	Hs00736166_m1
CDC25A	Hs00947994_m1	ILK	Hs00177914_m1	SMAD1	Hs00195432_m1
CDK4	Hs00262861_m1	INPP5D	Hs00183290_m1	SMAD2	Hs00183425_m1
CDKN1B	Hs00153277_m1	ITGB3	Hs01001469_m1	SMAD3	Hs00969210_m1
CHUK	Hs00175141_m1	LTBP1	Hs00386448_m1	SMAD4	Hs00929647_m1
DNMT1	Hs00154749_m1	LTBP4	Hs00186025_m1	SMAD6	Hs00178579_m1
E2F1	Hs00153451_m1	MAN1A1	Hs00195458_m1	SOS1	Hs00362308_m1
E2F2	Hs00231667_m1	MDM2	Hs00234753_m1	SUZ12	Hs00248742_m1
E2F3	Hs00605457_m1	MECOM	Hs00602795_m1	TGFB1	Hs00998133_m1
EZH1	Hs00940463_m1	MTCP1	Hs00180870_m1	TGFBR1	Hs00610319_m1
EZH2	Hs01016789_m1	MYC	Hs00905030_m1	TGFBR2	Hs00559660_m1
FCGR2A	Hs01017702_g1	NOG	Hs00271352_s1	TGFBRAP1	Hs00188614_m1
FKBP1B	Hs00188021_m1	PCGF6	Hs00827882_m1	TGIF1	Hs00820148_g1
FOXJ3	Hs00208978_m1	PDK2	Hs00176865_m1	TP53	Hs01034249_m1
FOXM1	Hs01073586_m1	PDPK1	Hs00176884_m1	VCAN	Hs00171642_m1
				YWHAG	Hs00705917_s1

 * Catalogue of TaqMan gene expression assays available from Applied Biosystems website (http://bioinfo.appliedbiosystems.com/genome-database/gene-expression.html)

Appendix 2

List of the TaqMan probes used for the Fluidigm analysis (Chapter 6)

Gene Symbol	Assay ID*	Gene Symbol	Assay ID*
ADCY7	Hs00936808_m1	MYC	Hs00905030_m1
ADRA2A	Hs01099503_s1	PKD2	Hs00165517_m1
ADRB2	Hs00240532_s1	PKN2	Hs00178944_m1
BCL2L1	Hs01067345_g1	PKN3	Hs00179481_m1
BCL2L11	Hs01076940_m1	PLCB1	Hs00248563_m1
CACNA1D	Hs01073321_m1	PLCG1	Hs01008225_m1
CCND1	Hs00765553_m1	PLCG2	Hs00182192_m1
CDKN1A	Hs00355782_m1	PLCL2	Hs00392897_m1
CDKN1B	Hs00153277_m1	PRKACA	Hs00923827_mH
CHRM3	Hs00327458_m1	PRKAR2A	Hs00177760_m1
GNAI1	Hs01053355_m1	PRKAR2B	Hs00176966_m1
GNB2	Hs00929275_g1	PRKCE	Hs00178455_m1
GNB3	Hs01564092_m1	PRKCH	Hs00178933_m1
GNG11	Hs00914578_m1	PTEN	Hs03673482_s1
HTR1F	Hs00265296_s1	STX3	Hs00188210_m1
ID1	Hs00357821_g1	GAPDH	Hs99999905_m1
ITPR1	Hs00181881_m1	GUSB	Hs99999908_m1
ITPR3	Hs01573555_m1		

*Catalogue of TaqMan gene expression assays available from Applied Biosystems website (http://bioinfo.appliedbiosystems.com/genome-database/gene-expression.html)

Appendix 3

Sequences of primer pairs used for the mouse contamination screen

Gene Symbol	Forward Primer	Reverse Primer	Amplicon size (bp)
hCYP4A22	AGCCTTCCACAATGACATCC	CATGGACTCACCAGCATCAC	78
hTAL1	TGTTGGGAAAGGCGAATAGT	TGGAATTGGGTTCTGGTTTC	130
mTal1	CGGAGGAGTGGAGGTCCTA	ACTCCAGGCCGCAACTAAC	72
mHoxA11	CTACCCAGGCCTTCACTCAG	CTAGCATCCCTACCCTGCTG	72

Appendix 4

Sequences of primer pairs used in the qPCR validation of ChIP-seq datasets

Gene Symbol	Forward Primer	Reverse Primer	Amplicon size (bp)	Chromosome	Co-ordinate start
AKT1	TAAAGTGACCCGGCGCTTCT	AGCCTGAAAGTCAACCTAAGCCAC	132	14	105261680
AKT2	ACTCTGAGAGAGCGAGAGCG	GAGGCCCATCTCTAACCCA	111	19	40791102
AKT3	ATGACTCAGCCTGGAAGGAAGT	ATGTGAGGTCAGGCCAGTTGAA	118	1	244006353
AMHR2	TGCAGAGCAAAGAGGAGGTT	TCATCCTATCCCAGCCTCTG	139	12	53817441
BCL2L1	GGAGGTGGCTGGTATGGAT	AGTCCCTTTAGGGTTTCGGA	100	20	30310501
BMP4	GCTAGATACTCAGCCGGAGC	GAAGATGCGAGAAGGCAGAG	135	14	54423354
BMP6	CAAGCGGTTCTCCTGGTGAT	TTCCTGGCCCTCAGTCCTT	100	6	7726830
BMP8B	CTGGCTCCTGGACGAGAGGA	ACAGACGGATTGGCCGAGAGT	86	1	40254233
BMPR1B	TTGAGAAGTGTAATCCCTCCGC	CGGCTGCCTCTGGAATTT	102	4	95678919
CBLB	CTCCCGAAGAAGGAGCAAC	TCTCCTCCTCCTGCTAGTGC	148	3	105587687
CCND1	CAACAGTAACGTCACACGGAC	ACTCTGCTGCTCGCTGCTA	100	11	69455673
CDC25A	AATCCAAACAAACGTGGCGGGT	AAGAGGTGTAGGTCGGCTTGGTT	142	3	48229646
CDK4	GACAGGAGGTGCTTCGACTG	GTAGCCACACCTCTGCTCCT	129	12	58145964
CDKN1B	CTTCTTCGTCAGCCTCCCT	AGCCGCTCTCCAAACCTT	145	12	12870197
CHUK	CGCGAGAATGAATGCGTCACTT	TTAAAGAGCGGACCGAGTACCT	129	10	101989176
DNMT1	TGCCCGAGGCATTCATTCATTC	AGGACTTTACTCAAGGGCTCTCAC	96	19	10305555
EZH2	GCTGTAAGGGACGCCACTG	CCGTGTGTTCAGCGAAAGA	70	7	148581070
FKBP1B	GATGCGCGAATGAATGGA	GCCTGACGACATTAACCCAC	79	2	24272371
FOXO1	GACGGAAACTGGGAGGAAG	GAGGAGCCTCGATGTGGAT	85	13	41240434
FOXP1	TAAATAAGGCGCGCGAGAG	GTTTACCCAGCAGGTGGATG	100	3	71632940
GAB2	TAGGGACCCGCGAGGATACAGAAA	GGCCTCAGGAGAAAGCGAGACT	70	11	78128552
GADD45A	AAATTGCTGGAGCAGGCTGA	CAACCACCTTCCGCCTCATTT	71	1	68150544
GDF15	GAACTCAGGACGGTGAATGG	GGAGTGCAACTCTGAGGGTC	132	19	18496768
GLIPR1	GCTGAAGTCGGCTAGGTTTG	TGTAGCAAGTGTGACACGCA	134	12	75874313
GNAI1	GCCACTGTACCCAGAGATTCA	CACGACCAGGCTCTGCTT	87	7	79763904
GNG5	GCAGAGCTTGTGGGAGTTTC	GCACAGTAGCTCCTACGCCT	130	1	84972048
GRAP2	AGTTCCTGTGTGGTTTGCCT	CCACTTTGTGCTGTCTTCCA	149	22	40296886
HIPK2	TGTCCGCGGTTCATGGCAAC	AGGCTCAAGATGGCAGATTCCGA	87	7	139477317
IKBKB	TCCGTGACGTCAGAGCAGGAA	AAATGCGGGAATCTTAACGCGGG	70	8	42128629
INPP5D	CTGGAACCATGGCAACATC	GGGAGGGAGCAGGTTGTACT	146	2	233924989
ITGB3	AGCAATAGTTTCCCACCGC	CCTCCGGCTTCTCTAGATCC	74	17	45331008
MAN1A1	TGAGACTGCTGCCTCTTTCCTA	AGTGTACCGTGCTGCTCTACT	131	6	119670726
MDM2	CTGTGTGTCGGAAAGATGGA	GCGCTCGTACGCACTAATC	124	12	69201756
MECOM	AGAGTGATCTGATCGGAAGCCA	TATAGGACTTGGGTCTTGGCGA	111	3	168865122
MTCP1	CGAGGTCACTCGAACCGAAGTTAT	AATCTCGTACAACTCTCCCTGGAC	74	X	154299101
MYC	CAGGGCTTCTCAGAGGCTT	TGCCTCTCGCTGGAATTACT	114	8	128748130

Gene Symbol	Forward Primer	Reverse Primer	Amplicon size (bp)	Chromosome	Co-ordinate start
NOG	GACCCTGCGAGACAGACAA	CTGGCTGAGCTGGTTCTAGG	137	17	54670860
PCGF6	ATGGTCGGGAGAGACACCA	GGCCGCTAAACCACATTTC	127	10	105110691
PDK2	ATTGGTTCGCCCGGAGTTGTT	CATTCCAACGCCCAGCCAAT	83	17	48172439
PDPK1	AGCAAAGTTGCGCCTCTGAGTT	TCATTCCGGCATCCCAGACC	150	16	2587665
PIK3C2B	GGCACGAAACTCCAAAGTGT	GCCATCACTTCCCTGAGAAA	137	1	204459352
PIK3CG	AGACGCACACTGGGTAGGTT	AGGAAGTGCCACTATTCCGA	150	7	106505541
PRKCA	GGACCATGGCTGACGTTT	TTTGTGGTCCTTCACCTCG	119	17	64298726
PTEN	CAGTCAAATCTCTGCGAACG	GTGTAGAGGGAAATGCAGGG	145	10	89622670
PTPN11	ATGCTTTGGACACTGTGCGT	CACTTCCTGCTTCCCGTCAG	94	12	112856518
RAB10	TGAGCTCGTCGACTTAGGG	TGAGGACAGTCTCTCTCCCG	85	2	26256779
RAF1	CGACTTCTAACTTGGCTCGG	AGTGCCTCTCCTGAAAGCAA	141	3	12705525
RB1	AGTCACCCACCAGACTCTTTGT	TTAAACTGGGAAACCTGGCGTG	139	13	48877611
RBBP7	TTTCCCAAGCGCGTCACACT	AGAGAGAGAGAAGAGCGCCTCA	142	X	16888237
RBL2	GTTGAATGTTCTCACGGTGG	GGCTGCTGGGAAACGTAGT	150	16	53468161
RHEBL1	AAACGAGGTCAGGGTGTGAG	GCGTTAGACGTGCTGTGTTG	145	12	49463608
SERPINE1	AGAGCGCTGTCAAGAAGACC	AGGTTGGAGGGAGTTTGCTT	145	7	100770170
SHC4	CCAACTCTTAGGCGCAGAAC	TCGAAAGTTCCTGTGTGCAG	142	15	49255441
SMAD1	TATTGGCAGCTGAGGAGTGGA	TGGAGCCTGGATTGATCTGGT	76	4	146402644
SMAD2	TGTTCCTTCCTCTTCCGATG	TTTCAGTTCCGCCTCCAA	77	18	45456730
SMAD3	GGATCTGCGCATCAAAGCTA	CCAAGTGGCAGCAGAAGTTT	121	15	67357983
SMAD4	AGGTCCAGATTCAGAGCCG	GGGTAATTTCAGGGTTTGGC	147	18	48556383
SMAD6	CTGCAATGGGACGTGTTCT	CCTCCCATCATATGCCAGTC	110	15	66994470
SOS1	TAGCGCTGGAGCTTCCTACT	ctcGAGCTTTCCTCCTCCTT	118	2	39347992
SUZ12	TCTCCTCCTCCCTTCCCTT	AGCAAGCGCCTGCAATAC	103	17	30263856
TGFB1	TTTCCTCTTCTCCCGACCAGC	CAGAGTCTGAGACGAGCCGC	95	19	41859516
TGFBR2	GGAACTCCTGAGTGGTGTGG	GCTCTCTCGTAGCTGCCAAT	104	3	30647794
TGFBRAP1	CCCAGTTTCTGTCACCTACACAGGA	ATTTCCCAGGGCCCAAGCATTT	93	2	105946291
TGIF1	TTCCTTCGGCTGCGTTTCTGT	TAGGGCGTTCTACCAACAACGACT	133	18	3451475
TP53	TGGACGGTGGCTCTAGACTT	CCCTTACTTGTCATGGCGAC	139	17	7590656
VCAN	TCTTGTCAGGAAGAAACGCC	GCAGCTTCAGCCACTCCTAA	138	5	82767084
YWHAG	AGAGGACCGACCCACAGAGC	TTGTCCGTCAGTGGCACGCA	122	7	75988148
CYP4A22 (1)†	AGCCTTCCACAATGACATCC	CATGGACTCACCAGCATCAC	78	1	47380428
CYP4A22 (2)†	AGCTGTTCTCCCACTGTTCC	CCCTGTCCACCTGTGCTTA	144	1	47602922
TAL1‡	TGTTGGGAAAGGCGAATAGT	TGGAATTGGGTTCTGGTTTC	130	1	47695443
SIL‡	GCTCCTACCCTGCAAACAGAC	GGAAACCAGGAGCACAAAGC	71	1	47552266
HSSILM10A*	TCTCTTTGAACACAGGGCAATG	TATTAGTCTAGGTGTACTGGCAGTTG	71	1	47560374
HSSILM51B*	TGAATGCTTCCCTTGTGATG	GTAATGTTTCCTTACTGGTTAGCAAC	71	1	47527705

† H3K4me3 ChIP negative controls.

‡ H3K27me3 ChIP negative controls.

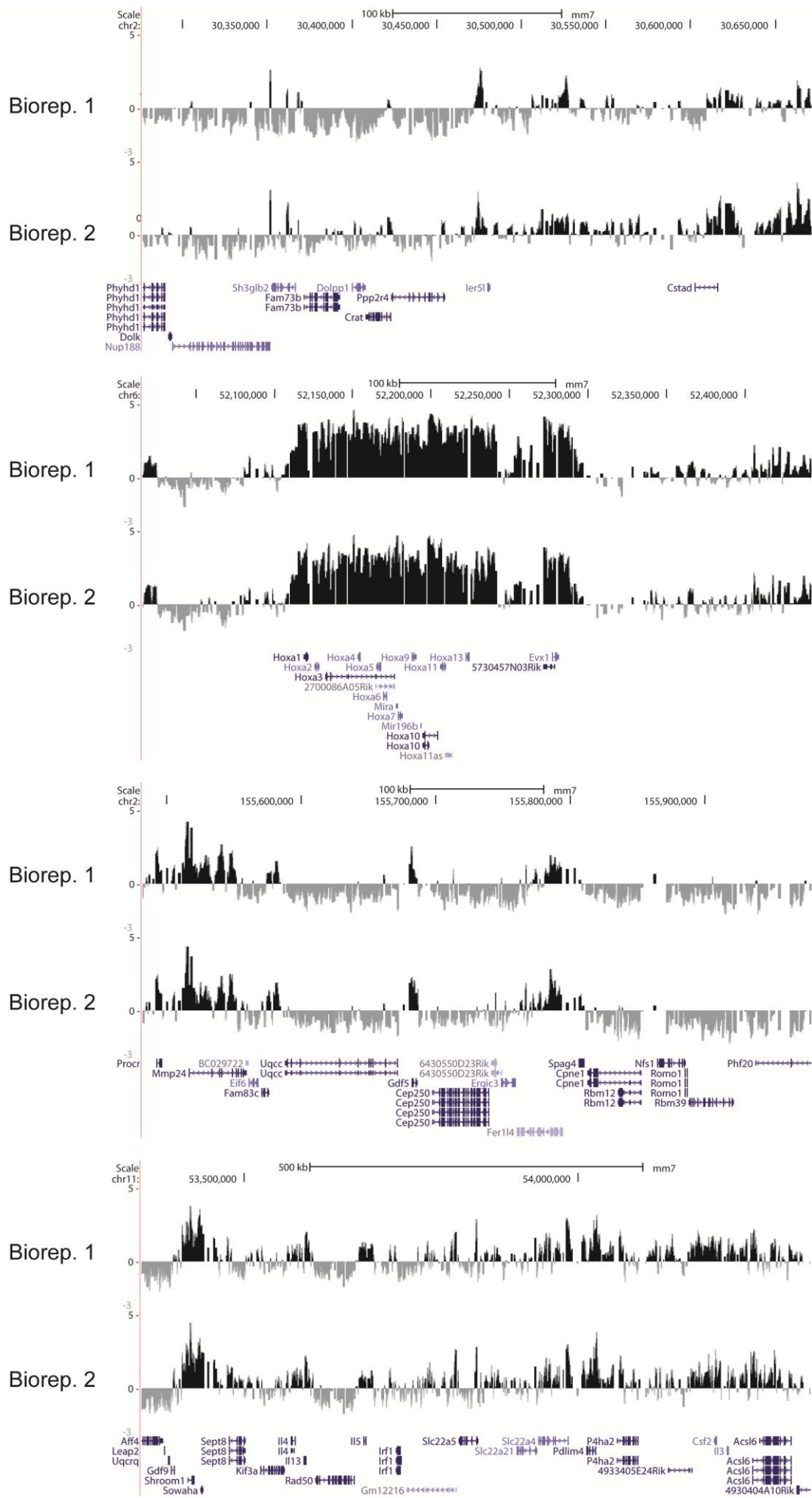* H3K4me3/H3K27me3 ChIP negative controls.

Appendix 5

A. Conventional H3K27me3 ChIP-chip bioreplicates
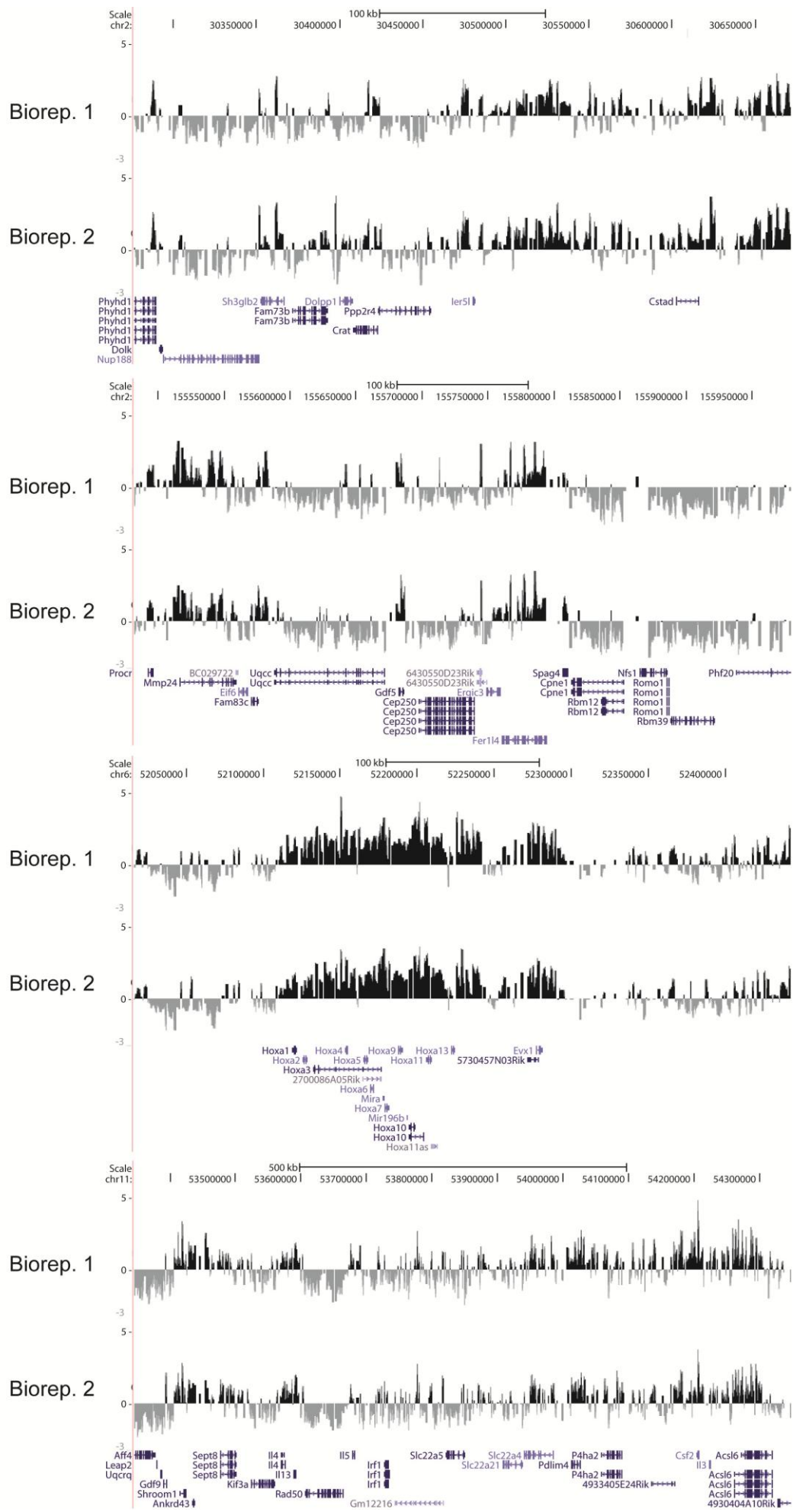
B. Low-cell H3K27me3 ChIP-chip bioreplicates

C. Conventional H3K4me3 ChIP-chip bioreplicates

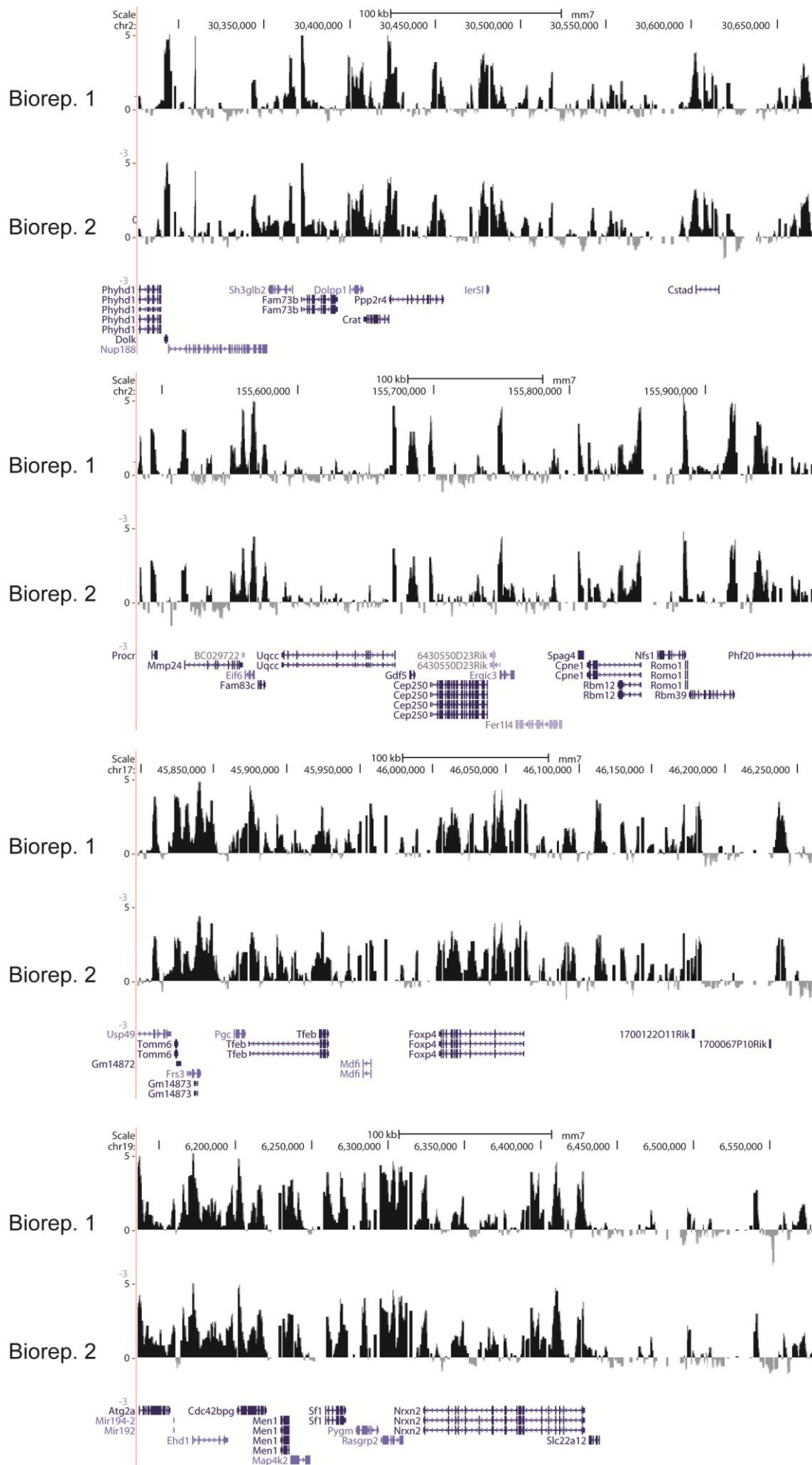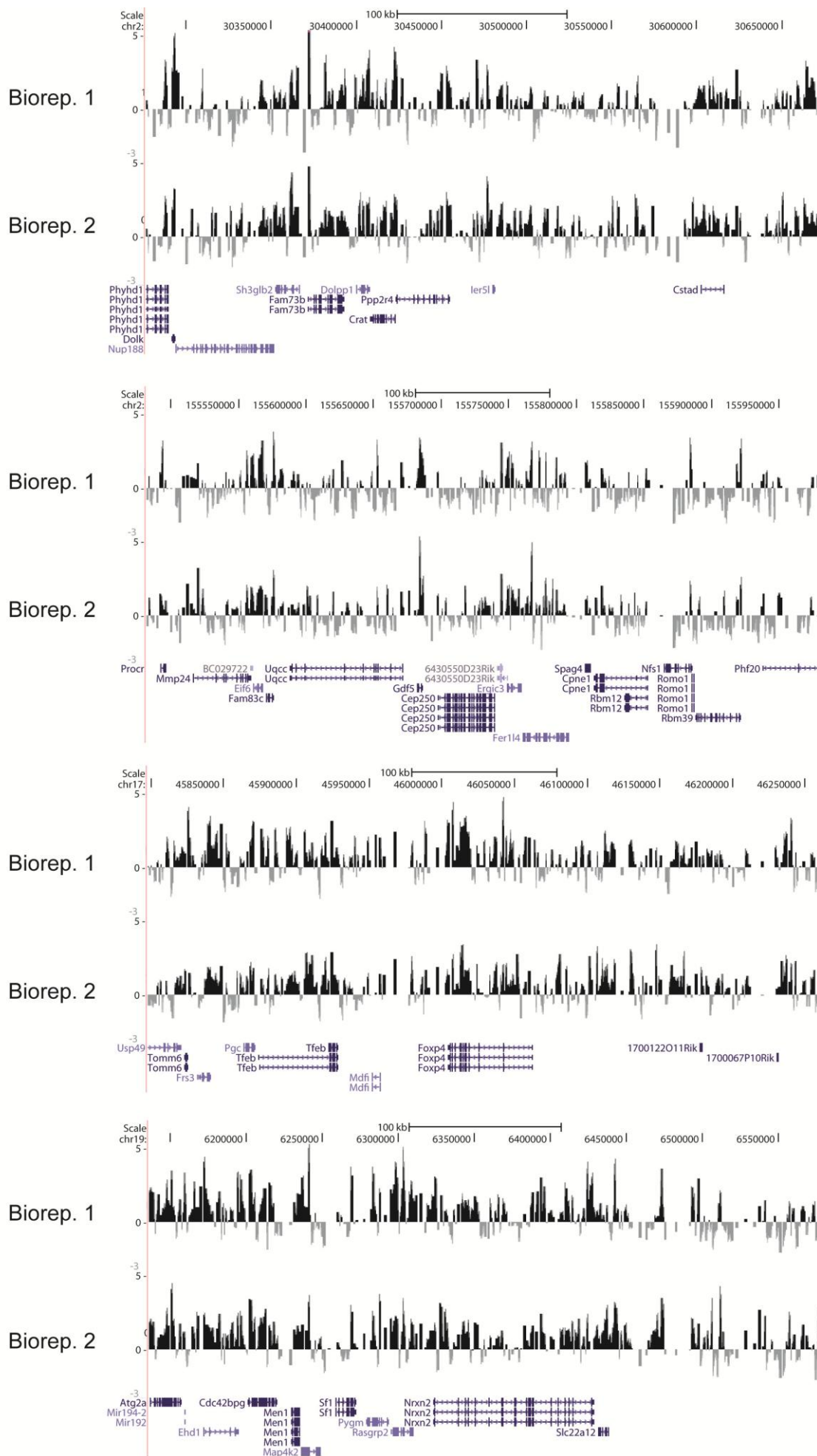D. Low-cell H3K4me3 ChIP-chip bioreplicates

D

Bibliography

ACEVEDO, L. G., INIGUEZ, A. L., HOLSTER, H. L., ZHANG, X., GREEN, R. & FARNHAM, P. J. 2007. Genome-scale ChIP-chip analysis using 10,000 human cells. Biotechniques, 43, 791-7.

ADAMO, A., SESE, B., BOUE, S., CASTANO, J., PARAMONOV, I., BARRERO, M. J. & IZPISUA BELMONTE, J. C. 2011. LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. Nat Cell Biol, 13, 652-9.

ADLI, M., ZHU, J. & BERNSTEIN, B. E. 2010. Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. Nat Methods, 7, 615-8.

ADOLFSSON, J., BORGE, O. J., BRYDER, D., THEILGAARD-MONCH, K., ASTRAND-GRUNDSTROM, I., SITNICKA, E., SASAKI, Y. & JACOBSEN, S. E. 2001. Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity. Immunity, 15, 659-69.

AGGER, K., CLOOS, P. A., CHRISTENSEN, J., PASINI, D., ROSE, S., RAPPSILBER, J., ISSAEVA, I., CANAANI, E., SALCINI, A. E. & HELIN, K. 2007. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature, 449, 731-4.

AIUTI, A., WEBB, I. J., BLEUL, C., SPRINGER, T. & GUTIERREZ-RAMOS, J. C. 1997. The chemokine SDF-1 is a chemoattractant for human CD34+ hematopoietic progenitor cells and provides a new mechanism to explain the mobilization of CD34+ progenitors to peripheral blood. J Exp Med, 185, 111-20.

AKALIN, A., FREDMAN, D., ARNER, E., DONG, X., BRYNE, J. C., SUZUKI, H., DAUB, C. O., HAYASHIZAKI, Y. & LENHARD, B. 2009. Transcriptional features of genomic regulatory blocks. Genome Biol, 10, R38.

AKASHI, K. 2005. Lineage promiscuity and plasticity in hematopoietic development. Ann N Y Acad Sci, 1044, 125-31.

AKASHI, K., TRAVER, D., MIYAMOTO, T. & WEISSMAN, I. L. 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature, 404, 193-7.

ALLFREY, V. G., FAULKNER, R. & MIRSKY, A. E. 1964. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. Proc Natl Acad Sci U S A, 51, 786-94.

ALLISON, D. B., CUI, X., PAGE, G. P. & SABRIPOUR, M. 2006. Microarray data analysis: from disarray to consolidation and consensus. Nat Rev Genet, 7, 55-65.

ALMER, A., RUDOLPH, H., HINNEN, A. & HORZ, W. 1986. Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. EMBO J, 5, 2689-96.

ALON, U. 2007. An introduction to systems biology : design principles of biological circuits, Boca Raton, FL, Chapman & Hall/CRC.

AL-ZI'ABI, M. O., BOWOLAKSONO, A. & OKUDA, K. 2009. Survival role of locally produced acetylcholine in the bovine corpus luteum. Biol Reprod, 80, 823-32.

ANG, Y. S., TSAI, S. Y., LEE, D. F., MONK, J., SU, J., RATNAKUMAR, K., DING, J., GE, Y., DARR, H., CHANG, B., WANG, J., RENDL, M., BERNSTEIN, E., SCHANIEL, C. & LEMISCHKA, I. R. 2011. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell, 145, 183-97.

ARMSTRONG, S. A., STAUNTON, J. E., SILVERMAN, L. B., PIETERS, R., DEN BOER, M. L., MINDEN, M. D., SALLAN, S. E., LANDER, E. S., GOLUB, T. R. & KORSMEYER, S. J. 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet, 30, 41-7.

ATTAR, E. C. & SCADDEN, D. T. 2004. Regulation of hematopoietic stem cell growth. Leukemia, 18, 1760-8.

ATTEMA, J. L., PAPATHANASIOU, P., FORSBERG, E. C., XU, J., SMALE, S. T. & WEISSMAN, I. L. 2007. Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. Proc Natl Acad Sci U S A, 104, 12371-6.

AUERBACH, R. K., EUSKIRCHEN, G., ROZOWSKY, J., LAMARRE-VINCENT, N., MOQTADERI, Z., LEFRANCOIS, P., STRUHL, K., GERSTEIN, M. & SNYDER, M. 2009. Mapping accessible chromatin regions using Sono-Seq. Proc Natl Acad Sci U S A, 106, 14926-31.

AZAM, M., SEELIGER, M. A., GRAY, N. S., KURIYAN, J. & DALEY, G. Q. 2008. Activation of tyrosine kinases by mutation of the gatekeeper threonine. Nat Struct Mol Biol, 15, 1109-18.

AZUARA, V., PERRY, P., SAUER, S., SPIVAKOV, M., JORGENSEN, H. F., JOHN, R. M., GOUTI, M., CASANOVA, M., WARNES, G., MERKENSCHLAGER, M. & FISHER, A. G. 2006. Chromatin signatures of pluripotent cell lines. Nat Cell Biol, 8, 532-8.

BANERJI, J., OLSON, L. & SCHAFFNER, W. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell, 33, 729-40.

BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell, 27, 299-308.

BANNISTER, A. J., SCHNEIDER, R., MYERS, F. A., THORNE, A. W., CRANE-ROBINSON, C. & KOUZARIDES, T. 2005. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. J Biol Chem, 280, 17732-6.

BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. 2007. High-resolution profiling of histone methylations in the human genome. Cell, 129, 823-37.

BARTRAM, C. R., DE KLEIN, A., HAGEMEIJER, A., VAN AGTHOVEN, T., GEURTS VAN KESSEL, A., BOOTSMA, D., GROSVELD, G., FERGUSON-SMITH, M. A., DAVIES, T., STONE, M. & ET AL. 1983. Translocation of c-ab1 oncogene correlates with the presence of a Philadelphia chromosome in chronic myelocytic leukaemia. Nature, 306, 277-80.

BATARD, P., MONIER, M. N., FORTUNEL, N., DUCOS, K., SANSILVESTRI-MOREL, P., PHAN, T., HATZFELD, A. & HATZFELD, J. A. 2000. TGF-(beta)1 maintains hematopoietic immaturity by a reversible negative control of cell cycle and induces CD34 antigen up-modulation. J Cell Sci, 113 (Pt 3), 383-90.

BAZZONI, G., CARLESSO, N., GRIFFIN, J. D. & HEMLER, M. E. 1996. Bcr/Abl expression stimulates integrin function in hematopoietic cell lines. J Clin Invest, 98, 521-8.

BECKER, A. J., MC, C. E. & TILL, J. E. 1963. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature, 197, 452-4.

BECKER, P. B. 1999. Chromatin protocols, Totowa, N.J., Humana Press.

BEDI, A., ZEHNBAUER, B. A., BARBER, J. P., SHARKIS, S. J. & JONES, R. J. 1994. Inhibition of apoptosis by BCR-ABL in chronic myeloid leukemia. Blood, 83, 2038-44.

BELL, A. C. & FELSENFELD, G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature, 405, 482-5.

BELL, A. C., WEST, A. G. & FELSENFELD, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell, 98, 387-96.

BENJAMINI, Y. & HOCHBERG, Y. 1995. CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. Journal of the Royal Statistical Society Series B-Methodological, 57, 289-300.

BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., KEIRA CHEETHAM, R., COX, A. J., ELLIS, D. J., FLATBUSH, M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature, 456, 53-9.

BERGER, U., MAYWALD, O., PFIRRMANN, M., LAHAYE, T., HOCHHAUS, A., REITER, A., HASFORD, J., HEIMPEL, H., HOSSFELD, D. K., KOLB, H. J., LOFFLER, H., PRALLE, H., QUEISSER, W. & HEHLMANN, R. 2005. Gender aspects in chronic myeloid leukemia: long-term results from randomized studies. Leukemia, 19, 984-9.

BERNSTEIN, B. E., HUMPHREY, E. L., ERLICH, R. L., SCHNEIDER, R., BOUMAN, P., LIU, J. S., KOUZARIDES, T. & SCHREIBER, S. L. 2002. Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A, 99, 8695-700.

BERNSTEIN, B. E., LIU, C. L., HUMPHREY, E. L., PERLSTEIN, E. O. & SCHREIBER, S. L. 2004. Global nucleosome occupancy in yeast. Genome Biol, 5, R62.

BERNSTEIN, B. E., MIKKELSEN, T. S., XIE, X., KAMAL, M., HUEBERT, D. J., CUFF, J., FRY, B., MEISSNER, A., WERNIG, M., PLATH, K., JAENISCH, R., WAGSCHAL, A., FEIL, R., SCHREIBER, S. L. & LANDER, E. S. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell, 125, 315-26.

BERNSTEIN, I. D., ANDREWS, R. G. & ZSEBO, K. M. 1991. Recombinant human stem cell factor enhances the formation of colonies by CD34+ and CD34+lin- cells, and the generation of colony-forming cell progeny from CD34+lin- cells cultured with interleukin-3, granulocyte colony-stimulating factor, or granulocyte-macrophage colony-stimulating factor. Blood, 77, 2316-21.

BERTONE, P., STOLC, V., ROYCE, T. E., ROZOWSKY, J. S., URBAN, A. E., ZHU, X., RINN, J. L., TONGPRASIT, W., SAMANTA, M., WEISSMAN, S., GERSTEIN, M. & SNYDER, M. 2004. Global identification of human transcribed sequences with genome tiling arrays. Science, 306, 2242-6.

BERTONE, P., STOLC, V., ROYCE, T. E., ROZOWSKY, J. S., URBAN, A. E., ZHU, X., RINN, J. L., TONGPRASIT, W., SAMANTA, M., WEISSMAN, S., GERSTEIN, M. & SNYDER, M. 2004. Global identification of human transcribed sequences with genome tiling arrays. Science, 306, 2242-6.

BHARDWAJ, G., MURDOCH, B., WU, D., BAKER, D. P., WILLIAMS, K. P., CHADWICK, K., LING, L. E., KARANU, F. N. & BHATIA, M. 2001. Sonic hedgehog induces the proliferation of primitive human hematopoietic cells via BMP regulation. Nat Immunol, 2, 172-80.

BHATIA, M., BONNET, D., WU, D., MURDOCH, B., WRANA, J., GALLACHER, L. & DICK, J. E. 1999. Bone morphogenetic proteins regulate the developmental program of human hematopoietic stem cells. J Exp Med, 189, 1139-48.

BHATIA, M., WANG, J. C., KAPP, U., BONNET, D. & DICK, J. E. 1997. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. Proc Natl Acad Sci U S A, 94, 5320-5.

BIENZ, M. 1998. TCF: transcriptional activator or repressor? Curr Opin Cell Biol, 10, 366-72.

BIRNEY, E., STAMATOYANNOPOULOS, J. A., DUTTA, A., GUIGO, R., GINGERAS, T. R., MARGULIES, E. H., WENG, Z., SNYDER, M., DERMITZAKIS, E. T., THURMAN, R. E., KUEHN, M. S., TAYLOR, C. M., NEPH, S., KOCH, C. M., ASTHANA, S., MALHOTRA, A., et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 447, 799-816.

BLACKWOOD, E. M. & KADONAGA, J. T. 1998. Going the distance: a current view of enhancer action. Science, 281, 60-3.

BLANK, U., KARLSSON, G. & KARLSSON, S. 2008. Signaling pathways governing stem-cell fate. Blood, 111, 492-503.

BLANK, U., KARLSSON, G., MOODY, J. L., UTSUGISAWA, T., MAGNUSSON, M., SINGBRANT, S., LARSSON, J. & KARLSSON, S. 2006. Smad7 promotes self-renewal of hematopoietic stem cells. Blood, 108, 4246-54.

BLOOM, J., AMADOR, V., BARTOLINI, F., DEMARTINO, G. & PAGANO, M. 2003. Proteasome-mediated degradation of p21 via N-terminal ubiquitinylation. Cell, 115, 71-82.

BOISSET, J. C., VAN CAPPELLEN, W., ANDRIEU-SOLER, C., GALJART, N., DZIERZAK, E. & ROBIN, C. 2010. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. Nature, 464, 116-20.

BOSMA, G. C., CUSTER, R. P. & BOSMA, M. J. 1983. A severe combined immunodeficiency mutation in the mouse. Nature, 301, 527-30.

BOYER, L. A., PLATH, K., ZEITLINGER, J., BRAMBRINK, T., MEDEIROS, L. A., LEE, T. I., LEVINE, S. S., WERNIG, M., TAJONAR, A., RAY, M. K., BELL, G. W., OTTE, A. P., VIDAL, M., GIFFORD, D. K., YOUNG, R. A. & JAENISCH, R. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature, 441, 349-53.

BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008. High-resolution mapping and characterization of open chromatin across the genome. Cell, 132, 311-22.

BRACKEN, A. P., DIETRICH, N., PASINI, D., HANSEN, K. H. & HELIN, K. 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev, 20, 1123-36.

BRACKEN, A. P., PASINI, D., CAPRA, M., PROSPERINI, E., COLLI, E. & HELIN, K. 2003. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J, 22, 5323-35.

BROSKE, A. M., VOCKENTANZ, L., KHARAZI, S., HUSKA, M. R., MANCINI, E., SCHELLER, M., KUHL, C., ENNS, A., PRINZ, M., JAENISCH, R., NERLOV, C., LEUTZ, A., ANDRADE-NAVARRO, M. A., JACOBSEN, S. E. & ROSENBAUER, F. 2009. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet, 41, 1207-15.

BRUNS, I., CZIBERE, A., FISCHER, J. C., ROELS, F., CADEDDU, R. P., BUEST, S., BRUENNERT, D., HUENERLITUERKOGLU, A. N., STOECKLEIN, N. H., SINGH, R., ZERBINI, L. F., JAGER, M., KOBBE, G., GATTERMANN, N., KRONENWETT, R., BRORS, B. & HAAS, R. 2009. The hematopoietic stem cell in chronic phase CML is characterized by a transcriptional profile resembling normal myeloid progenitor cells and reflecting loss of quiescence. Leukemia, 23, 892-9.

BUCK, M. J. & LIEB, J. D. 2004. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics, 83, 349-60.

BUCK, M. J., NOBEL, A. B. & LIEB, J. D. 2005. ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data. Genome Biol, 6, R97.

BURGOLD, T., SPREAFICO, F., DE SANTA, F., TOTARO, M. G., PROSPERINI, E., NATOLI, G. & TESTA, G. 2008. The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. PLoS One, 3, e3034.

BUTLER, J. M., NOLAN, D. J., VERTES, E. L., VARNUM-FINNEY, B., KOBAYASHI, H., HOOPER, A. T., SEANDEL, M., SHIDO, K., WHITE, I. A., KOBAYASHI, M., WITTE, L., MAY, C., SHAWBER, C., KIMURA, Y., KITAJEWSKI, J., ROSENWAKS, Z., BERNSTEIN, I. D. & RAFII, S. 2010. Endothelial cells are essential for the self-renewal and repopulation of Notch-dependent hematopoietic stem cells. Cell Stem Cell, 6, 251-64.

CALVI, L. M., ADAMS, G. B., WEIBRECHT, K. W., WEBER, J. M., OLSON, D. P., KNIGHT, M. C., MARTIN, R. P., SCHIPANI, E., DIVIETI, P., BRINGHURST, F. R., MILNER, L. A., KRONENBERG, H. M. & SCADDEN, D. T. 2003. Osteoblastic cells regulate the haematopoietic stem cell niche. Nature, 425, 841-6.

CAO, R., TSUKADA, Y. & ZHANG, Y. 2005. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol Cell, 20, 845-54.

CAO, R. & ZHANG, Y. 2004. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell, 15, 57-67.

CARETTI, G., DI PADOVA, M., MICALES, B., LYONS, G. E. & SARTORELLI, V. 2004. The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes Dev, 18, 2627-38.

CARNINCI, P., SANDELIN, A., LENHARD, B., KATAYAMA, S., SHIMOKAWA, K., PONJAVIC, J., SEMPLE, C. A., TAYLOR, M. S., ENGSTROM, P. G., FRITH, M. C., FORREST, A. R., ALKEMA, W. B., TAN, S. L., PLESSY, C., KODZIUS, R., RAVASI, T., KASUKAWA, T., et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet, 38, 626-35.

CASHMAN, J., CLARK-LEWIS, I., EAVES, A. & EAVES, C. 2002. Stromal-derived factor 1 inhibits the cycling of very primitive human hematopoietic cells in vitro and in NOD/SCID mice. Blood, 99, 792-9.

CASHMAN, J. D., EAVES, A. C. & EAVES, C. J. 1992. Granulocyte-macrophage colony-stimulating factor modulation of the inhibitory effect of transforming growth factor-beta on normal and leukemic human hematopoietic progenitor cells. Leukemia, 6, 886-92.

CATLIN, S. N., BUSQUE, L., GALE, R. E., GUTTORP, P. & ABKOWITZ, J. L. 2011. The replication rate of human hematopoietic stem cells in vivo. Blood, 117, 4460-6.

CHADWICK, K., SHOJAEI, F., GALLACHER, L. & BHATIA, M. 2005. Smad7 alters cell fate decisions of human hematopoietic repopulating cells. Blood, 105, 1905-15.

CHAI, B., HUANG, J., CAIRNS, B. R. & LAURENT, B. C. 2005. Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair. Genes Dev, 19, 1656-61.

CHALLEN, G. A., SUN, D., JEONG, M., LUO, M., JELINEK, J., BERG, J. S., BOCK, C., VASANTHAKUMAR, A., GU, H., XI, Y., LIANG, S., LU, Y., DARLINGTON, G. J., MEISSNER, A., ISSA, J. P., GODLEY, L. A., LI, W. & GOODELL, M. A. 2012. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet, 44, 23-31.

CHAMBERLAIN, S. J., YEE, D. & MAGNUSON, T. 2008. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem Cells, 26, 1496-505.

CHAN, W. W., WISE, S. C., KAUFMAN, M. D., AHN, Y. M., ENSINGER, C. L., HAACK, T., HOOD, M. M., JONES, J., LORD, J. W., LU, W. P., MILLER, D., PATT, W. C., SMITH, B. D., PETILLO, P. A., RUTKOSKI, T. J., TELIKEPALLI, H., VOGETI, L., YAO, T., CHUN, L., CLARK, R., EVANGELISTA, P., GAVRILESCU, L. C., LAZARIDES, K., ZALESKAS, V. M., STEWART, L. J., VAN ETTEN, R. A. & FLYNN, D. L. 2011. Conformational control inhibition of the BCR-ABL1 tyrosine kinase, including the gatekeeper T315I mutant, by the switch-control inhibitor DCC-2036. Cancer Cell, 19, 556-68.

CHEE, M., YANG, R., HUBBELL, E., BERNO, A., HUANG, X. C., STERN, D., WINKLER, J., LOCKHART, D. J., MORRIS, M. S. & FODOR, S. P. 1996. Accessing genetic information with high-density DNA arrays. Science, 274, 610-4.

CHEN, L. & WIDOM, J. 2005. Mechanism of transcriptional silencing in yeast. Cell, 120, 37-48.

CHENG, J., KAPRANOV, P., DRENKOW, J., DIKE, S., BRUBAKER, S., PATEL, S., LONG, J., STERN, D., TAMMANA, H., HELT, G., SEMENTCHENKO, V., PICCOLBONI, A., BEKIRANOV, S., BAILEY, D. K., GANESH, M., GHOSH, S., BELL, I., GERHARD, D. S. & GINGERAS, T. R. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science, 308, 1149-54.

CHENG, T., RODRIGUES, N., SHEN, H., YANG, Y., DOMBKOWSKI, D., SYKES, M. & SCADDEN, D. T. 2000. Hematopoietic stem cell quiescence maintained by p21cip1/waf1. Science, 287, 1804-8.

CHESHIER, S. H., MORRISON, S. J., LIAO, X. & WEISSMAN, I. L. 1999. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. Proc Natl Acad Sci U S A, 96, 3120-5.

CHEUNG, P., TANNER, K. G., CHEUNG, W. L., SASSONE-CORSI, P., DENU, J. M. & ALLIS, C. D. 2000. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell, 5, 905-15.

CHOI, K., KENNEDY, M., KAZAROV, A., PAPADIMITRIOU, J. C. & KELLER, G. 1998. A common precursor for hematopoietic and endothelial cells. Development, 125, 725-32.

CHOMEL, J. C., BONNET, M. L., SOREL, N., BERTRAND, A., MEUNIER, M. C., FICHELSON, S., MELKUS, M., BENNACEUR-GRISCELLI, A., GUILHOT, F. & TURHAN, A. G. 2011. Leukemic stem cell persistence in chronic myeloid leukemia patients with sustained undetectable molecular residual disease. Blood, 118, 3657-60.

CHONG, J. L., WENZEL, P. L., SAENZ-ROBLES, M. T., NAIR, V., FERREY, A., HAGAN, J. P., GOMEZ, Y. M., SHARMA, N., CHEN, H. Z., OUSEPH, M., WANG, S. H., TRIKHA, P., CULP, B., MEZACHE, L., WINTON, D. J., SANSOM, O. J., CHEN, D., BREMNER, R., CANTALUPO, P. G., ROBINSON, M. L., PIPAS, J. M. & LEONE, G. 2009. E2f1-3 switch from activators in progenitor cells to repressors in differentiating cells. Nature, 462, 930-4.

CHRISTENSEN, J., AGGER, K., CLOOS, P. A., PASINI, D., ROSE, S., SENNELS, L., RAPPSILBER, J., HANSEN, K. H., SALCINI, A. E. & HELIN, K. 2007. RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. Cell, 128, 1063-76.

CHUNG, C. W. & WITHERINGTON, J. 2011. Progress in the discovery of small-molecule inhibitors of bromodomain--histone interactions. J Biomol Screen, 16, 1170-85.

CIVIN, C. I., STRAUSS, L. C., BROVALL, C., FACKLER, M. J., SCHWARTZ, J. F. & SHAPER, J. H. 1984. Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. J Immunol, 133, 157-65.

CLARK, S. J., HARRISON, J., PAUL, C. L. & FROMMER, M. 1994. High sensitivity mapping of methylated cytosines. Nucleic Acids Res, 22, 2990-7.

CLEVERS, H. 2006. Wnt/beta-catenin signaling in development and disease. Cell, 127, 469-80.

COBAS, M., WILSON, A., ERNST, B., MANCINI, S. J., MACDONALD, H. R., KEMLER, R. & RADTKE, F. 2004. Beta-catenin is dispensable for hematopoiesis and lymphopoiesis. J Exp Med, 199, 221-9.

COLLAS, P. 2011. A chromatin immunoprecipition protocol for small cell numbers. Methods Mol Biol, 791, 179-93.

CORBIN, A. S., AGARWAL, A., LORIAUX, M., CORTES, J., DEININGER, M. W. & DRUKER, B. J. 2011. Human chronic myeloid leukemia stem cells are insensitive to imatinib despite inhibition of BCR-ABL activity. J Clin Invest, 121, 396-409.

CORONA, D. F. & TAMKUN, J. W. 2004. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. Biochim Biophys Acta, 1677, 113-9.

CORTELLINO, S., XU, J., SANNAI, M., MOORE, R., CARETTI, E., CIGLIANO, A., LE COZ, M., DEVARAJAN, K., WESSELS, A., SOPRANO, D., ABRAMOWITZ, L. K., BARTOLOMEI, M. S., RAMBOW, F., BASSI, M. R., BRUNO, T., FANCIULLI, M., RENNER, C., KLEIN-SZANTO, A. J., MATSUMOTO, Y., KOBI, D., DAVIDSON, I., ALBERTI, C., LARUE, L. & BELLACOSA, A. 2011. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell, 146, 67-79.

COTE, J., QUINN, J., WORKMAN, J. L. & PETERSON, C. L. 1994. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science, 265, 53-60.

CRAWFORD, G. E., DAVIS, S., SCACHERI, P. C., RENAUD, G., HALAWI, M. J., ERDOS, M. R., GREEN, R., MELTZER, P. S., WOLFSBERG, T. G. & COLLINS, F. S. 2006. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat Methods, 3, 503-9.

CRAWFORD, G. E., HOLT, I. E., WHITTLE, J., WEBB, B. D., TAI, D., DAVIS, S., MARGULIES, E. H., CHEN, Y., BERNAT, J. A., GINSBURG, D., ZHOU, D., LUO, S., VASICEK, T. J., DALY, M. J., WOLFSBERG, T. G. & COLLINS, F. S. 2006. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res, 16, 123-31.

CUI, K., ZANG, C., ROH, T. Y., SCHONES, D. E., CHILDS, R. W., PENG, W. & ZHAO, K. 2009. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell, 4, 80-93.

CUTHBERT, G. L., DAUJAT, S., SNOWDEN, A. W., ERDJUMENT-BROMAGE, H., HAGIWARA, T., YAMADA, M., SCHNEIDER, R., GREGORY, P. D., TEMPST, P., BANNISTER, A. J. & KOUZARIDES, T. 2004. Histone deimination antagonizes arginine methylation. Cell, 118, 545-53.

DAAKA, Y. 2004. G proteins in cancer: the prostate cancer paradigm. Sci STKE, 2004, re2.

DAHL, J. A. & COLLAS, P. 2007. Q2ChIP, a quick and quantitative chromatin immunoprecipitation assay, unravels epigenetic dynamics of developmentally regulated genes in human carcinoma cells. Stem Cells, 25, 1037-46.

DAHL, J. A. & COLLAS, P. 2008. A rapid micro chromatin immunoprecipitation assay (microChIP). Nat Protoc, 3, 1032-45.

DALEY, G. Q., VAN ETTEN, R. A. & BALTIMORE, D. 1990. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science, 247, 824-30.

DALMA-WEISZHAUSZ, D. D., WARRINGTON, J., TANIMOTO, E. Y. & MIYADA, C. G. 2006. The affymetrix GeneChip platform: an overview. Methods Enzymol, 410, 3-28.

DAMESHEK, W. 1951. Some speculations on the myeloproliferative syndromes. Blood, 6, 372-5.

DAO, M. A., TAYLOR, N. & NOLTA, J. A. 1998. Reduction in levels of the cyclin-dependent kinase inhibitor p27(kip-1) coupled with transforming growth factor beta neutralization induces cell-cycle entry and increases retroviral transduction of primitive human hematopoietic cells. Proc Natl Acad Sci U S A, 95, 13006-11.

DAS, P. M., RAMACHANDRAN, K., VANWERT, J. & SINGAL, R. 2004. Chromatin immunoprecipitation assay. Biotechniques, 37, 961-9.

DAWSON, M. A. & KOUZARIDES, T. 2012. Cancer epigenetics: from mechanism to therapy. Cell, 150, 12-27.

DAY, D. S., LUQUETTE, L. J., PARK, P. J. & KHARCHENKO, P. V. 2010. Estimating enrichment of repetitive elements from high-throughput sequence data. Genome Biol, 11, R69.

DAZZI, F., HASSERJIAN, R., GORDON, M. Y., BOECKLIN, F., COTTER, F., CORBO, M., CAPELLI, D. & GOLDMAN, J. M. 2000. Normal and chronic phase CML hematopoietic cells repopulate NOD/SCID bone marrow with different kinetics and cell lineage representation. Hematol J, 1, 307-15.

DE BRUIJN, M. F., MA, X., ROBIN, C., OTTERSBACH, K., SANCHEZ, M. J. & DZIERZAK, E. 2002. Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta. Immunity, 16, 673-83.

DE LA SERNA, I. L., OHKAWA, Y., BERKES, C. A., BERGSTROM, D. A., DACWAG, C. S., TAPSCOTT, S. J. & IMBALZANO, A. N. 2005. MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. Mol Cell Biol, 25, 3997-4009.

DEDON, P. C., SOULTS, J. A., ALLIS, C. D. & GOROVSKY, M. A. 1991. A simplified formaldehyde fixation and immunoprecipitation technique for studying protein-DNA interactions. Anal Biochem, 197, 83-90.

DHAMI, P., BRUCE, A. W., JIM, J. H., DILLON, S. C., HALL, A., COOPER, J. L., BONHOURE, N., CHIANG, K., ELLIS, P. D., LANGFORD, C., ANDREWS, R. M. & VETRIE, D. 2010. Genomic approaches uncover increasing complexities in the regulatory landscape at the human SCL (TAL1) locus. PLoS One, 5, e9059.

DHAMI, P., COFFEY, A. J., ABBS, S., VERMEESCH, J. R., DUMANSKI, J. P., WOODWARD, K. J., ANDREWS, R. M., LANGFORD, C. & VETRIE, D. 2005. Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. Am J Hum Genet, 76, 750-62.

DI LORENZO, A. & BEDFORD, M. T. 2011. Histone arginine methylation. FEBS Lett, 585, 2024-31.

DIAMANDIS, P., WILDENHAIN, J., CLARKE, I. D., SACHER, A. G., GRAHAM, J., BELLOWS, D. S., LING, E. K., WARD, R. J., JAMIESON, L. G., TYERS, M. & DIRKS, P. B. 2007. Chemical genetics reveals a complex functional ground state of neural stem cells. Nat Chem Biol, 3, 268-73.

DIERKS, C., BEIGI, R., GUO, G. R., ZIRLIK, K., STEGERT, M. R., MANLEY, P., TRUSSELL, C., SCHMITT-GRAEFF, A., LANDWERLIN, K., VEELKEN, H. & WARMUTH, M. 2008. Expansion of Bcr-Abl-positive leukemic stem cells is dependent on Hedgehog pathway activation. Cancer Cell, 14, 238-49.

DILLON, S. 2008. Development of chromatin immunoprecipitation microarray technology for the identification of regulatory elements in the human genome. PhD, University of Cambridge.

DJEBALI, S., DAVIS, C. A., MERKEL, A., DOBIN, A., LASSMANN, T., MORTAZAVI, A., TANZER, A., LAGARDE, J., LIN, W., SCHLESINGER, F., XUE, C., MARINOV, G. K., KHATUN, J., WILLIAMS, B. A., ZALESKI, C., ROZOWSKY, J., RODER, M., KOKOCINSKI, F., et al. 2012. Landscape of transcription in human cells. Nature, 489, 101-8.

DORSAM, R. T. & GUTKIND, J. S. 2007. G-protein-coupled receptors and cancer. Nat Rev Cancer, 7, 79-94.

DOU, Y., MILNE, T. A., RUTHENBURG, A. J., LEE, S., LEE, J. W., VERDINE, G. L., ALLIS, C. D. & ROEDER, R. G. 2006. Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol, 13, 713-9.

DOU, Y., MILNE, T. A., TACKETT, A. J., SMITH, E. R., FUKUDA, A., WYSOCKA, J., ALLIS, C. D., CHAIT, B. T., HESS, J. L. & ROEDER, R. G. 2005. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell, 121, 873-85.

DOULATOV, S., NOTTA, F., EPPERT, K., NGUYEN, L. T., OHASHI, P. S. & DICK, J. E. 2010. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nat Immunol, 11, 585-93.

DOULATOV, S., NOTTA, F., LAURENTI, E. & DICK, J. E. 2012. Hematopoiesis: a human perspective. Cell Stem Cell, 10, 120-36.

DOVER, J., SCHNEIDER, J., TAWIAH-BOATENG, M. A., WOOD, A., DEAN, K., JOHNSTON, M. & SHILATIFARD, A. 2002. Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem, 277, 28368-71.

DRUKER, B. J., GUILHOT, F., O'BRIEN, S. G., GATHMANN, I., KANTARJIAN, H., GATTERMANN, N., DEININGER, M. W., SILVER, R. T., GOLDMAN, J. M., STONE, R. M., CERVANTES, F., HOCHHAUS, A., POWELL, B. L., GABRILOVE, J. L., ROUSSELOT, P., et al. 2006. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. N Engl J Med, 355, 2408-17.

DRUKER, B. J., TAMURA, S., BUCHDUNGER, E., OHNO, S., SEGAL, G. M., FANNING, S., ZIMMERMANN, J. & LYDON, N. B. 1996. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat Med, 2, 561-6.

DUBOIS, C. M., RUSCETTI, F. W., STANKOVA, J. & KELLER, J. R. 1994. Transforming growth factor-beta regulates c-kit message stability and cell-surface protein expression in hematopoietic progenitors. Blood, 83, 3138-45.

DUCOS, K., PANTERNE, B., FORTUNEL, N., HATZFELD, A., MONIER, M. N. & HATZFELD, J. 2000. p21(cip1) mRNA is controlled by endogenous transforming growth factor-beta1 in quiescent human hematopoietic stem/progenitor cells. J Cell Physiol, 184, 80-5.

DUNCAN, A. W., RATTIS, F. M., DIMASCIO, L. N., CONGDON, K. L., PAZIANOS, G., ZHAO, C., YOON, K., COOK, J. M., WILLERT, K., GAIANO, N. & REYA, T. 2005. Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. Nat Immunol, 6, 314-22.

DUNHAM, I., KUNDAJE, A., ALDRED, S. F., COLLINS, P. J., DAVIS, C. A., DOYLE, F., EPSTEIN, C. B., FRIETZE, S., HARROW, J., KAUL, R., KHATUN, J., LAJOIE, B. R., LANDT, S. G., LEE, B. K., PAULI, F., ROSENBLOOM, K. R., SABO, P., SAFI, A., SANYAL, A., SHORESH, N., et al. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 57-74.

EAVES, C. J., CASHMAN, J. D., WOLPE, S. D. & EAVES, A. C. 1993. Unresponsiveness of primitive chronic myeloid leukemia cells to macrophage inflammatory protein 1 alpha, an inhibitor of primitive normal hematopoietic cells. Proc Natl Acad Sci U S A, 90, 12015-9.

EBRAHIMKHANI, M. R., OAKLEY, F., MURPHY, L. B., MANN, J., MOLES, A., PERUGORRIA, M. J., ELLIS, E., LAKEY, A. F., BURT, A. D., DOUGLASS, A., WRIGHT, M. C., WHITE, S. A., JAFFRE, F., MAROTEAUX, L. & MANN, D. A. 2011. Stimulating healthy tissue regeneration by targeting the 5-HT(2)B receptor in chronic liver disease. Nat Med, 17, 1668-73.

ELNITSKI, L., JIN, V. X., FARNHAM, P. J. & JONES, S. J. 2006. Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. Genome Res, 16, 1455-64.

ENCODE, C. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science, 306, 636-40.

ENDOH, M., ENDO, T. A., ENDOH, T., FUJIMURA, Y., OHARA, O., TOYODA, T., OTTE, A. P., OKANO, M., BROCKDORFF, N., VIDAL, M. & KOSEKI, H. 2008. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development, 135, 1513-24.

ERNST, J. & KELLIS, M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol, 28, 817-25.

ERNST, P., MABON, M., DAVIDSON, A. J., ZON, L. I. & KORSMEYER, S. J. 2004. An Mll-dependent Hox program drives hematopoietic progenitor expansion. Curr Biol, 14, 2063-9.

ERNST, T., CHASE, A. J., SCORE, J., HIDALGO-CURTIS, C. E., BRYANT, C., JONES, A. V., WAGHORN, K., ZOI, K., ROSS, F. M., REITER, A., HOCHHAUS, A., DREXLER, H. G., DUNCOMBE, A., CERVANTES, F., OSCIER, D., BOULTWOOD, J., GRAND, F. H. & CROSS, N. C. 2010. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet, 42, 722-6.

ESSAFI, A., FERNANDEZ DE MATTOS, S., HASSEN, Y. A., SOEIRO, I., MUFTI, G. J., THOMAS, N. S., MEDEMA, R. H. & LAM, E. W. 2005. Direct transcriptional regulation of Bim by FoxO3a mediates STI571-induced apoptosis in Bcr-Abl-expressing cells. Oncogene, 24, 2317-29.

FADERL, S., TALPAZ, M., ESTROV, Z., O'BRIEN, S., KURZROCK, R. & KANTARJIAN, H. M. 1999. The biology of chronic myeloid leukemia. N Engl J Med, 341, 164-72.

FARNHAM, P. J. 2009. Insights from genomic profiling of transcription factors. Nat Rev Genet, 10, 605-16.

FEDURCO, M., ROMIEU, A., WILLIAMS, S., LAWRENCE, I. & TURCATTI, G. 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res, 34, e22.

FELSENFELD, G., BURGESS-BEUSSE, B., FARRELL, C., GASZNER, M., GHIRLANDO, R., HUANG, S., JIN, C., LITT, M., MAGDINIER, F., MUTSKOV, V., NAKATANI, Y., TAGAMI, H., WEST, A. & YUSUFZAI, T. 2004. Chromatin boundaries and chromatin domains. Cold Spring Harb Symp Quant Biol, 69, 245-50.

FELSENFELD, G. & GROUDINE, M. 2003. Controlling the double helix. Nature, 421, 448-53.

FIALKOW, P. J., GARTLER, S. M. & YOSHIDA, A. 1967. Clonal origin of chronic myelocytic leukemia in man. Proc Natl Acad Sci U S A, 58, 1468-71.

FIALKOW, P. J., JACOBSON, R. J. & PAPAYANNOPOULOU, T. 1977. Chronic myelocytic leukemia: clonal origin in a stem cell common to the granulocyte, erythrocyte, platelet and monocyte/macrophage. Am J Med, 63, 125-30.

FIERRO, F. A., TAUBENBERGER, A., PUECH, P. H., EHNINGER, G., BORNHAUSER, M., MULLER, D. J. & ILLMER, T. 2008. BCR/ABL expression of myeloid progenitors increases beta1-integrin mediated adhesion to stromal cells. J Mol Biol, 377, 1082-93.

FIGUEROA, M. E., REIMERS, M., THOMPSON, R. F., YE, K., LI, Y., SELZER, R. R., FRIDRIKSSON, J., PAIETTA, E., WIERNIK, P., GREEN, R. D., GREALLY, J. M. & MELNICK, A. 2008. An integrative genomic and epigenomic approach for the study of transcriptional regulation. PLoS One, 3, e1882.

FISCHLE, W., WANG, Y., JACOBS, S. A., KIM, Y., ALLIS, C. D. & KHORASANIZADEH, S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev, 17, 1870-81.

FISKUS, W., WANG, Y., SREEKUMAR, A., BUCKLEY, K. M., SHI, H., JILLELLA, A., USTUN, C., RAO, R., FERNANDEZ, P., CHEN, J., BALUSU, R., KOUL, S., ATADJA, P., MARQUEZ, V. E. & BHALLA, K. N. 2009. Combined epigenetic therapy with the histone methyltransferase EZH2 inhibitor 3-deazaneplanocin A and the histone deacetylase inhibitor panobinostat against human AML cells. Blood, 114, 2733-43.

FITCH, S. R., KIMBER, G. M., WILSON, N. K., PARKER, A., MIRSHEKAR-SYAHKAL, B., GOTTGENS, B., MEDVINSKY, A., DZIERZAK, E. & OTTERSBACH, K. 2012. Signaling from the Sympathetic Nervous System Regulates Hematopoietic Stem Cell Emergence during Embryogenesis. Cell Stem Cell, 11, 554-66.

FLEMING, H. E., JANZEN, V., LO CELSO, C., GUO, J., LEAHY, K. M., KRONENBERG, H. M. & SCADDEN, D. T. 2008. Wnt signaling in the niche enforces hematopoietic stem cell quiescence and is necessary to preserve self-renewal in vivo. Cell Stem Cell, 2, 274-83.

FLORIAN, S., SONNECK, K., HAUSWIRTH, A. W., KRAUTH, M. T., SCHERNTHANER, G. H., SPERR, W. R. & VALENT, P. 2006. Detection of molecular targets on the surface of CD34+/CD38-- stem cells in various myeloid malignancies. Leuk Lymphoma, 47, 207-22.

FONSECA, R. B., MOHR, A. M., WANG, L., SIFRI, Z. C., RAMESHWAR, P. & LIVINGSTON, D. H. 2005. The impact of a hypercatecholamine state on erythropoiesis following severe injury and the role of IL-6. J Trauma, 59, 884-9; discussion 889-90.

FOUDI, A., HOCHEDLINGER, K., VAN BUREN, D., SCHINDLER, J. W., JAENISCH, R., CAREY, V. & HOCK, H. 2009. Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. Nat Biotechnol, 27, 84-90.

FOUSE, S. D., SHEN, Y., PELLEGRINI, M., COLE, S., MEISSNER, A., VAN NESTE, L., JAENISCH, R. & FAN, G. 2008. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell, 2, 160-9.

FROMMER, M., MCDONALD, L. E., MILLAR, D. S., COLLIS, C. M., WATT, F., GRIGG, G. W., MOLLOY, P. L. & PAUL, C. L. 1992. A genomic sequencing protocol that yields a positive

display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A, 89, 1827-31.

FULLWOOD, M. J. & RUAN, Y. 2009. ChIP-based methods for the identification of long-range chromatin interactions. J Cell Biochem, 107, 30-9.

FULOP, G. M. & PHILLIPS, R. A. 1990. The scid mutation in mice causes a general defect in DNA repair. Nature, 347, 479-82.

FYODOROV, D. V. & KADONAGA, J. T. 2001. The many faces of chromatin remodeling: SWItching beyond transcription. Cell, 106, 523-5.

GAN, T., JUDE, C. D., ZAFFUTO, K. & ERNST, P. 2010. Developmentally induced Mll1 loss reveals defects in postnatal haematopoiesis. Leukemia, 24, 1732-41.

GARTEL, A. L. & TYNER, A. L. 1999. Transcriptional regulation of the p21((WAF1/CIP1)) gene. Exp Cell Res, 246, 280-9.

GEARHART, M. D., CORCORAN, C. M., WAMSTAD, J. A. & BARDWELL, V. J. 2006. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. Mol Cell Biol, 26, 6880-9.

GENTLEMAN, R. 2005. Bioinformatics and computational biology solutions using R and Bioconductor, New York, Springer Science+Business Media.

GEORGANTAS, R. W., 3RD, TANADVE, V., MALEHORN, M., HEIMFELD, S., CHEN, C., CARR, L., MARTINEZ-MURILLO, F., RIGGINS, G., KOWALSKI, J. & CIVIN, C. I. 2004. Microarray and serial analysis of gene expression analyses identify known and novel transcripts overexpressed in hematopoietic stem cells. Cancer Res, 64, 4434-41.

GERSHENZON, N. I. & IOSHIKHES, I. P. 2005. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics, 21, 1295-300.

GESBERT, F. & GRIFFIN, J. D. 2000. Bcr/Abl activates transcription of the Bcl-X gene through STAT5. Blood, 96, 2269-76.

GIRESI, P. G., KIM, J., MCDANIELL, R. M., IYER, V. R. & LIEB, J. D. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res, 17, 877-85.

GOLDKNOPF, I. L., TAYLOR, C. W., BAUM, R. M., YEOMAN, L. C., OLSON, M. O., PRESTAYKO, A. W. & BUSCH, H. 1975. Isolation and characterization of protein A24, a "histone-like" non-histone chromosomal protein. J Biol Chem, 250, 7182-7.

GORRE, M. E., MOHAMMED, M., ELLWOOD, K., HSU, N., PAQUETTE, R., RAO, P. N. & SAWYERS, C. L. 2001. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science, 293, 876-80.

GRAHAM, S. M., VASS, J. K., HOLYOAKE, T. L. & GRAHAM, G. J. 2007. Transcriptional analysis of quiescent and proliferating CD34+ human hemopoietic cells from normal and chronic myeloid leukemia sources. Stem Cells, 25, 3111-20.

GRAY, P. A., FU, H., LUO, P., ZHAO, Q., YU, J., FERRARI, A., TENZEN, T., YUK, D. I., TSUNG, E. F., CAI, Z., ALBERTA, J. A., CHENG, L. P., LIU, Y., STENMAN, J. M., VALERIUS, M. T., BILLINGS, N., KIM, H. A., GREENBERG, M. E., MCMAHON, A. P., ROWITCH, D. H., STILES, C. D. & MA, Q. 2004. Mouse brain organization revealed through direct genome-scale TF expression analysis. Science, 306, 2255-7.

GREENMAN, C., STEPHENS, P., SMITH, R., DALGLIESH, G. L., HUNTER, C., BIGNELL, G., DAVIES, H., TEAGUE, J., BUTLER, A., STEVENS, C., EDKINS, S., O'MEARA, S., VASTRIK, I., SCHMIDT, E. E., AVIS, T., BARTHORPE, S., BHAMRA, G., BUCK, G., CHOUDHURY, B., et al. 2007. Patterns of somatic mutation in human cancer genomes. Nature, 446, 153-8.

GROFFEN, J., STEPHENSON, J. R., HEISTERKAMP, N., DE KLEIN, A., BARTRAM, C. R. & GROSVELD, G. 1984. Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. Cell, 36, 93-9.

GROSS, D. S. & GARRARD, W. T. 1988. Nuclease hypersensitive sites in chromatin. Annu Rev Biochem, 57, 159-97.

GRUBACH, L., JUHL-CHRISTENSEN, C., RETHMEIER, A., OLESEN, L. H., AGGERHOLM, A., HOKLAND, P. & OSTERGAARD, M. 2008. Gene expression profiling of Polycomb, Hox and Meis genes in patients with acute myeloid leukaemia. Eur J Haematol, 81, 112-22.

GUENTHER, M. G., LEVINE, S. S., BOYER, L. A., JAENISCH, R. & YOUNG, R. A. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. Cell, 130, 77-88.

GUTKIND, J. S., NOVOTNY, E. A., BRANN, M. R. & ROBBINS, K. C. 1991. Muscarinic acetylcholine receptor subtypes as agonist-dependent oncogenes. Proc Natl Acad Sci U S A, 88, 4703-7.

HAHN, S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol, 11, 394-403.

HAMILTON, A., HELGASON, G. V., SCHEMIONEK, M., ZHANG, B., MYSSINA, S., ALLAN, E. K., NICOLINI, F. E., MULLER-TIDOW, C., BHATIA, R., BRUNTON, V. G., KOSCHMIEDER, S. & HOLYOAKE, T. L. 2012. Chronic myeloid leukemia stem cells are not dependent on Bcr-Abl kinase activity for their survival. Blood, 119, 1501-10.

HAN, H., CORTEZ, C. C., YANG, X., NICHOLS, P. W., JONES, P. A. & LIANG, G. 2011. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. Hum Mol Genet, 20, 4299-310.

HANLON, S. E. & LIEB, J. D. 2004. Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. Curr Opin Genet Dev, 14, 697-705.

HANYALOGLU, A. C. & VON ZASTROW, M. 2008. Regulation of GPCRs by endocytic membrane trafficking and its potential implications. Annu Rev Pharmacol Toxicol, 48, 537-68.

HARGREAVES, D. C., HORNG, T. & MEDZHITOV, R. 2009. Control of inducible gene expression by signal-dependent transcriptional elongation. Cell, 138, 129-45.

HARLOW, E. & LANE, D. 1999. Using antibodies : a laboratory manual, Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

HARPER, J. W., ELLEDGE, S. J., KEYOMARSI, K., DYNLACHT, B., TSAI, L. H., ZHANG, P., DOBROWOLSKI, S., BAI, C., CONNELL-CROWLEY, L., SWINDELL, E. & ET AL. 1995. Inhibition of cyclin-dependent kinases by p21. Mol Biol Cell, 6, 387-400.

HARRIS, M. B., MOSTECKI, J. & ROTHMAN, P. B. 2005. Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. J Biol Chem, 280, 13114-21.

HASHIMSHONY, T., ZHANG, J., KESHET, I., BUSTIN, M. & CEDAR, H. 2003. The role of DNA methylation in setting up chromatin structure during development. Nat Genet, 34, 187-92.

HATZFELD, J., LI, M. L., BROWN, E. L., SOOKDEO, H., LEVESQUE, J. P., O'TOOLE, T., GURNEY, C., CLARK, S. C. & HATZFELD, A. 1991. Release of early human hematopoietic progenitors from quiescence by antisense transforming growth factor beta 1 or Rb oligonucleotides. J Exp Med, 174, 925-9.

HAUG, J. S., HE, X. C., GRINDLEY, J. C., WUNDERLICH, J. P., GAUDENZ, K., ROSS, J. T., PAULSON, A., WAGNER, K. P., XIE, Y., ZHU, R., YIN, T., PERRY, J. M., HEMBREE, M. J., REDENBAUGH, E. P., RADICE, G. L., SEIDEL, C. & LI, L. 2008. N-cadherin expression level distinguishes reserved versus primed states of hematopoietic stem cells. Cell Stem Cell, 2, 367-79.

HAWKINS, R. D., HON, G. C., LEE, L. K., NGO, Q., LISTER, R., PELIZZOLA, M., EDSALL, L. E., KUAN, S., LUU, Y., KLUGMAN, S., ANTOSIEWICZ-BOURGET, J., YE, Z., ESPINOZA, C., AGARWAHL, S., SHEN, L., RUOTTI, V., WANG, W., STEWART, R., THOMSON, J. A., ECKER, J. R. & REN, B. 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell, 6, 479-91.

HAYLOCK, D. N., TO, L. B., DOWSE, T. L., JUTTNER, C. A. & SIMMONS, P. J. 1992. Ex vivo expansion and maturation of peripheral blood CD34+ cells into the myeloid lineage. Blood, 80, 1405-12.

HAZLEHURST, L. A., DAMIANO, J. S., BUYUKSAL, I., PLEDGER, W. J. & DALTON, W. S. 2000. Adhesion to fibronectin via beta1 integrins regulates p27kip1 levels and contributes to cell adhesion mediated drug resistance (CAM-DR). Oncogene, 19, 4319-27.

HE, W., DORN, D. C., ERDJUMENT-BROMAGE, H., TEMPST, P., MOORE, M. A. & MASSAGUE, J. 2006. Hematopoiesis controlled by distinct TIF1gamma and Smad4 branches of the TGFbeta pathway. Cell, 125, 929-41.

HE, Y. F., LI, B. Z., LI, Z., LIU, P., WANG, Y., TANG, Q., DING, J., JIA, Y., CHEN, Z., LI, L., SUN, Y., LI, X., DAI, Q., SONG, C. X., ZHANG, K., HE, C. & XU, G. L. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science, 333, 1303-7.

HEASLEY, L. E. 2001. Autocrine and paracrine signaling through neuropeptide receptors in human cancer. Oncogene, 20, 1563-9.

HEHLMANN, R., HOCHHAUS, A. & BACCARANI, M. 2007. Chronic myeloid leukaemia. Lancet, 370, 342-50.

HEINTZMAN, N. D., HON, G. C., HAWKINS, R. D., KHERADPOUR, P., STARK, A., HARP, L. F., YE, Z., LEE, L. K., STUART, R. K., CHING, C. W., CHING, K. A., ANTOSIEWICZ-BOURGET, J. E., LIU, H., ZHANG, X., GREEN, R. D., LOBANENKOV, V. V., STEWART, R., THOMSON, J. A., CRAWFORD, G. E., KELLIS, M. & REN, B. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature, 459, 108-12.

HEINTZMAN, N. D., STUART, R. K., HON, G., FU, Y., CHING, C. W., HAWKINS, R. D., BARRERA, L. O., VAN CALCAR, S., QU, C., CHING, K. A., WANG, W., WENG, Z., GREEN, R. D., CRAWFORD, G. E. & REN, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet, 39, 311-8.

HEISTERKAMP, N., JENSTER, G., TEN HOEVE, J., ZOVICH, D., PATTENGALE, P. K. & GROFFEN, J. 1990. Acute leukaemia in bcr/abl transgenic mice. Nature, 344, 251-3.

HELDIN, C. H., MIYAZONO, K. & TEN DIJKE, P. 1997. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. Nature, 390, 465-71.

HELLMAN, A. & CHESS, A. 2007. Gene body-specific methylation on the active X chromosome. Science, 315, 1141-3.

HENIKOFF, S. 2000. Heterochromatin function in complex genomes. Biochim Biophys Acta, 1470, O1-8.

HERRMANN, H., CERNY-REITERER, S., GLEIXNER, K. V., BLATT, K., HERNDLHOFER, S., RABITSCH, W., JAGER, E., MITTERBAUER-HOHENDANNER, G., STREUBEL, B., SELZER, E., SCHWARZINGER, I., SPERR, W. R. & VALENT, P. 2012. CD34(+)/CD38(-) stem cells in chronic myeloid leukemia express Siglec-3 (CD33) and are responsive to the CD33-targeting drug gemtuzumab/ozogamicin. Haematologica, 97, 219-26.

HILLIER, L. W., MARTH, G. T., QUINLAN, A. R., DOOLING, D., FEWELL, G., BARNETT, D., FOX, P., GLASSCOCK, J. I., HICKENBOTHAM, M., HUANG, W., MAGRINI, V. J., RICHT, R. J., SANDER, S. N., STEWART, D. A., STROMBERG, M., TSUNG, E. F., WYLIE, T., SCHEDL, T., WILSON, R. K. & MARDIS, E. R. 2008. Whole-genome sequencing and variant discovery in C. elegans. Nat Methods, 5, 183-8.

HOELBL, A., KOVACIC, B., KERENYI, M. A., SIMMA, O., WARSCH, W., CUI, Y., BEUG, H., HENNIGHAUSEN, L., MORIGGL, R. & SEXL, V. 2006. Clarifying the role of Stat5 in lymphoid development and Abelson-induced transformation. Blood, 107, 4898-906.

HOLLIDAY, R. & PUGH, J. E. 1975. DNA modification mechanisms and gene activity during development. Science, 187, 226-32.

HOLYOAKE, T., JIANG, X., EAVES, C. & EAVES, A. 1999. Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. Blood, 94, 2056-64.

HOLYOAKE, T. L., JIANG, X., JORGENSEN, H. G., GRAHAM, S., ALCORN, M. J., LAIRD, C., EAVES, A. C. & EAVES, C. J. 2001. Primitive quiescent leukemic cells from patients with chronic myeloid leukemia spontaneously initiate factor-independent growth in vitro in association with up-regulation of expression of interleukin-3. Blood, 97, 720-8.

HOOPER, M., HARDY, K., HANDYSIDE, A., HUNTER, S. & MONK, M. 1987. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature, 326, 292-5.

HUBER, T. L., KOUSKOFF, V., FEHLING, H. J., PALIS, J. & KELLER, G. 2004. Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. Nature, 432, 625-30.

HUDSON, M. E. & SNYDER, M. 2006. High-throughput methods of regulatory element discovery. Biotechniques, 41, 673, 675, 677 passim.

HUGHES, T., DEININGER, M., HOCHHAUS, A., BRANFORD, S., RADICH, J., KAEDA, J., BACCARANI, M., CORTES, J., CROSS, N. C., DRUKER, B. J., GABERT, J., GRIMWADE, D.,

HEHLMANN, R., KAMEL-REID, S., LIPTON, J. H., LONGTINE, J., MARTINELLI, G., SAGLIO, G., SOVERINI, S., STOCK, W. & GOLDMAN, J. M. 2006. Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results. Blood, 108, 28-37.

HUNTLY, B. J., SHIGEMATSU, H., DEGUCHI, K., LEE, B. H., MIZUNO, S., DUCLOS, N., ROWAN, R., AMARAL, S., CURLEY, D., WILLIAMS, I. R., AKASHI, K. & GILLILAND, D. G. 2004. MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. Cancer Cell, 6, 587-96.

HURTZ, C., HATZI, K., CERCHIETTI, L., BRAIG, M., PARK, E., KIM, Y. M., HERZOG, S., RAMEZANI-RAD, P., JUMAA, H., MULLER, M. C., HOFMANN, W. K., HOCHHAUS, A., YE, B. H., AGARWAL, A., DRUKER, B. J., SHAH, N. P., MELNICK, A. M. & MUSCHEN, M. 2011. BCL6-mediated repression of p53 is critical for leukemia stem cell survival in chronic myeloid leukemia. J Exp Med, 208, 2163-74.

IHGS, C. 2004. Finishing the euchromatic sequence of the human genome. Nature, 431, 931-45.

IKUTA, K. & WEISSMAN, I. L. 1992. Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation. Proc Natl Acad Sci U S A, 89, 1502-6.

ILARIA, R. L., JR. & VAN ETTEN, R. A. 1996. P210 and P190(BCR/ABL) induce the tyrosine phosphorylation and DNA binding activity of multiple specific STAT family members. J Biol Chem, 271, 31704-10.

ILLINGWORTH, R. S. & BIRD, A. P. 2009. CpG islands--'a rough guide'. FEBS Lett, 583, 1713-20.

IRIZARRY, R. A., BOLSTAD, B. M., COLLIN, F., COPE, L. M., HOBBS, B. & SPEED, T. P. 2003. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res, 31, e15.

IRIZARRY, R. A., LADD-ACOSTA, C., WEN, B., WU, Z., MONTANO, C., ONYANGO, P., CUI, H., GABO, K., RONGIONE, M., WEBSTER, M., JI, H., POTASH, J. B., SABUNCIYAN, S. & FEINBERG, A. P. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet, 41, 178-86.

ITO, K., BERNARDI, R., MOROTTI, A., MATSUOKA, S., SAGLIO, G., IKEDA, Y., ROSENBLATT, J., AVIGAN, D. E., TERUYA-FELDSTEIN, J. & PANDOLFI, P. P. 2008. PML targeting eradicates quiescent leukaemia-initiating cells. Nature, 453, 1072-8.

ITO, K., HIRAO, A., ARAI, F., TAKUBO, K., MATSUOKA, S., MIYAMOTO, K., OHMURA, M., NAKA, K., HOSOKAWA, K., IKEDA, Y. & SUDA, T. 2006. Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. Nat Med, 12, 446-51.

ITO, M., HIRAMATSU, H., KOBAYASHI, K., SUZUE, K., KAWAHATA, M., HIOKI, K., UEYAMA, Y., KOYANAGI, Y., SUGAMURA, K., TSUJI, K., HEIKE, T. & NAKAHATA, T. 2002. NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. Blood, 100, 3175-82.

ITO, T., BULGER, M., PAZIN, M. J., KOBAYASHI, R. & KADONAGA, J. T. 1997. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell, 90, 145-55.

ITO, T., KWON, H. Y., ZIMDAHL, B., CONGDON, K. L., BLUM, J., LENTO, W. E., ZHAO, C., LAGOO, A., GERRARD, G., FORONI, L., GOLDMAN, J., GOH, H., KIM, S. H., KIM, D. W., CHUAH, C., OEHLER, V. G., RADICH, J. P., JORDAN, C. T. & REYA, T. 2010. Regulation of myeloid leukaemia by the cell-fate determinant Musashi. Nature, 466, 765-8.

IVANOVA, N. B., DIMOS, J. T., SCHANIEL, C., HACKNEY, J. A., MOORE, K. A. & LEMISCHKA, I. R. 2002. A stem cell molecular signature. Science, 298, 601-4.

IVANOVS, A., RYBTSOV, S., WELCH, L., ANDERSON, R. A., TURNER, M. L. & MEDVINSKY, A. 2011. Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. J Exp Med, 208, 2417-27.

IWASAKI, H. & AKASHI, K. 2007. Myeloid lineage commitment from the hematopoietic stem cell. Immunity, 26, 726-40.

IWASE, S., LAN, F., BAYLISS, P., DE LA TORRE-UBIETA, L., HUARTE, M., QI, H. H., WHETSTINE, J. R., BONNI, A., ROBERTS, T. M. & SHI, Y. 2007. The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. Cell, 128, 1077-88.

IYER, V. R., HORAK, C. E., SCAFE, C. S., BOTSTEIN, D., SNYDER, M. & BROWN, P. O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature, 409, 533-8.

JABBOUR, E., KANTARJIAN, H., O'BRIEN, S., SHAN, J., QUINTAS-CARDAMA, A., FADERL, S., GARCIA-MANERO, G., RAVANDI, F., RIOS, M. B. & CORTES, J. 2011. The achievement of an early complete cytogenetic response is a major determinant for outcome in patients with early chronic phase chronic myeloid leukemia treated with tyrosine kinase inhibitors. Blood, 118, 4541-6; quiz 4759.

JACKSON-GRUSBY, L., BEARD, C., POSSEMATO, R., TUDOR, M., FAMBROUGH, D., CSANKOVSZKI, G., DAUSMAN, J., LEE, P., WILSON, C., LANDER, E. & JAENISCH, R. 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet, 27, 31-9.

JACOBSEN, S. E., RUSCETTI, F. W., DUBOIS, C. M., LEE, J., BOONE, T. C. & KELLER, J. R. 1991. Transforming growth factor-beta trans-modulates the expression of colony stimulating factor receptors on murine hematopoietic progenitor cell lines. Blood, 77, 1706-16.

JAMIESON, C. H., AILLES, L. E., DYLLA, S. J., MUIJTJENS, M., JONES, C., ZEHNDER, J. L., GOTLIB, J., LI, K., MANZ, M. G., KEATING, A., SAWYERS, C. L. & WEISSMAN, I. L. 2004. Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. N Engl J Med, 351, 657-67.

JARAS, M., JOHNELS, P., HANSEN, N., AGERSTAM, H., TSAPOGAS, P., RISSLER, M., LASSEN, C., OLOFSSON, T., BJERRUM, O. W., RICHTER, J. & FIORETOS, T. 2010. Isolation and killing of candidate chronic myeloid leukemia stem cells by antibody targeting of IL-1 receptor accessory protein. Proc Natl Acad Sci U S A, 107, 16280-5.

JENUWEIN, T. & ALLIS, C. D. 2001. Translating the histone code. Science, 293, 1074-80.

JI, H., EHRLICH, L. I., SEITA, J., MURAKAMI, P., DOI, A., LINDAU, P., LEE, H., ARYEE, M. J., IRIZARRY, R. A., KIM, K., ROSSI, D. J., INLAY, M. A., SERWOLD, T., KARSUNKY, H., HO, L., DALEY, G. Q., WEISSMAN, I. L. & FEINBERG, A. P. 2010. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature, 467, 338-42.

JI, J., VIJAYARAGAVAN, K., BOSSE, M., MENENDEZ, P., WEISEL, K. & BHATIA, M. 2008. OP9 stroma augments survival of hematopoietic precursors and progenitors during hematopoietic differentiation from human embryonic stem cells. Stem Cells, 26, 2485-95.

JIANG, H., SHUKLA, A., WANG, X., CHEN, W. Y., BERNSTEIN, B. E. & ROEDER, R. G. 2011. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell, 144, 513-25.

JIANG, X., LOPEZ, A., HOLYOAKE, T., EAVES, A. & EAVES, C. 1999. Autocrine production and action of IL-3 and granulocyte colony-stimulating factor in chronic myeloid leukemia. Proc Natl Acad Sci U S A, 96, 12804-9.

JIANG, X., ZHAO, Y., FORREST, D., SMITH, C., EAVES, A. & EAVES, C. 2008. Stem cell biomarkers in chronic myeloid leukemia. Dis Markers, 24, 201-16.

JIANG, X., ZHAO, Y., SMITH, C., GASPARETTO, M., TURHAN, A., EAVES, A. & EAVES, C. 2007. Chronic myeloid leukemia stem cells possess multiple unique features of resistance to BCR-ABL targeted therapies. Leukemia, 21, 926-35.

JONES, P. A. 1999. The DNA methylation paradox. Trends Genet, 15, 34-7.

JONES, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet, 13, 484-92.

JONES, P. A. & BAYLIN, S. B. 2007. The epigenomics of cancer. Cell, 128, 683-92.

JONES, P. A. & LIANG, G. 2009. Rethinking how DNA methylation patterns are maintained. Nat Rev Genet, 10, 805-11.

JULIUS, D., LIVELLI, T. J., JESSELL, T. M. & AXEL, R. 1989. Ectopic expression of the serotonin 1c receptor and the triggering of malignant transformation. Science, 244, 1057-62.

KAELIN, W. G., JR. 2005. The concept of synthetic lethality in the context of anticancer therapy. Nat Rev Cancer, 5, 689-98.

KALINKOVICH, A., SPIEGEL, A., SHIVTIEL, S., KOLLET, O., JORDANEY, N., PIACIBELLO, W. & LAPIDOT, T. 2009. Blood-forming stem cells are nervous: direct and indirect regulation of immature human CD34+ cells by the nervous system. Brain Behav Immun, 23, 1059-65.

KAMEL-REID, S. & DICK, J. E. 1988. Engraftment of immune-deficient mice with human hematopoietic stem cells. Science, 242, 1706-9.

KAPRANOV, P., CAWLEY, S. E., DRENKOW, J., BEKIRANOV, S., STRAUSBERG, R. L., FODOR, S. P. & GINGERAS, T. R. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. Science, 296, 916-9.

KARANU, F. N., MURDOCH, B., GALLACHER, L., WU, D. M., KOREMOTO, M., SAKANO, S. & BHATIA, M. 2000. The notch ligand jagged-1 represents a novel growth factor of human hematopoietic stem cells. J Exp Med, 192, 1365-72.

KARANU, F. N., MURDOCH, B., MIYABAYASHI, T., OHNO, M., KOREMOTO, M., GALLACHER, L., WU, D., ITOH, A., SAKANO, S. & BHATIA, M. 2001. Human homologues of Delta-1 and Delta-4 function as mitogenic regulators of primitive human hematopoietic cells. Blood, 97, 1960-7.

KARPEN, G. H. & ALLSHIRE, R. C. 1997. The case for epigenetic effects on centromere identity and function. Trends Genet, 13, 489-96.

KARPF, A. R. & MATSUI, S. 2005. Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. Cancer Res, 65, 8635-9.

KASS, S. U., LANDSBERGER, N. & WOLFFE, A. P. 1997. DNA methylation directs a time-dependent repression of transcription initiation. Curr Biol, 7, 157-65.

KATAN-KHAYKOVICH, Y. & STRUHL, K. 2005. Heterochromatin formation involves changes in histone modifications over multiple cell generations. EMBO J, 24, 2138-49.

KATAYAMA, Y., BATTISTA, M., KAO, W. M., HIDALGO, A., PEIRED, A. J., THOMAS, S. A. & FRENETTE, P. S. 2006. Signals from the sympathetic nervous system regulate hematopoietic stem cell egress from bone marrow. Cell, 124, 407-21.

KAWASAKI, H., ECKNER, R., YAO, T. P., TAIRA, K., CHIU, R., LIVINGSTON, D. M. & YOKOYAMA, K. K. 1998. Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation. Nature, 393, 284-9.

KEENE, M. A., CORCES, V., LOWENHAUPT, K. & ELGIN, S. C. 1981. DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription. Proc Natl Acad Sci U S A, 78, 143-6.

KHARCHENKO, P. V., TOLSTORUKOV, M. Y. & PARK, P. J. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat Biotechnol, 26, 1351-9.

KIEL, M. J., HE, S., ASHKENAZI, R., GENTRY, S. N., TETA, M., KUSHNER, J. A., JACKSON, T. L. & MORRISON, S. J. 2007. Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU. Nature, 449, 238-42.

KIEL, M. J., YILMAZ, O. H., IWASHITA, T., TERHORST, C. & MORRISON, S. J. 2005. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell, 121, 1109-21.

KIM, T. H., BARRERA, L. O., ZHENG, M., QU, C., SINGER, M. A., RICHMOND, T. A., WU, Y., GREEN, R. D. & REN, B. 2005. A high-resolution map of active promoters in the human genome. Nature, 436, 876-80.

KIMURA, S., ROBERTS, A. W., METCALF, D. & ALEXANDER, W. S. 1998. Hematopoietic stem cell deficiencies in mice lacking c-Mpl, the receptor for thrombopoietin. Proc Natl Acad Sci U S A, 95, 1195-200.

KIRSTETTER, P., ANDERSON, K., PORSE, B. T., JACOBSEN, S. E. & NERLOV, C. 2006. Activation of the canonical Wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. Nat Immunol, 7, 1048-56.

KIZER, K. O., PHATNANI, H. P., SHIBATA, Y., HALL, H., GREENLEAF, A. L. & STRAHL, B. D. 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol Cell Biol, 25, 3305-16.

KODAMA, H., NOSE, M., NIIDA, S. & NISHIKAWA, S. 1994. Involvement of the c-kit receptor in the adhesion of hematopoietic stem cells to stromal cells. Exp Hematol, 22, 979-84.

KOLASINSKA-ZWIERZ, P., DOWN, T., LATORRE, I., LIU, T., LIU, X. S. & AHRINGER, J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet, 41, 376-81.

KONDO, M., WEISSMAN, I. L. & AKASHI, K. 1997. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell, 91, 661-72.

KOUZARIDES, T. 2007. Chromatin modifications and their function. Cell, 128, 693-705.

KRIVTSOV, A. V. & ARMSTRONG, S. A. 2007. MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer, 7, 823-33.

KRIVTSOV, A. V., FENG, Z., LEMIEUX, M. E., FABER, J., VEMPATI, S., SINHA, A. U., XIA, X., JESNECK, J., BRACKEN, A. P., SILVERMAN, L. B., KUTOK, J. L., KUNG, A. L. & ARMSTRONG, S. A. 2008. H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell, 14, 355-68.

KRIVTSOV, A. V., TWOMEY, D., FENG, Z., STUBBS, M. C., WANG, Y., FABER, J., LEVINE, J. E., WANG, J., HAHN, W. C., GILLILAND, D. G., GOLUB, T. R. & ARMSTRONG, S. A. 2006. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature, 442, 818-22.

KRONENWETT, R., BUTTERWECK, U., STEIDL, U., KLISZEWSKI, S., NEUMANN, F., BORK, S., BLANCO, E. D., ROES, N., GRAF, T., BRORS, B., EILS, R., MAERCKER, C., KOBBE, G., GATTERMANN, N. & HAAS, R. 2005. Distinct molecular phenotype of malignant CD34(+) hematopoietic stem and progenitor cells in chronic myelogenous leukemia. Oncogene, 24, 5313-24.

KU, M., KOCHE, R. P., RHEINBAY, E., MENDENHALL, E. M., ENDOH, M., MIKKELSEN, T. S., PRESSER, A., NUSBAUM, C., XIE, X., CHI, A. S., ADLI, M., KASIF, S., PTASZEK, L. M., COWAN, C. A., LANDER, E. S., KOSEKI, H. & BERNSTEIN, B. E. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet, 4, e1000242.

KUHN, R. M., KAROLCHIK, D., ZWEIG, A. S., TRUMBOWER, H., THOMAS, D. J., THAKKAPALLAYIL, A., SUGNET, C. W., STANKE, M., SMITH, K. E., SIEPEL, A., ROSENBLOOM, K. R., RHEAD, B., RANEY, B. J., POHL, A., PEDERSEN, J. S., HSU, F., HINRICHS, A. S., HARTE, R. A., DIEKHANS, M., CLAWSON, H., BEJERANO, G., BARBER, G. P., BAERTSCH, R., HAUSSLER, D. & KENT, W. J. 2007. The UCSC genome browser database: update 2007. Nucleic Acids Res, 35, D668-73.

KUMARI, A., BRENDEL, C., HOCHHAUS, A., NEUBAUER, A. & BURCHERT, A. 2012. Low BCR-ABL expression levels in hematopoietic precursor cells enable persistence of chronic myeloid leukemia under imatinib. Blood, 119, 530-9.

KUNG, A. L., REBEL, V. I., BRONSON, R. T., CH'NG, L. E., SIEFF, C. A., LIVINGSTON, D. M. & YAO, T. P. 2000. Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. Genes Dev, 14, 272-7.

KWON, H., IMBALZANO, A. N., KHAVARI, P. A., KINGSTON, R. E. & GREEN, M. R. 1994. Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. Nature, 370, 477-81.

LANCRIN, C., SROCZYNSKA, P., STEPHENSON, C., ALLEN, T., KOUSKOFF, V. & LACAUD, G. 2009. The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. Nature, 457, 892-5.

LANDEIRA, D., SAUER, S., POOT, R., DVORKINA, M., MAZZARELLA, L., JORGENSEN, H. F., PEREIRA, C. F., LELEU, M., PICCOLO, F. M., SPIVAKOV, M., BROOKES, E., POMBO, A., FISHER, C., SKARNES, W. C., SNOEK, T., BEZSTAROSTI, K., DEMMERS, J., KLOSE, R. J., CASANOVA, M., TAVARES, L., BROCKDORFF, N., MERKENSCHLAGER, M. & FISHER, A. G. 2010. Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nat Cell Biol, 12, 618-24.

LANEUVILLE, P. 1995. Abl tyrosine protein kinase. Semin Immunol, 7, 255-66.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 10, R25.

LANGST, G., BONTE, E. J., CORONA, D. F. & BECKER, P. B. 1999. Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. Cell, 97, 843-52.

LANSDORP, P. M., SUTHERLAND, H. J. & EAVES, C. J. 1990. Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow. J Exp Med, 172, 363-6.

LAPIDOT, T., PFLUMIO, F., DOEDENS, M., MURDOCH, B., WILLIAMS, D. E. & DICK, J. E. 1992. Cytokine stimulation of multilineage hematopoiesis from immature human cells engrafted in SCID mice. Science, 255, 1137-41.

LAROCHELLE, A., SAVONA, M., WIGGINS, M., ANDERSON, S., ICHWAN, B., KEYVANFAR, K., MORRISON, S. J. & DUNBAR, C. E. 2011. Human and rhesus macaque hematopoietic stem cells cannot be purified based only on SLAM family markers. Blood, 117, 1550-4.

LAROCHELLE, A., VORMOOR, J., HANENBERG, H., WANG, J. C., BHATIA, M., LAPIDOT, T., MORITZ, T., MURDOCH, B., XIAO, X. L., KATO, I., WILLIAMS, D. A. & DICK, J. E. 1996. Identification of primitive human hematopoietic cells capable of repopulating NOD/SCID mouse bone marrow: implications for gene therapy. Nat Med, 2, 1329-37.

LARSSON, J., BLANK, U., HELGADOTTIR, H., BJORNSSON, J. M., EHINGER, M., GOUMANS, M. J., FAN, X., LEVEEN, P. & KARLSSON, S. 2003. TGF-beta signaling-deficient hematopoietic stem cells have normal self-renewal and regenerative ability in vivo despite increased proliferative capacity in vitro. Blood, 102, 3129-35.

LARSSON, J., BLANK, U., KLINTMAN, J., MAGNUSSON, M. & KARLSSON, S. 2005. Quiescence of hematopoietic stem cells and maintenance of the stem cell pool is not dependent on TGF-beta signaling in vivo. Exp Hematol, 33, 592-6.

LATCHMAN, D. S. 2005. Gene regulation : a eukaryotic perspective, New York, Taylor & Francis.

LAURENT, B. C. 2006. Chromatin dynamics in cellular function, Berlin ; New York, Springer.

LEE, C. K., SHIBATA, Y., RAO, B., STRAHL, B. D. & LIEB, J. D. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet, 36, 900-5.

LEE, E. R., MURDOCH, F. E. & FRITSCH, M. K. 2007. High histone acetylation and decreased polycomb repressive complex 2 member levels regulate gene specific transcriptional changes during early embryonic stem cell differentiation induced by retinoic acid. Stem Cells, 25, 2191-9.

LEE, J. S., SMITH, E. & SHILATIFARD, A. 2010. The language of histone crosstalk. Cell, 142, 682-5.

LEE, J. T. & JAENISCH, R. 1997. The (epi)genetic control of mammalian X-chromosome inactivation. Curr Opin Genet Dev, 7, 274-80.

LEE, M. G., VILLA, R., TROJER, P., NORMAN, J., YAN, K. P., REINBERG, D., DI CROCE, L. & SHIEKHATTAR, R. 2007. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science, 318, 447-50.

LEE, T. I., JENNER, R. G., BOYER, L. A., GUENTHER, M. G., LEVINE, S. S., KUMAR, R. M., CHEVALIER, B., JOHNSTONE, S. E., COLE, M. F., ISONO, K., KOSEKI, H., FUCHIKAMI, T., ABE, K., MURRAY, H. L., ZUCKER, J. P., YUAN, B., BELL, G. W., HERBOLSHEIMER, E., et al. A. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell, 125, 301-13.

LESSARD, J., SCHUMACHER, A., THORSTEINSDOTTIR, U., VAN LOHUIZEN, M., MAGNUSON, T. & SAUVAGEAU, G. 1999. Functional antagonism of the Polycomb-Group genes eed and Bmi1 in hemopoietic cell proliferation. Genes Dev, 13, 2691-703.

LEVIN, R. H., WHANG, J., TJIO, J. H., CARBONE, P. P., FREI, E., 3RD & FREIREICH, E. J. 1963. Persistent Mitosis of Transfused Homologous Leukocytes in Children Receiving Antileukemic Therapy. Science, 142, 1305-11.

LI, E., BESTOR, T. H. & JAENISCH, R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell, 69, 915-26.

LI, G., MARGUERON, R., KU, M., CHAMBON, P., BERNSTEIN, B. E. & REINBERG, D. 2010. Jarid2 and PRC2, partners in regulating gene expression. Genes Dev, 24, 368-80.

LI, H., RUAN, J. & DURBIN, R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res, 18, 1851-8.

LI, L., HE, S., SUN, J. M. & DAVIE, J. R. 2004. Gene regulation by Sp1 and Sp3. Biochem Cell Biol, 82, 460-71.

LI, Q., PETERSON, K. R., FANG, X. & STAMATOYANNOPOULOS, G. 2002. Locus control regions. Blood, 100, 3077-86.

LI, Q., ZHANG, M., DUAN, Z. & STAMATOYANNOPOULOS, G. 1999. Structural analysis and mapping of DNase I hypersensitivity of HS5 of the beta-globin locus control region. Genomics, 61, 183-93.

LIEB, J. D., LIU, X., BOTSTEIN, D. & BROWN, P. O. 2001. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet, 28, 327-34.

LIM, D. A., HUANG, Y. C., SWIGUT, T., MIRICK, A. L., GARCIA-VERDUGO, J. M., WYSOCKA, J., ERNST, P. & ALVAREZ-BUYLLA, A. 2009. Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells. Nature, 458, 529-33.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature, 462, 315-22.

LIU, B., SUN, Y., JIANG, F., ZHANG, S., WU, Y., LAN, Y., YANG, X. & MAO, N. 2003. Disruption of Smad5 gene leads to enhanced proliferation of high-proliferative potential precursors during embryonic hematopoiesis. Blood, 101, 124-33.

LO CELSO, C., FLEMING, H. E., WU, J. W., ZHAO, C. X., MIAKE-LYE, S., FUJISAKI, J., COTE, D., ROWE, D. W., LIN, C. P. & SCADDEN, D. T. 2009. Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. Nature, 457, 92-6.

LO, W. S., TRIEVEL, R. C., ROJAS, J. R., DUGGAN, L., HSU, J. Y., ALLIS, C. D., MARMORSTEIN, R. & BERGER, S. L. 2000. Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. Mol Cell, 5, 917-26.

LOH, Y. H., WU, Q., CHEW, J. L., VEGA, V. B., ZHANG, W., CHEN, X., BOURQUE, G., GEORGE, J., LEONG, B., LIU, J., WONG, K. Y., SUNG, K. W., LEE, C. W., ZHAO, X. D., CHIU, K. P., LIPOVICH, L., KUZNETSOV, V. A., ROBSON, P., STANTON, L. W., WEI, C. L., RUAN, Y., LIM, B. & NG, H. H. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet, 38, 431-40.

LOVRIC, J. 2011. Introducing proteomics : from concepts to sample separation, mass spectrometry and data analysis, Chichester, West Sussex ; Hoboken, NJ, Wiley-Blackwell.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature, 389, 251-60.

LUGO, T. G., PENDERGAST, A. M., MULLER, A. J. & WITTE, O. N. 1990. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. Science, 247, 1079-82.

LUGUS, J. J., PARK, C., MA, Y. D. & CHOI, K. 2009. Both primitive and definitive blood cells are derived from Flk-1+ mesoderm. Blood, 113, 563-6.

LUSCOMBE, N. M., AUSTIN, S. E., BERMAN, H. M. & THORNTON, J. M. 2000. An overview of the structures of protein-DNA complexes. Genome Biol, 1, REVIEWS001.

LYONS, A. B. & PARISH, C. R. 1994. Determination of lymphocyte division by flow cytometry. J Immunol Methods, 171, 131-7.

MACDONALD, N., WELBURN, J. P., NOBLE, M. E., NGUYEN, A., YAFFE, M. B., CLYNES, D., MOGGS, J. G., ORPHANIDES, G., THOMSON, S., EDMUNDS, J. W., CLAYTON, A. L., ENDICOTT, J. A. & MAHADEVAN, L. C. 2005. Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. Mol Cell, 20, 199-211.

MAGNUSSON, M., BRUN, A. C., MIYAKE, N., LARSSON, J., EHINGER, M., BJORNSSON, J. M., WUTZ, A., SIGVARDSSON, M. & KARLSSON, S. 2007. HOXA10 is a critical regulator for hematopoietic stem cells and erythroid/megakaryocyte development. Blood, 109, 3687-96.

MAGUER-SATTA, V., BURL, S., LIU, L., DAMEN, J., CHAHINE, H., KRYSTAL, G., EAVES, A. & EAVES, C. 1998. BCR-ABL accelerates C2-ceramide-induced apoptosis. Oncogene, 16, 237-48.

MAJETI, R., PARK, C. Y. & WEISSMAN, I. L. 2007. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. Cell Stem Cell, 1, 635-45.

MAJEWSKI, I. J., BLEWITT, M. E., DE GRAAF, C. A., MCMANUS, E. J., BAHLO, M., HILTON, A. A., HYLAND, C. D., SMYTH, G. K., CORBIN, J. E., METCALF, D., ALEXANDER, W. S. & HILTON, D. J. 2008. Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. PLoS Biol, 6, e93.

MAJEWSKI, I. J., RITCHIE, M. E., PHIPSON, B., CORBIN, J., PAKUSCH, M., EBERT, A., BUSSLINGER, M., KOSEKI, H., HU, Y., SMYTH, G. K., ALEXANDER, W. S., HILTON, D. J. & BLEWITT, M. E. 2010. Opposing roles of polycomb repressive complexes in hematopoietic stem and progenitor cells. Blood, 116, 731-9.

MALI, R. S., RAMDAS, B., MA, P., SHI, J., MUNUGALAVADLA, V., SIMS, E., WEI, L., VEMULA, S., NABINGER, S. C., GOODWIN, C. B., CHAN, R. J., TRAINA, F., VISCONTE, V., TIU, R. V., LEWIS, T. A., STERN, A. M., WEN, Q., CRISPINO, J. D., BOSWELL, H. S. & KAPUR, R. 2011. Rho kinase regulates the survival and transformation of cells bearing oncogenic forms of KIT, FLT3, and BCR-ABL. Cancer Cell, 20, 357-69.

MANCINI, S. J., MANTEI, N., DUMORTIER, A., SUTER, U., MACDONALD, H. R. & RADTKE, F. 2005. Jagged1-dependent Notch signaling is dispensable for hematopoietic stem cell self-renewal and differentiation. Blood, 105, 2340-2.

MARGUERON, R., LI, G., SARMA, K., BLAIS, A., ZAVADIL, J., WOODCOCK, C. L., DYNLACHT, B. D. & REINBERG, D. 2008. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol Cell, 32, 503-18.

MARINISSEN, M. J., CHIARIELLO, M. & GUTKIND, J. S. 2001. Regulation of gene expression by the small GTPase Rho through the ERK6 (p38 gamma) MAP kinase pathway. Genes Dev, 15, 535-53.

MARMORSTEIN, R. 2001. Protein modules that manipulate histone tails for chromatin regulation. Nat Rev Mol Cell Biol, 2, 422-32.

MARTENS, J. A. & WINSTON, F. 2003. Recent advances in understanding chromatin remodeling by Swi/Snf complexes. Curr Opin Genet Dev, 13, 136-42.

MASTON, G. A., EVANS, S. K. & GREEN, M. R. 2006. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet, 7, 29-59.

MATSUMOTO, A., TAKEISHI, S., KANIE, T., SUSAKI, E., ONOYAMA, I., TATEISHI, Y., NAKAYAMA, K. & NAKAYAMA, K. I. 2011. p57 is required for quiescence and maintenance of adult hematopoietic stem cells. Cell Stem Cell, 9, 262-71.

MCGALL, G. H. & CHRISTIANS, F. C. 2002. High-density genechip oligonucleotide probe arrays. Adv Biochem Eng Biotechnol, 77, 21-42.

MCGARVEY, K. M., FAHRNER, J. A., GREENE, E., MARTENS, J., JENUWEIN, T. & BAYLIN, S. B. 2006. Silenced tumor suppressor genes reactivated by DNA demethylation do not return to a fully euchromatic chromatin state. Cancer Res, 66, 3541-9.

MCGHEE, J. D., WOOD, W. I., DOLAN, M., ENGEL, J. D. & FELSENFELD, G. 1981. A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. Cell, 27, 45-55.

MCGINTY, R. K., KIM, J., CHATTERJEE, C., ROEDER, R. G. & MUIR, T. W. 2008. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. Nature, 453, 812-6.

MCMAHON, K. A., HIEW, S. Y., HADJUR, S., VEIGA-FERNANDES, H., MENZEL, U., PRICE, A. J., KIOUSSIS, D., WILLIAMS, O. & BRADY, H. J. 2007. Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal. Cell Stem Cell, 1, 338-45.

MCWEENEY, S. K., PEMBERTON, L. C., LORIAUX, M. M., VARTANIAN, K., WILLIS, S. G., YOCHUM, G., WILMOT, B., TURPAZ, Y., PILLAI, R., DRUKER, B. J., SNEAD, J. L., MACPARTLIN, M., O'BRIEN, S. G., MELO, J. V., LANGE, T., HARRINGTON, C. A. & DEININGER, M. W. 2010. A gene expression signature of CD34+ cells to predict major cytogenetic response in chronic-phase chronic myeloid leukemia patients treated with imatinib. Blood, 115, 315-25.

MEISSNER, A., MIKKELSEN, T. S., GU, H., WERNIG, M., HANNA, J., SIVACHENKO, A., ZHANG, X., BERNSTEIN, B. E., NUSBAUM, C., JAFFE, D. B., GNIRKE, A., JAENISCH, R. &

LANDER, E. S. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature, 454, 766-70.

MELO, J. V. 1996. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. Blood, 88, 2375-84.

MENDEZ-FERRER, S., MICHURINA, T. V., FERRARO, F., MAZLOOM, A. R., MACARTHUR, B. D., LIRA, S. A., SCADDEN, D. T., MA'AYAN, A., ENIKOLOPOV, G. N. & FRENETTE, P. S. 2010. Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. Nature, 466, 829-34.

METZKER, M. L. 2010. Sequencing technologies - the next generation. Nat Rev Genet, 11, 31-46.

MI, H., DONG, Q., MURUGANUJAN, A., GAUDET, P., LEWIS, S. & THOMAS, P. D. 2010. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res, 38, D204-10.

MIDDLETON, M. K., ZUKAS, A. M., RUBINSTEIN, T., JACOB, M., ZHU, P., ZHAO, L., BLAIR, I. & PURE, E. 2006. Identification of 12/15-lipoxygenase as a suppressor of myeloproliferative disease. J Exp Med, 203, 2529-40.

MIGNINI, F., STRECCIONI, V. & AMENTA, F. 2003. Autonomic innervation of immune organs and neuroimmune modulation. Auton Autacoid Pharmacol, 23, 1-25.

MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T. K., KOCHE, R. P., LEE, W., MENDENHALL, E., O'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. & BERNSTEIN, B. E. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature, 448, 553-60.

MILLER, C. L. & EAVES, C. J. 1997. Expansion in vitro of adult murine hematopoietic stem cells with transplantable lympho-myeloid reconstituting ability. Proc Natl Acad Sci U S A, 94, 13648-53.

MILNER, L. A., KOPAN, R., MARTIN, D. I. & BERNSTEIN, I. D. 1994. A human homologue of the Drosophila developmental gene, Notch, is expressed in CD34+ hematopoietic precursors. Blood, 83, 2057-62.

MIN, J., ZHANG, Y. & XU, R. M. 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev, 17, 1823-8.

MINOCHE, A. E., DOHM, J. C. & HIMMELBAUER, H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. Genome Biol, 12, R112.

MINSKY, N., SHEMA, E., FIELD, Y., SCHUSTER, M., SEGAL, E. & OREN, M. 2008. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nat Cell Biol, 10, 483-8.

MOCHIZUKI-KASHIO, M., MISHIMA, Y., MIYAGI, S., NEGISHI, M., SARAYA, A., KONUMA, T., SHINGA, J., KOSEKI, H. & IWAMA, A. 2011. Dependency on the polycomb gene Ezh2 distinguishes fetal from adult hematopoietic stem cells. Blood, 118, 6553-61.

MOHN, F., WEBER, M., REBHAN, M., ROLOFF, T. C., RICHTER, J., STADLER, M. B., BIBEL, M. & SCHUBELER, D. 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell, 30, 755-66.

MOHTY, M., YONG, A. S., SZYDLO, R. M., APPERLEY, J. F. & MELO, J. V. 2007. The polycomb group BMI1 gene is a molecular marker for predicting prognosis of chronic myeloid leukemia. Blood, 110, 380-3.

MORGAN, X. C., NI, S., MIRANKER, D. P. & IYER, V. R. 2007. Predicting combinatorial binding of transcription factors to regulatory elements in the human genome by association rule mining. BMC Bioinformatics, 8, 445.

MORIN, R. D., JOHNSON, N. A., SEVERSON, T. M., MUNGALL, A. J., AN, J., GOYA, R., PAUL, J. E., BOYLE, M., WOOLCOCK, B. W., KUCHENBAUER, F., YAP, D., HUMPHRIES, R. K., GRIFFITH, O. L., SHAH, S., ZHU, H., KIMBARA, M., SHASHKIN, P., CHARLOT, J. F., et al. 2010. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet, 42, 181-5.

MORRISON, S. J., WANDYCZ, A. M., HEMMATI, H. D., WRIGHT, D. E. & WEISSMAN, I. L. 1997. Identification of a lineage of multipotent hematopoietic progenitors. Development, 124, 1929-39.

MOSAMMAPARAST, N. & SHI, Y. 2010. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. Annu Rev Biochem, 79, 155-79.

MUGHAL, T. I. & GOLDMAN, J. M. 1999. Understanding leukaemia and related cancers, Osney Mead, Oxford ; Malden, MA, Blackwell Science Ltd.

MUJTABA, S., MANZUR, K. L., GURNON, J. R., KANG, M., VAN ETTEN, J. L. & ZHOU, M. M. 2008. Epigenetic transcriptional repression of cellular genes by a viral SET protein. Nat Cell Biol, 10, 1114-22.

MURRAY, L., CHEN, B., GALY, A., CHEN, S., TUSHINSKI, R., UCHIDA, N., NEGRIN, R., TRICOT, G., JAGANNATH, S., VESOLE, D. & ET AL. 1995. Enrichment of human hematopoietic stem cell activity in the CD34+Thy-1+Lin- subpopulation from mobilized peripheral blood. Blood, 85, 368-78.

MUTHU, K., IYER, S., HE, L. K., SZILAGYI, A., GAMELLI, R. L., SHANKAR, R. & JONES, S. B. 2007. Murine hematopoietic stem cells and progenitors express adrenergic receptors. J Neuroimmunol, 186, 27-36.

NAKA, K., HOSHII, T., MURAGUCHI, T., TADOKORO, Y., OOSHIO, T., KONDO, Y., NAKAO, S., MOTOYAMA, N. & HIRAO, A. 2010. TGF-beta-FOXO signalling maintains leukaemia-initiating cells in chronic myeloid leukaemia. Nature, 463, 676-80.

NAKANISHI, S., SANDERSON, B. W., DELVENTHAL, K. M., BRADFORD, W. D., STAEHLING-HAMPTON, K. & SHILATIFARD, A. 2008. A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. Nat Struct Mol Biol, 15, 881-8.

NAKAO, A., AFRAKHTE, M., MOREN, A., NAKAYAMA, T., CHRISTIAN, J. L., HEUCHEL, R., ITOH, S., KAWABATA, M., HELDIN, N. E., HELDIN, C. H. & TEN DIJKE, P. 1997. Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. Nature, 389, 631-5.

NARLIKAR, G. J., PHELAN, M. L. & KINGSTON, R. E. 2001. Generation and interconversion of multiple distinct nucleosomal states as a mechanism for catalyzing chromatin fluidity. Mol Cell, 8, 1219-30.

NATHAN, D., INGVARSDOTTIR, K., STERNER, D. E., BYLEBYL, G. R., DOKMANOVIC, M., DORSEY, J. A., WHELAN, K. A., KRSMANOVIC, M., LANE, W. S., MELUH, P. B., JOHNSON, E. S. & BERGER, S. L. 2006. Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. Genes Dev, 20, 966-76.

NEERING, S. J., BUSHNELL, T., SOZER, S., ASHTON, J., ROSSI, R. M., WANG, P. Y., BELL, D. R., HEINRICH, D., BOTTARO, A. & JORDAN, C. T. 2007. Leukemia stem cells in a genetically defined murine model of blast-crisis CML. Blood, 110, 2578-85.

NEKRASOV, M., KLYMENKO, T., FRATERMAN, S., PAPP, B., OKTABA, K., KOCHER, T., COHEN, A., STUNNENBERG, H. G., WILM, M. & MULLER, J. 2007. Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J, 26, 4078-88.

NEUMULLER, R. A. & KNOBLICH, J. A. 2009. Dividing cellular asymmetry: asymmetric cell division and its implications for stem cells and cancer. Genes Dev, 23, 2675-99.

NICKEL, B. E., ALLIS, C. D. & DAVIE, J. R. 1989. Ubiquitinated histone H2B is preferentially located in transcriptionally active chromatin. Biochemistry, 28, 958-63.

NILSSON, S. K., JOHNSTON, H. M., WHITTY, G. A., WILLIAMS, B., WEBB, R. J., DENHARDT, D. T., BERTONCELLO, I., BENDALL, L. J., SIMMONS, P. J. & HAYLOCK, D. N. 2005. Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. Blood, 106, 1232-9.

NIX, D. A., COURDY, S. J. & BOUCHER, K. M. 2008. Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. BMC Bioinformatics, 9, 523.

NORDON, R. E., GINSBERG, S. S. & EAVES, C. J. 1997. High-resolution cell division tracking demonstrates the FLt3-ligand-dependence of human marrow CD34+CD38- cell production in vitro. Br J Haematol, 98, 528-39.

NOSAKA, T., KAWASHIMA, T., MISAWA, K., IKUTA, K., MUI, A. L. & KITAMURA, T. 1999. STAT5 as a molecular regulator of proliferation, differentiation and apoptosis in hematopoietic cells. EMBO J, 18, 4754-65.

NOTTA, F., DOULATOV, S., LAURENTI, E., POEPPL, A., JURISICA, I. & DICK, J. E. 2011. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. Science, 333, 218-21.

NOVERSHTERN, N., SUBRAMANIAN, A., LAWTON, L. N., MAK, R. H., HAINING, W. N., MCCONKEY, M. E., HABIB, N., YOSEF, N., CHANG, C. Y., SHAY, T., FRAMPTON, G. M., DRAKE, A. C., LESKOV, I., NILSSON, B., PREFFER, F., DOMBKOWSKI, D., EVANS, J. W., LIEFELD, T., SMUTKO, J. S., CHEN, J., FRIEDMAN, N., YOUNG, R. A., GOLUB, T. R., REGEV, A. & EBERT, B. L. 2011. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell, 144, 296-309.

NOWAK, S. J. & CORCES, V. G. 2000. Phosphorylation of histone H3 correlates with transcriptionally active loci. Genes Dev, 14, 3003-13.

NOWAK, S. J. & CORCES, V. G. 2004. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. Trends Genet, 20, 214-20.

NTZIACHRISTOS, P., TSIRIGOS, A., VAN VLIERBERGHE, P., NEDJIC, J., TRIMARCHI, T., FLAHERTY, M. S., FERRES-MARCO, D., DA ROS, V., TANG, Z., SIEGLE, J., ASP, P., HADLER, M., RIGO, I., DE KEERSMAECKER, K., PATEL, J., HUYNH, T., UTRO, F., et al. 2012. Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nat Med, 18, 298-301.

OEHLER, V. G., YEUNG, K. Y., CHOI, Y. E., BUMGARNER, R. E., RAFTERY, A. E. & RADICH, J. P. 2009. The derivation of diagnostic markers of chronic myeloid leukemia progression from microarray data. Blood, 114, 3292-8.

OGBOURNE, S. & ANTALIS, T. M. 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem J, 331 (Pt 1), 1-14.

O'HARE, T., SHAKESPEARE, W. C., ZHU, X., EIDE, C. A., RIVERA, V. M., WANG, F., ADRIAN, L. T., ZHOU, T., HUANG, W. S., XU, Q., METCALF, C. A., 3RD, TYNER, J. W., LORIAUX, M. M., CORBIN, A. S., WARDWELL, S., NING, Y., KEATS, J. A., WANG, Y., et al. 2009. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. Cancer Cell, 16, 401-12.

OKANO, M., BELL, D. W., HABER, D. A. & LI, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell, 99, 247-57.

OMATSU, Y., SUGIYAMA, T., KOHARA, H., KONDOH, G., FUJII, N., KOHNO, K. & NAGASAWA, T. 2010. The essential functions of adipo-osteogenic progenitors as the hematopoietic stem and progenitor cell niche. Immunity, 33, 387-99.

O'NEILL, L. P. & TURNER, B. M. 1996. Immunoprecipitation of chromatin. Methods Enzymol, 274, 189-97.

O'NEILL, L. P. & TURNER, B. M. 2003. Immunoprecipitation of native chromatin: NChIP. Methods, 31, 76-82.

O'NEILL, L. P., VERMILYEA, M. D. & TURNER, B. M. 2006. Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. Nat Genet, 38, 835-41.

OOI, S. K., QIU, C., BERNSTEIN, E., LI, K., JIA, D., YANG, Z., ERDJUMENT-BROMAGE, H., TEMPST, P., LIN, S. P., ALLIS, C. D., CHENG, X. & BESTOR, T. H. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature, 448, 714-7.

ORFORD, K., KHARCHENKO, P., LAI, W., DAO, M. C., WORHUNSKY, D. J., FERRO, A., JANZEN, V., PARK, P. J. & SCADDEN, D. T. 2008. Differential H3K4 methylation identifies developmentally poised hematopoietic genes. Dev Cell, 14, 798-809.

ORKIN, S. H. 2000. Diversification of haematopoietic stem cells to specific lineages. Nat Rev Genet, 1, 57-64.

ORKIN, S. H. 2003. Priming the hematopoietic pump. Immunity, 19, 633-4.

ORKIN, S. H. & ZON, L. I. 2008. Hematopoiesis: an evolving paradigm for stem cell biology. Cell, 132, 631-44.

ORLANDO, V., STRUTT, H. & PARO, R. 1997. Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods, 11, 205-14.

OSAWA, M., HANADA, K., HAMADA, H. & NAKAUCHI, H. 1996. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science, 273, 242-5.

OZER, H., ARMITAGE, J. O., BENNETT, C. L., CRAWFORD, J., DEMETRI, G. D., PIZZO, P. A., SCHIFFER, C. A., SMITH, T. J., SOMLO, G., WADE, J. C., WADE, J. L., 3RD, WINN, R. J., WOZNIAK, A. J. & SOMERFIELD, M. R. 2000. 2000 update of recommendations for the use of hematopoietic colony-stimulating factors: evidence-based, clinical practice guidelines. American Society of Clinical Oncology Growth Factors Expert Panel. J Clin Oncol, 18, 3558-85.

PAGANO, M., TAM, S. W., THEODORAS, A. M., BEER-ROMERO, P., DEL SAL, G., CHAU, V., YEW, P. R., DRAETTA, G. F. & ROLFE, M. 1995. Role of the ubiquitin-proteasome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27. Science, 269, 682-5.

PAJCINI, K. V., SPECK, N. A. & PEAR, W. S. 2011. Notch signaling in mammalian hematopoietic stem cells. Leukemia, 25, 1525-32.

PAN, G., TIAN, S., NIE, J., YANG, C., RUOTTI, V., WEI, H., JONSDOTTIR, G. A., STEWART, R. & THOMSON, J. A. 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell, 1, 299-312.

PARK, I. K., HE, Y., LIN, F., LAERUM, O. D., TIAN, Q., BUMGARNER, R., KLUG, C. A., LI, K., KUHR, C., DOYLE, M. J., XIE, T., SCHUMMER, M., SUN, Y., GOLDSMITH, A., CLARKE, M. F., WEISSMAN, I. L., HOOD, L. & LI, L. 2002. Differential gene expression profiling of adult murine hematopoietic stem cells. Blood, 99, 488-98.

PARK, P. J. 2009. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet, 10, 669-80.

PARMAR, K., MAUCH, P., VERGILIO, J. A., SACKSTEIN, R. & DOWN, J. D. 2007. Distribution of hematopoietic stem cells in the bone marrow according to regional hypoxia. Proc Natl Acad Sci U S A, 104, 5431-6.

PASINI, D., BRACKEN, A. P., HANSEN, J. B., CAPILLO, M. & HELIN, K. 2007. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol, 27, 3769-79.

PASINI, D., BRACKEN, A. P., JENSEN, M. R., LAZZERINI DENCHI, E. & HELIN, K. 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO J, 23, 4061-71.

PASINI, D., CLOOS, P. A., WALFRIDSSON, J., OLSSON, L., BUKOWSKI, J. P., JOHANSEN, J. V., BAK, M., TOMMERUP, N., RAPPSILBER, J. & HELIN, K. 2010. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature, 464, 306-10.

PASINI, D., HANSEN, K. H., CHRISTENSEN, J., AGGER, K., CLOOS, P. A. & HELIN, K. 2008. Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes Dev, 22, 1345-55.

PAULER, F. M., SLOANE, M. A., HUANG, R., REGHA, K., KOERNER, M. V., TAMIR, I., SOMMER, A., ASZODI, A., JENUWEIN, T. & BARLOW, D. P. 2009. H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res, 19, 221-33.

PAVLIDIS, P., LI, Q. & NOBLE, W. S. 2003. The effect of replication on gene expression microarray experiments. Bioinformatics, 19, 1620-7.

PENG, C., ZOU, P., MA, Y. & HU, Z. 2005. Effect of angiotensin II on cord blood CD34+ cells expansion in vitro. J Huazhong Univ Sci Technolog Med Sci, 25, 26-8.

PEREIRA, A., MCLAREN, A., BELL, W. R., COPOLOV, D. & DEAN, B. 2003. Potential clozapine target sites on peripheral hematopoietic cells and stromal cells of the bone marrow. Pharmacogenomics J, 3, 227-34.

PERRY, S. S., ZHAO, Y., NIE, L., COCHRANE, S. W., HUANG, Z. & SUN, X. H. 2007. Id1, but not Id3, directs long-term repopulating hematopoietic stem-cell maintenance. Blood, 110, 2351-60.

PESTINA, T. I., CLEVELAND, J. L., YANG, C., ZAMBETTI, G. P. & JACKSON, C. W. 2001. Mpl ligand prevents lethal myelosuppression by inhibiting p53-dependent apoptosis. Blood, 98, 2084-90.

PETZER, A. L., EAVES, C. J., BARNETT, M. J. & EAVES, A. C. 1997. Selective expansion of primitive normal hematopoietic cells in cytokine-supplemented cultures of purified cells from patients with chronic myeloid leukemia. Blood, 90, 64-9.

PETZER, A. L., EAVES, C. J., LANSDORP, P. M., PONCHIO, L., BARNETT, M. J. & EAVES, A. C. 1996. Characterization of primitive subpopulations of normal and leukemic cells present in the blood of patients with newly diagnosed as well as established chronic myeloid leukemia. Blood, 88, 2162-71.

PIERCE, K. L., PREMONT, R. T. & LEFKOWITZ, R. J. 2002. Seven-transmembrane receptors. Nat Rev Mol Cell Biol, 3, 639-50.

POPP, C., DEAN, W., FENG, S., COKUS, S. J., ANDREWS, S., PELLEGRINI, M., JACOBSEN, S. E. & REIK, W. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature, 463, 1101-5.

PORRAS, O. H. & STUTZIN, A. 2011. Glutamate-induced metabolic changes influence the cytoplasmic redox state of hippocampal neurons. Biochem Biophys Res Commun, 411, 82-7.

PRIVALSKY, M. L. 2004. The role of corepressors in transcriptional regulation by nuclear hormone receptors. Annu Rev Physiol, 66, 315-60.

QUAIL, M. A., KOZAREWA, I., SMITH, F., SCALLY, A., STEPHENS, P. J., DURBIN, R., SWERDLOW, H. & TURNER, D. J. 2008. A large genome center's improvements to the Illumina sequencing system. Nat Methods, 5, 1005-10.

QUINTAS-CARDAMA, A. & CORTES, J. 2009. Molecular biology of bcr-abl1-positive chronic myeloid leukemia. Blood, 113, 1619-30.

RABINOVICH, A., JIN, V. X., RABINOVICH, R., XU, X. & FARNHAM, P. J. 2008. E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. Genome Res, 18, 1763-77.

RADICH, J. P., DAI, H., MAO, M., OEHLER, V., SCHELTER, J., DRUKER, B., SAWYERS, C., SHAH, N., STOCK, W., WILLMAN, C. L., FRIEND, S. & LINSLEY, P. S. 2006. Gene expression changes associated with progression and response in chronic myeloid leukemia. Proc Natl Acad Sci U S A, 103, 2794-9.

RADTKE, F., WILSON, A., STARK, G., BAUER, M., VAN MEERWIJK, J., MACDONALD, H. R. & AGUET, M. 1999. Deficient T cell fate specification in mice with an induced inactivation of Notch1. Immunity, 10, 547-58.

RAHL, P. B., LIN, C. Y., SEILA, A. C., FLYNN, R. A., MCCUINE, S., BURGE, C. B., SHARP, P. A. & YOUNG, R. A. 2010. c-Myc regulates transcriptional pause release. Cell, 141, 432-45.

RAJEEVAN, M. S., RANAMUKHAARACHCHI, D. G., VERNON, S. D. & UNGER, E. R. 2001. Use of real-time quantitative PCR to validate the results of cDNA array and differential display PCR technologies. Methods, 25, 443-51.

RAMALHO-SANTOS, M., YOON, S., MATSUZAKI, Y., MULLIGAN, R. C. & MELTON, D. A. 2002. "Stemness": transcriptional profiling of embryonic and adult stem cells. Science, 298, 597-600.

RAMARAJ, P., SINGH, H., NIU, N., CHU, S., HOLTZ, M., YEE, J. K. & BHATIA, R. 2004. Effect of mutational inactivation of tyrosine kinase activity on BCR/ABL-induced abnormalities in cell growth and adhesion in human hematopoietic progenitors. Cancer Res, 64, 5322-31.

RAMARAJ, P., SINGH, H., NIU, N., CHU, S., HOLTZ, M., YEE, J. K. & BHATIA, R. 2004. Effect of mutational inactivation of tyrosine kinase activity on BCR/ABL-induced abnormalities in cell growth and adhesion in human hematopoietic progenitors. Cancer Res, 64, 5322-31.

RECILLAS-TARGA, F., PIKAART, M. J., BURGESS-BEUSSE, B., BELL, A. C., LITT, M. D., WEST, A. G., GASZNER, M. & FELSENFELD, G. 2002. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc Natl Acad Sci U S A, 99, 6883-8.

REDDICONTO, G., TOTO, C., PALAMA, I., DE LEO, S., DE LUCA, E., DE MATTEIS, S., DINI, L., PASSERINI, C. G., DI RENZO, N., MAFFIA, M. & COLUCCIA, A. M. 2012. Targeting of GSK3beta promotes imatinib-mediated apoptosis in quiescent CD34+ chronic myeloid leukemia progenitors, preserving normal stem cells. Blood, 119, 2335-45.

REITER, E. & LEFKOWITZ, R. J. 2006. GRKs and beta-arrestins: roles in receptor silencing, trafficking and signaling. Trends Endocrinol Metab, 17, 159-65.

REMENYI, A., SCHOLER, H. R. & WILMANNS, M. 2004. Combinatorial control of gene expression. Nat Struct Mol Biol, 11, 812-5.

REN, B., ROBERT, F., WYRICK, J. J., APARICIO, O., JENNINGS, E. G., SIMON, I., ZEITLINGER, J., SCHREIBER, J., HANNETT, N., KANIN, E., VOLKERT, T. L., WILSON, C. J., BELL, S. P. & YOUNG, R. A. 2000. Genome-wide location and function of DNA binding proteins. Science, 290, 2306-9.

REN, R. 2005. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. Nat Rev Cancer, 5, 172-83.

REYA, T., DUNCAN, A. W., AILLES, L., DOMEN, J., SCHERER, D. C., WILLERT, K., HINTZ, L., NUSSE, R. & WEISSMAN, I. L. 2003. A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature, 423, 409-14.

RICHARDS, E. J. & ELGIN, S. C. 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell, 108, 489-500.

RIGGS, A. D. 1975. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet, 14, 9-25.

RINGROSE, L. 2007. Polycomb comes of age: genome-wide profiling of target sites. Curr Opin Cell Biol, 19, 290-7.

RINGROSE, L. & PARO, R. 2007. Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development, 134, 223-32.

RINN, J. L., EUSKIRCHEN, G., BERTONE, P., MARTONE, R., LUSCOMBE, N. M., HARTMAN, S., HARRISON, P. M., NELSON, F. K., MILLER, P., GERSTEIN, M., WEISSMAN, S. & SNYDER, M. 2003. The transcriptional activity of human Chromosome 22. Genes Dev, 17, 529-40.

RITTER, S. L. & HALL, R. A. 2009. Fine-tuning of GPCR activity by receptor-interacting proteins. Nat Rev Mol Cell Biol, 10, 819-30.

ROBERTSON, G., HIRST, M., BAINBRIDGE, M., BILENKY, M., ZHAO, Y., ZENG, T., EUSKIRCHEN, G., BERNIER, B., VARHOL, R., DELANEY, A., THIESSEN, N., GRIFFITH, O. L., HE, A., MARRA, M., SNYDER, M. & JONES, S. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods, 4, 651-7.

ROCKETT, J. C. & HELLMANN, G. M. 2004. Confirming microarray data--is it really necessary? Genomics, 83, 541-9.

RODRIGUEZ, J., MUNOZ, M., VIVES, L., FRANGOU, C. G., GROUDINE, M. & PEINADO, M. A. 2008. Bivalent domains enforce transcriptional memory of DNA methylated genes in cancer cells. Proc Natl Acad Sci U S A, 105, 19809-14.

ROSENBAUER, F. & TENEN, D. G. 2007. Transcription factors in myeloid development: balancing differentiation with transformation. Nat Rev Immunol, 7, 105-17.

ROTHENBERG, E. V. 2007. Cell lineage regulators in B and T cell development. Nat Immunol, 8, 441-4.

ROWLEY, J. D. 1973. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature, 243, 290-3.

ROZOWSKY, J., EUSKIRCHEN, G., AUERBACH, R. K., ZHANG, Z. D., GIBSON, T., BJORNSON, R., CARRIERO, N., SNYDER, M. & GERSTEIN, M. B. 2009. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol, 27, 66-75.

RUIZ I ALTABA, A. 1997. Catching a Gli-mpse of Hedgehog. Cell, 90, 193-6.

RUSCETTI, F. W., AKEL, S. & BARTELMEZ, S. H. 2005. Autocrine transforming growth factor-beta regulation of hematopoiesis: many outcomes that depend on the context. Oncogene, 24, 5751-63.

RUSSELL, S., MEADOWS, L. A. & RUSSELL, R. R. 2009. Microarray technology in practice, Elsevier.

RUSSO, V. E. A., MARTIENSSEN, R. A. & RIGGS, A. D. 1996. Epigenetic mechanisms of gene regulation, Plainview, N.Y., Cold Spring Harbor Laboratory Press.

RYBTSOV, S., SOBIESIAK, M., TAOUDI, S., SOUILHOL, C., SENSERRICH, J., LIAKHOVITSKAIA, A., IVANOVS, A., FRAMPTON, J., ZHAO, S. & MEDVINSKY, A. 2011. Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. J Exp Med, 208, 1305-15.

SAKAMOTO, K., YAMAGUCHI, S., ANDO, R., MIYAWAKI, A., KABASAWA, Y., TAKAGI, M., LI, C. L., PERBAL, B. & KATSUBE, K. 2002. The nephroblastoma overexpressed gene (NOV/ccn3) protein associates with Notch1 extracellular domain and inhibits myoblast differentiation via Notch signaling pathway. J Biol Chem, 277, 29399-405.

SANCHEZ, C., SANCHEZ, I., DEMMERS, J. A., RODRIGUEZ, P., STROUBOULIS, J. & VIDAL, M. 2007. Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. Mol Cell Proteomics, 6, 820-34.

SANTOS-ROSA, H., SCHNEIDER, R., BANNISTER, A. J., SHERRIFF, J., BERNSTEIN, B. E., EMRE, N. C., SCHREIBER, S. L., MELLOR, J. & KOUZARIDES, T. 2002. Active genes are tri-methylated at K4 of histone H3. Nature, 419, 407-11.

SARMA, K., MARGUERON, R., IVANOV, A., PIRROTTA, V. & REINBERG, D. 2008. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Mol Cell Biol, 28, 2718-31.

SATTLER, M., MOHI, M. G., PRIDE, Y. B., QUINNAN, L. R., MALOUF, N. A., PODAR, K., GESBERT, F., IWASAKI, H., LI, S., VAN ETTEN, R. A., GU, H., GRIFFIN, J. D. & NEEL, B. G. 2002. Critical role for Gab2 in transformation by BCR/ABL. Cancer Cell, 1, 479-92.

SAUVAGEAU, G., THORSTEINSDOTTIR, U., EAVES, C. J., LAWRENCE, H. J., LARGMAN, C., LANSDORP, P. M. & HUMPHRIES, R. K. 1995. Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo. Genes Dev, 9, 1753-65.

SCANDURA, J. M., BOCCUNI, P., MASSAGUE, J. & NIMER, S. D. 2004. Transforming growth factor beta-induced cell cycle arrest of human hematopoietic cells requires p57KIP2 up-regulation. Proc Natl Acad Sci U S A, 101, 15231-6.

SCHILTZ, R. L., MIZZEN, C. A., VASSILEV, A., COOK, R. G., ALLIS, C. D. & NAKATANI, Y. 1999. Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates. J Biol Chem, 274, 1189-92.

SCHLESINGER, Y., STRAUSSMAN, R., KESHET, I., FARKASH, S., HECHT, M., ZIMMERMAN, J., EDEN, E., YAKHINI, Z., BEN-SHUSHAN, E., REUBINOFF, B. E., BERGMAN, Y., SIMON, I. & CEDAR, H. 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet, 39, 232-6.

SCHMITZ, S. U., ALBERT, M., MALATESTA, M., MOREY, L., JOHANSEN, J. V., BAK, M., TOMMERUP, N., ABARRATEGUI, I. & HELIN, K. 2011. Jarid1b targets genes regulating development and is involved in neural differentiation. EMBO J, 30, 4586-600.

SCHOEFTNER, S., SENGUPTA, A. K., KUBICEK, S., MECHTLER, K., SPAHN, L., KOSEKI, H., JENUWEIN, T. & WUTZ, A. 2006. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J, 25, 3110-22.

SCHWARTZ, Y. B. & PIRROTTA, V. 2007. Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet, 8, 9-22.

SELIGSON, D. B., HORVATH, S., MCBRIAN, M. A., MAH, V., YU, H., TZE, S., WANG, Q., CHIA, D., GOODGLICK, L. & KURDISTANI, S. K. 2009. Global levels of histone modifications predict prognosis in different cancers. Am J Pathol, 174, 1619-28.

SHAH, N. P., TRAN, C., LEE, F. Y., CHEN, P., NORRIS, D. & SAWYERS, C. L. 2004. Overriding imatinib resistance with a novel ABL kinase inhibitor. Science, 305, 399-401.

SHEN, X., LIU, Y., HSU, Y. J., FUJIWARA, Y., KIM, J., MAO, X., YUAN, G. C. & ORKIN, S. H. 2008. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell, 32, 491-502.

SHEPHERD, B. E., GUTTORP, P., LANSDORP, P. M. & ABKOWITZ, J. L. 2004. Estimating human hematopoietic stem cell kinetics using granulocyte telomere lengths. Exp Hematol, 32, 1040-50.

SHERR, C. J. & ROBERTS, J. M. 1995. Inhibitors of mammalian G1 cyclin-dependent kinases. Genes Dev, 9, 1149-63.

SHI, Y., LAN, F., MATSON, C., MULLIGAN, P., WHETSTINE, J. R., COLE, P. A. & CASERO, R. A. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell, 119, 941-53.

SHIIO, Y. & EISENMAN, R. N. 2003. Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A, 100, 13225-30.

SHILATIFARD, A. 2008. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Curr Opin Cell Biol, 20, 341-8.

SHOJAEI, F., TROWBRIDGE, J., GALLACHER, L., YUEFEI, L., GOODALE, D., KARANU, F., LEVAC, K. & BHATIA, M. 2005. Hierarchical and ontogenic positions serve to define the molecular basis of human hematopoietic stem cell behavior. Dev Cell, 8, 651-63.

SHULTZ, L. D., LYONS, B. L., BURZENSKI, L. M., GOTT, B., CHEN, X., CHALEFF, S., KOTB, M., GILLIES, S. D., KING, M., MANGADA, J., GREINER, D. L. & HANDGRETINGER, R. 2005. Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. J Immunol, 174, 6477-89.

SHULTZ, L. D., SCHWEITZER, P. A., CHRISTIANSON, S. W., GOTT, B., SCHWEITZER, I. B., TENNENT, B., MCKENNA, S., MOBRAATEN, L., RAJAN, T. V., GREINER, D. L. & ET AL. 1995. Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice. J Immunol, 154, 180-91.

SIMSEK, T., KOCABAS, F., ZHENG, J., DEBERARDINIS, R. J., MAHMOUD, A. I., OLSON, E. N., SCHNEIDER, J. W., ZHANG, C. C. & SADEK, H. A. 2010. The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. Cell Stem Cell, 7, 380-90.

SINGBRANT, S., MOODY, J. L., BLANK, U., KARLSSON, G., UMANS, L., ZWIJSEN, A. & KARLSSON, S. 2006. Smad5 is dispensable for adult murine hematopoiesis. Blood, 108, 3707-12.

SITNICKA, E., BUZA-VIDAS, N., LARSSON, S., NYGREN, J. M., LIUBA, K. & JACOBSEN, S. E. 2003. Human CD34+ hematopoietic stem cells capable of multilineage engrafting NOD/SCID mice express flt3: distinct flt3 and c-kit expression and response patterns on mouse and candidate human hematopoietic stem cells. Blood, 102, 881-6.

SITNICKA, E., RUSCETTI, F. W., PRIESTLEY, G. V., WOLF, N. S. & BARTELMEZ, S. H. 1996. Transforming growth factor beta 1 directly and reversibly inhibits the initial cell divisions of long-term repopulating hematopoietic stem cells. Blood, 88, 82-8.

SKORSKI, T., KANAKARAJ, P., NIEBOROWSKA-SKORSKA, M., RATAJCZAK, M. Z., WEN, S. C., ZON, G., GEWIRTZ, A. M., PERUSSIA, B. & CALABRETTA, B. 1995. Phosphatidylinositol-3 kinase activity is regulated by BCR/ABL and is required for the growth of Philadelphia chromosome-positive cells. Blood, 86, 726-36.

SNEERINGER, C. J., SCOTT, M. P., KUNTZ, K. W., KNUTSON, S. K., POLLOCK, R. M., RICHON, V. M. & COPELAND, R. A. 2010. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proc Natl Acad Sci U S A, 107, 20980-5.

SOLAR, G. P., KERR, W. G., ZEIGLER, F. C., HESS, D., DONAHUE, C., DE SAUVAGE, F. J. & EATON, D. L. 1998. Role of c-mpl in early hematopoiesis. Blood, 92, 4-10.

SOLOMON, M. J. & VARSHAVSKY, A. 1985. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. Proc Natl Acad Sci U S A, 82, 6470-4.

SONG, J. J., GARLICK, J. D. & KINGSTON, R. E. 2008. Structural basis of histone H4 recognition by p55. Genes Dev, 22, 1313-8.

SPANGRUDE, G. J., HEIMFELD, S. & WEISSMAN, I. L. 1988. Purification and characterization of mouse hematopoietic stem cells. Science, 241, 58-62.

SPARMANN, A. & VAN LOHUIZEN, M. 2006. Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer, 6, 846-56.

SPENCER, V. A., SUN, J. M., LI, L. & DAVIE, J. R. 2003. Chromatin immunoprecipitation: a tool for studying histone acetylation and transcription factor binding. Methods, 31, 67-75.

SPIEGEL, A., SHIVTIEL, S., KALINKOVICH, A., LUDIN, A., NETZER, N., GOICHBERG, P., AZARIA, Y., RESNICK, I., HARDAN, I., BEN-HUR, H., NAGLER, A., RUBINSTEIN, M. & LAPIDOT, T. 2007. Catecholaminergic neurotransmitters regulate migration and repopulation of immature human CD34+ cells through Wnt signaling. Nat Immunol, 8, 1123-31.

SPYROU, C., STARK, R., LYNCH, A. G. & TAVARE, S. 2009. BayesPeak: Bayesian analysis of ChIP-seq data. BMC Bioinformatics, 10, 299.

STALDER, J., LARSEN, A., ENGEL, J. D., DOLAN, M., GROUDINE, M. & WEINTRAUB, H. 1980. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell, 20, 451-60.

STEGER, D. J., LEFTEROVA, M. I., YING, L., STONESTROM, A. J., SCHUPP, M., ZHUO, D., VAKOC, A. L., KIM, J. E., CHEN, J., LAZAR, M. A., BLOBEL, G. A. & VAKOC, C. R. 2008. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol Cell Biol, 28, 2825-39.

STEIDL, U., BORK, S., SCHAUB, S., SELBACH, O., SERES, J., AIVADO, M., SCHROEDER, T., ROHR, U. P., FENK, R., KLISZEWSKI, S., MAERCKER, C., NEUBERT, P., BORNSTEIN, S. R., HAAS, H. L., KOBBE, G., TENEN, D. G., HAAS, R. & KRONENWETT, R. 2004. Primary human CD34+ hematopoietic stem and progenitor cells express functionally active receptors of neuromediators. Blood, 104, 81-8.

STEIDL, U., KRONENWETT, R., ROHR, U. P., FENK, R., KLISZEWSKI, S., MAERCKER, C., NEUBERT, P., AIVADO, M., KOCH, J., MODLICH, O., BOJAR, H., GATTERMANN, N. & HAAS, R. 2002. Gene expression profiling identifies significant differences between the molecular phenotypes of bone marrow-derived and circulating human CD34+ hematopoietic stem cells. Blood, 99, 2037-44.

STERNER, D. E. & BERGER, S. L. 2000. Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev, 64, 435-59.

STEWART, M. J., LITZ-JACKSON, S., BURGESS, G. S., WILLIAMSON, E. A., LEIBOWITZ, D. S. & BOSWELL, H. S. 1995. Role for E2F1 in p210 BCR-ABL downstream regulation of c-myc transcription initiation. Studies in murine myeloid cells. Leukemia, 9, 1499-507.

STIER, S., CHENG, T., FORKERT, R., LUTZ, C., DOMBKOWSKI, D. M., ZHANG, J. L. & SCADDEN, D. T. 2003. Ex vivo targeting of p21Cip1/Waf1 permits relative expansion of human hematopoietic stem cells. Blood, 102, 1260-6.

STOCK, J. K., GIADROSSI, S., CASANOVA, M., BROOKES, E., VIDAL, M., KOSEKI, H., BROCKDORFF, N., FISHER, A. G. & POMBO, A. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol, 9, 1428-35.

STOREY, J. D. & TIBSHIRANI, R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A, 100, 9440-5.

STRIFE, A., LAMBEK, C., WISNIEWSKI, D., WACHTER, M., GULATI, S. C. & CLARKSON, B. D. 1988. Discordant maturation as the primary biological defect in chronic myelogenous leukemia. Cancer Res, 48, 1035-41.

STRUHL, G. 1981. A gene product required for correct initiation of segmental determination in Drosophila. Nature, 293, 36-41.

SUKA, N., SUKA, Y., CARMEN, A. A., WU, J. & GRUNSTEIN, M. 2001. Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. Mol Cell, 8, 473-9.

SUTHERLAND, H. J., EAVES, C. J., EAVES, A. C., DRAGOWSKA, W. & LANSDORP, P. M. 1989. Characterization and partial purification of human marrow cells capable of initiating long-term hematopoiesis in vitro. Blood, 74, 1563-70.

SUTHERLAND, H. J., EAVES, C. J., LANSDORP, P. M., THACKER, J. D. & HOGGE, D. E. 1991. Differential regulation of primitive human hematopoietic cells in long-term cultures maintained on genetically engineered murine stromal cells. Blood, 78, 666-72.

SZUKIEWICZ, D., SZEWCZYK, G., PYZLAK, M., STANGRET, A. & MASLINSKA, D. 2009. Overexpression of histamine H(1)-receptor by human amniotic epithelial cells in chorioamnionitis correlates with augmented production of secretory leukocyte protease inhibitor. Inflamm Res, 58 Suppl 1, 57-8.

TABE, Y., JIN, L., TSUTSUMI-ISHII, Y., XU, Y., MCQUEEN, T., PRIEBE, W., MILLS, G. B., OHSAKA, A., NAGAOKA, I., ANDREEFF, M. & KONOPLEVA, M. 2007. Activation of integrin-linked kinase is a critical prosurvival pathway induced in leukemic cells by bone marrow-derived stromal cells. Cancer Res, 67, 684-94.

TABERLAY, P. C., KELLY, T. K., LIU, C. C., YOU, J. S., DE CARVALHO, D. D., MIRANDA, T. B., ZHOU, X. J., LIANG, G. & JONES, P. A. 2011. Polycomb-repressed genes have permissive enhancers that initiate reprogramming. Cell, 147, 1283-94.

TAICHMAN, R. S., REILLY, M. J. & EMERSON, S. G. 1996. Human osteoblasts support human hematopoietic progenitor cells in vitro bone marrow cultures. Blood, 87, 518-24.

TAKUBO, K., GODA, N., YAMADA, W., IRIUCHISHIMA, H., IKEDA, E., KUBOTA, Y., SHIMA, H., JOHNSON, R. S., HIRAO, A., SUEMATSU, M. & SUDA, T. 2010. Regulation of the HIF-1alpha level is essential for hematopoietic stem cells. Cell Stem Cell, 7, 391-402.

TAN, J., YANG, X., ZHUANG, L., JIANG, X., CHEN, W., LEE, P. L., KARUTURI, R. K., TAN, P. B., LIU, E. T. & YU, Q. 2007. Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. Genes Dev, 21, 1050-63.

TAN, M., LUO, H., LEE, S., JIN, F., YANG, J. S., MONTELLIER, E., BUCHOU, T., CHENG, Z., ROUSSEAUX, S., RAJAGOPAL, N., LU, Z., YE, Z., ZHU, Q., WYSOCKA, J., YE, Y., KHOCHBIN, S., REN, B. & ZHAO, Y. 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell, 146, 1016-28.

TANAKA, S., MIYAGI, S., SASHIDA, G., CHIBA, T., YUAN, J., MOCHIZUKI-KASHIO, M., SUZUKI, Y., SUGANO, S., NAKASEKO, C., YOKOTE, K., KOSEKI, H. & IWAMA, A. 2012. Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. Blood, 120, 1107-17.

TAVERNA, S. D., LI, H., RUTHENBURG, A. J., ALLIS, C. D. & PATEL, D. J. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol, 14, 1025-40.

THOMAS, E. K., CANCELAS, J. A., CHAE, H. D., COX, A. D., KELLER, P. J., PERROTTI, D., NEVIANI, P., DRUKER, B. J., SETCHELL, K. D., ZHENG, Y., HARRIS, C. E. & WILLIAMS, D. A. 2007. Rac guanosine triphosphatases represent integrating molecular therapeutic targets for BCR-ABL-induced myeloproliferative disease. Cancer Cell, 12, 467-78.

THOMAS, P. D., KEJARIWAL, A., CAMPBELL, M. J., MI, H., DIEMER, K., GUO, N., LADUNGA, I., ULITSKY-LAZAREVA, B., MURUGANUJAN, A., RABKIN, S., VANDERGRIFF, J. A. & DOREMIEUX, O. 2003. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res, 31, 334-41.

THOMSON, J. P., SKENE, P. J., SELFRIDGE, J., CLOUAIRE, T., GUY, J., WEBB, S., KERR, A. R., DEATON, A., ANDREWS, R., JAMES, K. D., TURNER, D. J., ILLINGWORTH, R. & BIRD, A. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature, 464, 1082-6.

THURMAN, R. E., RYNES, E., HUMBERT, R., VIERSTRA, J., MAURANO, M. T., HAUGEN, E., SHEFFIELD, N. C., STERGACHIS, A. B., WANG, H., VERNOT, B., GARG, K., JOHN, S., SANDSTROM, R., BATES, D., BOATMAN, L., CANFIELD, T. K., DIEGEL, M., DUNN, D., et al. 2012. The accessible chromatin landscape of the human genome. Nature, 489, 75-82.

TILL, J. E. & MC, C. E. 1961. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat Res, 14, 213-22.

TIWARI, V. K., MCGARVEY, K. M., LICCHESI, J. D., OHM, J. E., HERMAN, J. G., SCHUBELER, D. & BAYLIN, S. B. 2008. PcG proteins, DNA methylation, and gene repression by chromatin looping. PLoS Biol, 6, 2911-27.

TOLHUIS, B., PALSTRA, R. J., SPLINTER, E., GROSVELD, F. & DE LAAT, W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell, 10, 1453-65.

TOTHOVA, Z., KOLLIPARA, R., HUNTLY, B. J., LEE, B. H., CASTRILLON, D. H., CULLEN, D. E., MCDOWELL, E. P., LAZO-KALLANIAN, S., WILLIAMS, I. R., SEARS, C., ARMSTRONG, S. A., PASSEGUE, E., DEPINHO, R. A. & GILLILAND, D. G. 2007. FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. Cell, 128, 325-39.

TOUGH, I. M., JACOBS, P. A., COURT BROWN, W. M., BAIKIE, A. G. & WILLIAMSON, E. R. 1963. Cytogenetic studies on bone-marrow in chronic myeloid leukaemia. Lancet, 1, 844-6.

TRAYCOFF, C. M., HALSTEAD, B., RICE, S., MCMAHEL, J., SROUR, E. F. & CORNETTA, K. 1998. Chronic myelogenous leukaemia CD34+ cells exit G0/G1 phases of cell cycle more rapidly than normal marrow CD34+ cells. Br J Haematol, 102, 759-67.

TROWBRIDGE, J. J., SNOW, J. W., KIM, J. & ORKIN, S. H. 2009. DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell, 5, 442-9.

TROWBRIDGE, J. J., XENOCOSTAS, A., MOON, R. T. & BHATIA, M. 2006. Glycogen synthase kinase-3 is an in vivo regulator of hematopoietic stem cell repopulation. Nat Med, 12, 89-98.

TSAI, C. A., HSUEH, H. M. & CHEN, J. J. 2003. Estimation of false discovery rates in multiple testing: application to gene microarray data. Biometrics, 59, 1071-81.

TSCHIERSCH, B., HOFMANN, A., KRAUSS, V., DORN, R., KORGE, G. & REUTER, G. 1994. The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J, 13, 3822-31.

TSUKADA, Y., FANG, J., ERDJUMENT-BROMAGE, H., WARREN, M. E., BORCHERS, C. H., TEMPST, P. & ZHANG, Y. 2006. Histone demethylation by a family of JmjC domain-containing proteins. Nature, 439, 811-6.

TSUKIYAMA, T. & WU, C. 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell, 83, 1011-20.

TSUMURA, A., HAYAKAWA, T., KUMAKI, Y., TAKEBAYASHI, S., SAKAUE, M., MATSUOKA, C., SHIMOTOHNO, K., ISHIKAWA, F., LI, E., UEDA, H. R., NAKAYAMA, J. & OKANO, M. 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells, 11, 805-14.

TURNER, B. M. 1993. Decoding the nucleosome. Cell, 75, 5-8.

TZENG, Y. S., LI, H., KANG, Y. L., CHEN, W. C., CHENG, W. C. & LAI, D. M. 2011. Loss of Cxcl12/Sdf-1 in adult mice decreases the quiescent state of hematopoietic stem/progenitor cells and alters the pattern of hematopoietic regeneration after myelosuppression. Blood, 117, 429-39.

UDOMSAKDI, C., EAVES, C. J., SWOLIN, B., REID, D. S., BARNETT, M. J. & EAVES, A. C. 1992. Rapid decline of chronic myeloid leukemic cells in long-term culture due to a defect at the leukemic stem cell level. Proc Natl Acad Sci U S A, 89, 6192-6.

URBAN, J. D., CLARKE, W. P., VON ZASTROW, M., NICHOLS, D. E., KOBILKA, B., WEINSTEIN, H., JAVITCH, J. A., ROTH, B. L., CHRISTOPOULOS, A., SEXTON, P. M., MILLER, K. J., SPEDDING, M. & MAILMAN, R. B. 2007. Functional selectivity and classical concepts of quantitative pharmacology. J Pharmacol Exp Ther, 320, 1-13.

VAKOC, C. R., MANDAT, S. A., OLENCHOCK, B. A. & BLOBEL, G. A. 2005. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell, 19, 381-91.

VALOUEV, A., JOHNSON, D. S., SUNDQUIST, A., MEDINA, C., ANTON, E., BATZOGLOU, S., MYERS, R. M. & SIDOW, A. 2008. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods, 5, 829-34.

VAQUERIZAS, J. M., KUMMERFELD, S. K., TEICHMANN, S. A. & LUSCOMBE, N. M. 2009. A census of human transcription factors: function, expression and evolution. Nat Rev Genet, 10, 252-63.

VARAMBALLY, S., CAO, Q., MANI, R. S., SHANKAR, S., WANG, X., ATEEQ, B., LAXMAN, B., CAO, X., JING, X., RAMNARAYANAN, K., BRENNER, J. C., YU, J., KIM, J. H., HAN, B., TAN, P., KUMAR-SINHA, C., LONIGRO, R. J., PALANISAMY, N., MAHER, C. A. & CHINNAIYAN, A. M. 2008. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science, 322, 1695-9.

VENOLIA, L. & GARTLER, S. M. 1983. Comparison of transformation efficiency of human active and inactive X-chromosomal DNA. Nature, 302, 82-3.

VERMEULEN, M., MULDER, K. W., DENISSOV, S., PIJNAPPEL, W. W., VAN SCHAIK, F. M., VARIER, R. A., BALTISSEN, M. P., STUNNENBERG, H. G., MANN, M. & TIMMERS, H. T. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell, 131, 58-69.

VILAR, J. M. & SAIZ, L. 2005. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Curr Opin Genet Dev, 15, 136-44.

VIRE, E., BRENNER, C., DEPLUS, R., BLANCHON, L., FRAGA, M., DIDELOT, C., MOREY, L., VAN EYNDE, A., BERNARD, D., VANDERWINDEN, J. M., BOLLEN, M., ESTELLER, M., DI CROCE, L., DE LAUNOIT, Y. & FUKS, F. 2006. The Polycomb group protein EZH2 directly controls DNA methylation. Nature, 439, 871-4.

VISNJIC, D., KALAJZIC, Z., ROWE, D. W., KATAVIC, V., LORENZO, J. & AGUILA, H. L. 2004. Hematopoiesis is severely altered in mice with an induced osteoblast deficiency. Blood, 103, 3258-64.

VODYANIK, M. A., BORK, J. A., THOMSON, J. A. & SLUKVIN, II 2005. Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. Blood, 105, 617-26.

VOGEL, W., SCHEDING, S., KANZ, L. & BRUGGER, W. 2000. Clinical applications of CD34(+) peripheral blood progenitor cells (PBPC). Stem Cells, 18, 87-92.

WALSH, C. P., CHAILLET, J. R. & BESTOR, T. H. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet, 20, 116-7.

WALSH, C. P. & XU, G. L. 2006. Cytosine methylation and DNA repair. Curr Top Microbiol Immunol, 301, 283-315.

WANG, H., WANG, L., ERDJUMENT-BROMAGE, H., VIDAL, M., TEMPST, P., JONES, R. S. & ZHANG, Y. 2004. Role of histone H2A ubiquitination in Polycomb silencing. Nature, 431, 873-8.

WANG, J. C., DOEDENS, M. & DICK, J. E. 1997. Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay. Blood, 89, 3919-24.

WANG, L. D. & WAGERS, A. J. 2011. Dynamic niches in the origination and differentiation of haematopoietic stem cells. Nat Rev Mol Cell Biol, 12, 643-55.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, 10, 57-63.

WANG, Z., ZANG, C., ROSENFELD, J. A., SCHONES, D. E., BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., PENG, W., ZHANG, M. Q. & ZHAO, K. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet, 40, 897-903.

WEBER, M., DAVIES, J. J., WITTIG, D., OAKELEY, E. J., HAASE, M., LAM, W. L. & SCHUBELER, D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet, 37, 853-62.

WEI, Y., XIA, W., ZHANG, Z., LIU, J., WANG, H., ADSAY, N. V., ALBARRACIN, C., YU, D., ABBRUZZESE, J. L., MILLS, G. B., BAST, R. C., JR., HORTOBAGYI, G. N. & HUNG, M. C. 2008. Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. Mol Carcinog, 47, 701-6.

WEINMANN, A. S. & FARNHAM, P. J. 2002. Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. Methods, 26, 37-47.

WEISBERG, E., MANLEY, P. W., BREITENSTEIN, W., BRUGGEN, J., COWAN-JACOB, S. W., RAY, A., HUNTLY, B., FABBRO, D., FENDRICH, G., HALL-MEYERS, E., KUNG, A. L., MESTAN, J., DALEY, G. Q., CALLAHAN, L., CATLEY, L., CAVAZZA, C., AZAM, M., NEUBERG, D., WRIGHT, R. D., GILLILAND, D. G. & GRIFFIN, J. D. 2005. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. Cancer Cell, 7, 129-41.

WEISHAUPT, H. & ATTEMA, J. L. 2010. A Method to Study the Epigenetic Chromatin States of Rare Hematopoietic Stem and Progenitor Cells; MiniChIP-Chip. Biol Proced Online, 12, 1-17.

WEISHAUPT, H., SIGVARDSSON, M. & ATTEMA, J. L. 2010. Epigenetic chromatin states uniquely define the developmental plasticity of murine hematopoietic stem cells. Blood, 115, 247-56.

WELLS, J. & FARNHAM, P. J. 2002. Characterizing transcription factor binding sites using formaldehyde crosslinking and immunoprecipitation. Methods, 26, 48-56.

WEST, M. H. & BONNER, W. M. 1980. Histone 2B can be modified by the attachment of ubiquitin. Nucleic Acids Res, 8, 4671-80.

WHANG, J., FREI, E., 3RD, TJIO, J. H., CARBONE, P. P. & BRECHER, G. 1963. The Distribution of the Philadelphia Chromosome in Patients with Chronic Myelogenous Leukemia. Blood, 22, 664-73.

WHETSTINE, J. R., NOTTKE, A., LAN, F., HUARTE, M., SMOLIKOV, S., CHEN, Z., SPOONER, E., LI, E., ZHANG, G., COLAIACOVO, M. & SHI, Y. 2006. Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell, 125, 467-81.

WHITEHOUSE, I., FLAUS, A., CAIRNS, B. R., WHITE, M. F., WORKMAN, J. L. & OWEN-HUGHES, T. 1999. Nucleosome mobilization catalysed by the yeast SWI/SNF complex. Nature, 400, 784-7.

WIDSCHWENDTER, M., FIEGL, H., EGLE, D., MUELLER-HOLZNER, E., SPIZZO, G., MARTH, C., WEISENBERGER, D. J., CAMPAN, M., YOUNG, J., JACOBS, I. & LAIRD, P. W. 2007. Epigenetic stem cell signature in cancer. Nat Genet, 39, 157-8.

WILLERT, K., BROWN, J. D., DANENBERG, E., DUNCAN, A. W., WEISSMAN, I. L., REYA, T., YATES, J. R., 3RD & NUSSE, R. 2003. Wnt proteins are lipid-modified and can act as stem cell growth factors. Nature, 423, 448-52.

WILSON, A., LAURENTI, E., OSER, G., VAN DER WATH, R. C., BLANCO-BOSE, W., JAWORSKI, M., OFFNER, S., DUNANT, C. F., ESHKIND, L., BOCKAMP, E., LIO, P., MACDONALD, H. R. & TRUMPP, A. 2008. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell, 135, 1118-29.

WINKLER, I. G., BARBIER, V., WADLEY, R., ZANNETTINO, A. C., WILLIAMS, S. & LEVESQUE, J. P. 2010. Positioning of bone marrow hematopoietic and stromal cells relative to blood flow in vivo: serially reconstituting hematopoietic stem cells reside in distinct nonperfused niches. Blood, 116, 375-85.

WINSTON, F. & CARLSON, M. 1992. Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. Trends Genet, 8, 387-91.

WOLF, S. F., JOLLY, D. J., LUNNEN, K. D., FRIEDMANN, T. & MIGEON, B. R. 1984. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc Natl Acad Sci U S A, 81, 2806-10.

WORKMAN, J. L. & KINGSTON, R. E. 1998. Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annu Rev Biochem, 67, 545-79.

WRIGHT, W. E. & FUNK, W. D. 1993. CASTing for multicomponent DNA-binding complexes. Trends Biochem Sci, 18, 77-80.

WU, H. & ZHANG, Y. 2011. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev, 25, 2436-52.

WU, M., WANG, P. F., LEE, J. S., MARTIN-BROWN, S., FLORENS, L., WASHBURN, M. & SHILATIFARD, A. 2008. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol Cell Biol, 28, 7337-44.

XU, F., LI, X., WU, L., ZHANG, Q., YANG, R., YANG, Y., ZHANG, Z., HE, Q. & CHANG, C. 2011. Overexpression of the EZH2, RING1 and BMI1 genes is common in myelodysplastic syndromes: relation to adverse epigenetic alteration and poor prognostic scoring. Ann Hematol, 90, 643-53.

YAMASHITA, R., SUZUKI, Y., SUGANO, S. & NAKAI, K. 2005. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. Gene, 350, 129-36.

YAMAZAKI, S., EMA, H., KARLSSON, G., YAMAGUCHI, T., MIYOSHI, H., SHIODA, S., TAKETO, M. M., KARLSSON, S., IWAMA, A. & NAKAUCHI, H. 2011. Nonmyelinating Schwann cells maintain hematopoietic stem cell hibernation in the bone marrow niche. Cell, 147, 1146-58.

YANG, M., LI, K., NG, P. C., CHUEN, C. K., LAU, T. K., CHENG, Y. S., LIU, Y. S., LI, C. K., YUEN, P. M., JAMES, A. E., LEE, S. M. & FOK, T. F. 2007. Promoting effects of serotonin on

hematopoiesis: ex vivo expansion of cord blood CD34+ stem/progenitor cells, proliferation of bone marrow stromal cells, and antiapoptosis. Stem Cells, 25, 1800-6.

YE, T., KREBS, A. R., CHOUKRALLAH, M. A., KEIME, C., PLEWNIAK, F., DAVIDSON, I. & TORA, L. 2011. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res, 39, e35.

YOSHIDA, H., HAYASHI, S., KUNISADA, T., OGAWA, M., NISHIKAWA, S., OKAMURA, H., SUDO, T. & SHULTZ, L. D. 1990. The murine mutation osteopetrosis is in the coding region of the macrophage colony stimulating factor gene. Nature, 345, 442-4.

YOU, J. S., KELLY, T. K., DE CARVALHO, D. D., TABERLAY, P. C., LIANG, G. & JONES, P. A. 2011. OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. Proc Natl Acad Sci U S A, 108, 14497-502.

YUAN, W., XU, M., HUANG, C., LIU, N., CHEN, S. & ZHU, B. 2011. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem, 286, 7983-9.

YUAN, Z. M., HUANG, Y., ISHIKO, T., NAKADA, S., UTSUGISAWA, T., SHIOYA, H., UTSUGISAWA, Y., SHI, Y., WEICHSELBAUM, R. & KUFE, D. 1999. Function for p300 and not CBP in the apoptotic response to DNA damage. Oncogene, 18, 5714-7.

ZANG, C., SCHONES, D. E., ZENG, C., CUI, K., ZHAO, K. & PENG, W. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics, 25, 1952-8.

ZHANG, H., LI, H., XI, H. S. & LI, S. 2012. HIF1alpha is required for survival maintenance of chronic myeloid leukemia stem cells. Blood, 119, 2595-607.

ZHANG, W., GLOCKNER, S. C., GUO, M., MACHIDA, E. O., WANG, D. H., EASWARAN, H., VAN NESTE, L., HERMAN, J. G., SCHUEBEL, K. E., WATKINS, D. N., AHUJA, N. & BAYLIN, S. B. 2008. Epigenetic inactivation of the canonical Wnt antagonist SRY-box containing gene 17 in colorectal cancer. Cancer Res, 68, 2764-72.

ZHANG, X. & REN, R. 1998. Bcr-Abl efficiently induces a myeloproliferative disease and production of excess interleukin-3 and granulocyte-macrophage colony-stimulating factor in mice: a novel model for chronic myelogenous leukemia. Blood, 92, 3829-40.

ZHANG, X., SUBRAHMANYAM, R., WONG, R., GROSS, A. W. & REN, R. 2001. The NH(2)-terminal coiled-coil domain and tyrosine 177 play important roles in induction of a myeloproliferative disease in mice by Bcr-Abl. Mol Cell Biol, 21, 840-53.

ZHAO, C., BLUM, J., CHEN, A., KWON, H. Y., JUNG, S. H., COOK, J. M., LAGOO, A. & REYA, T. 2007. Loss of beta-catenin impairs the renewal of normal and CML stem cells in vivo. Cancer Cell, 12, 528-41.

ZHAO, C., CHEN, A., JAMIESON, C. H., FERESHTEH, M., ABRAHAMSSON, A., BLUM, J., KWON, H. Y., KIM, J., CHUTE, J. P., RIZZIERI, D., MUNCHHOF, M., VANARSDALE, T., BEACHY, P. A. & REYA, T. 2009. Hedgehog signalling is essential for maintenance of cancer stem cells in myeloid leukaemia. Nature, 458, 776-9.

ZHAO, R. C., JIANG, Y. & VERFAILLIE, C. M. 2001. A model of human p210(bcr/ABL)-mediated chronic myelogenous leukemia by transduction of primary normal human CD34(+) cells with a BCR/ABL-containing retroviral vector. Blood, 97, 2406-12.

ZHAO, X. D., HAN, X., CHEW, J. L., LIU, J., CHIU, K. P., CHOO, A., ORLOV, Y. L., SUNG, W. K., SHAHAB, A., KUZNETSOV, V. A., BOURQUE, G., OH, S., RUAN, Y., NG, H. H. & WEI, C. L. 2007. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell, 1, 286-98.

ZHOU, V. W., GOREN, A. & BERNSTEIN, B. E. 2011. Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet, 12, 7-18.

ZOU, P., YOSHIHARA, H., HOSOKAWA, K., TAI, I., SHINMYOZU, K., TSUKAHARA, F., MARU, Y., NAKAYAMA, K., NAKAYAMA, K. I. & SUDA, T. 2011. p57(Kip2) and p27(Kip1) cooperate to maintain hematopoietic stem cell quiescence through interactions with Hsc70. Cell Stem Cell, 9, 247-61.

ZOVEIN, A. C., HOFMANN, J. J., LYNCH, M., FRENCH, W. J., TURLO, K. A., YANG, Y., BECKER, M. S., ZANETTA, L., DEJANA, E., GASSON, J. C., TALLQUIST, M. D. & IRUELA-ARISPE, M. L. 2008. Fate tracing reveals the endothelial origin of hematopoietic stem cells. Cell Stem Cell, 3, 625-36.