AN EXPLORATION OF THE MOUSE EPIGENOME AT METAPHASE AND INTERPHASE DURING EMBRYONIC DIFFERENTIATION

by

RICHARD MICHAEL BOWKER

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Immunity & Infection

Institute of Biomedical Research

University of Birmingham

February 2015

Abstract

Histone modifications form an important part of the epigenetic landscape that controls many aspects of cellular function, including regulation of gene expression and cell differentiation. The persistence and inheritance of many of these modifications through the cell cycle and differentiation are still unknown. Here, I show global epigenetic karyotypes of metaphase chromosomes labelled to highlight specific marks. Metaphase is transcriptionally inactive and so epigenetic marks here are not simply reflective of gene transcription. I found that histone marks such as H3K27me3 are inherited through differentiation, whereas others such as H4K20me3 have re-organised distributions. FISH analysis allowed the alignment of genetic features with H3K4me3 and H3K27me3 distributions, showing that these marks are correlated with increased gene density, revealing a deeply intertwined distribution in ES cells, indicating bivalency.

Focusing on the *Hoxa* cluster using N-ChIP in ES cells allowed the analysis of histone modification prevalence at the single gene level in ES cells. Most histone modifications remain stable between G_1/S phase and G_2/M phase, although H3K9ac decreases in ES cells at G_2/M. Results for bivalent modifications show permissive chromatin environments denoted by high H3K4me3 and low H3K27me3 methylation at gene promoters that are expressed soon after the onset of differentiation, denoting a predictive chromatin signature. This signature was altered after five days of differentiation, where H3K4me3 increases and H3K27me3 decreases at most *Hoxa* promoters, concomitant with the rise in expression of some *Hoxa* genes, displaying the dynamic properties necessary to represent a mechanism for control of transcription during differentiation.

For Nina.

You gave everything new meaning.

I’m lucky to have such a curious and beautiful daughter.

Thank you

ACKNOWLEDGEMENTS

Thanks Laura. Thanks for helping me get through the maze and out the other side. I really appreciate you sticking with me through the good and the bad, and for being a general guru for all things scientific.

Thanks to the wonderful people in the Chromatin group; Jenny, Hannah S, Hannah G, Milan, John, Maaike, Elsa, Marianne and anyone else I'm forgetting. Big thanks to Phil Antczak for giving me a hand with the stats.

Thanks to my parents, who always gave me someone to turn to in times of crisis, as well as unwavering support and also some good advice.

Finally, thanks to Fede. Without your support and love, this would not be here today. I really cannot thank you enough. You're the most wonderful person I know, in every sense.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS USED

ES cell	Embryonic stem cell
ESC	Embryonic stem cell
CCE/R	Male mouse ES cell culture line
OS25	Male mouse ES cell culture line
MEF	Mouse embryonic fibroblast
LIF	Leukaemia inhibitory factor
iPS cell	Induced Pluripotent Stem cell
H3, H4	Histone H3, Histone H4
K4, K9, S10	Lysine 4, Lysine 9, Serine 10
me1, me2, me3	Monomethyl, dimethyl, trimethyl histone modification
ac	Acetyl histone modification
HMT	Histone methyltransferase
HDM	Histone demethylase
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
DNMT	DNA methyltransferase
CpG	Cytosine – Phosphate – Guanine
N-ChIP	Native chromatin immunoprecipitation
TSS	Transcription start site
Pol II	RNA Polymerase II

Chr	Chromosome
FISH	Fluorescence in situ hybridisation
DAPI	Diamidino-2-phenylindole, a fluorescent stain for DNA
FITC	Fluorescein isothiocyanate, a fluorescent dye
Mb	Megabase (10^6 bases)
G-band	Chromosome band that stains with Giemsa
d0	Undifferentiated ES cell
d2	2 days after removal of LIF – differentiated ES cell
RA	Retinoic acid

1. INTRODUCTION

1.1 THE CELL CYCLE

1.1.1 The components of the eukaryotic cell cycle

In eukaryotes, cell division is regulated by a series of proteins that prevent unregulated cell division and promote genome stability. This results in four discrete periods of time which a cell spends between one division and the next. G_1 (gap 1) phase starts after a parent cell has fissioned, and ends at the G_1 restriction point, which is a checkpoint where a cascade of proteins prevent the cell from moving to the next phase of the cycle while assessing the extent of any DNA damage accrued by the cell since mitosis (Rossow *et al.*, 1979, Branzei and Foiani, 2006). At this point, damaged or senescent cells can withdraw from the cell cycle and enter a quiescent state known as G_0, characterised by the continuing metabolic activity but suspended reproductive capacity of the cell (Iwamoto *et al.*, 2008). The network of proteins provides a tightly regulated system that enables cell cycle progression. The balance between progression and inhibition is kept by proteins that respond to DNA damage and genome instability, and prevent the uncontrolled division of a cell (Di Leonardo *et al.*, 1994, Khan *et al.*, 2003).

1.1.2 Genome integrity maintained by cell cycle checkpoints

Checkpoints in mitosis have evolved to mitigate the accumulation of copying errors during DNA replication and also to prevent damaged DNA from being copied to a daughter cell (Kaufmann and Paules, 1996, Deckbar *et al.*, 2011). In eukaryotes, cellular pathways are in place that detect any DNA errors or mutation, mediate the signal produced from these sensors, transduce the signal, and

instigate blockage of the cell cycle and DNA repair, and also effect alterations in gene expression (Navas *et al.*, 1996, Sclafani and Holzen, 2007). These sensors are often closely linked to other cellular processes such as cell cycle regulation and DNA replication (Navas *et al.*, 1996). CDK-Cyclin complexes are critical for the progression of the cell cycle. The four cyclins (A, B, D, E) associate with their respective CDKs (cyclin dependent kinases) to move cells through the G_1/S block (Cyclins D and E, CDKs 4 and 2) and the G_2/M block (Cyclins A and B, CDKs 2 and 1) (Morgan, 1997, Choi and Anders, 2013, Dai *et al.*, 2013). These CDK-cyclin complexes are part of an intricate regulatory mechanism that enforces each checkpoint.

1.1.3 The G_1/S checkpoint

At G_1/S, cells make the decision whether or not to divide. If DNA is undamaged, accumulation of Cdc25A causes dephosphorylation of CDKs 2 and 4, which are complexed with cyclins E and D (Iavarone and Massague, 1997, Iwamoto *et al.*, 2008). These subsequently phosphorylate pRB (retinoblastoma protein), which dissociates from the E2F transcription factor. This protein activates transcription of itself, cyclin genes and S phase-specific genes that promote entry into S phase (Mudryj *et al.*, 1991, Piao *et al.*, 2011). However, upon DNA damage, cellular proteins act to disrupt the action of the CDK-Cyclin E complex to stall the cell's passage into S phase (Branzei and Foiani, 2006). At this point, the DNA repair and cell cycle control mechanisms are intertwined, which indicates the critical importance of both these processes to the longevity of the cell.

1.1.4 DNA replication during synthesis phase

After the G_1 checkpoint, the nucleus begins to replicate its DNA. Thus begins the synthesis, or S, phase (Taylor, 1973). The main function of S phase is to faithfully reproduce the genome in order to have enough genetic material for two daughter cells. The speed of cell growth and division is another carefully controlled facet of mammalian cells, and can vary greatly even between cells of the same organism. Embryonic stem cells (ESCs), for example, can be distinguished from differentiated cells from the length of their G_1 phase (Arai *et al.*, 2011, Ruiz *et al.*, 2011). Adult differentiated cells have a tightly regulated and effective G_1 block, whereas in stem cells there is a less lengthy control on the passage from G_1 into S phase, and so cells progress quickly into DNA synthesis. **Figure 1.1** shows that a much larger proportion of an unsynchronised cell population are in G_1 phase in differentiated cells (MEFs) than ES cells (Tichy, 2011). However, it has also been found that embryonic stem cells still have the capacity to impose a strict G_1 block after induction of DNA damage with UV light, indicating that ES cells still maintain effective DNA repair pathways (Barta *et al.*, 2010). It has even been suggested that artificially blocking the cell cycle of ES cells is sufficient to cause the cells to differentiate (Ruiz *et al.*, 2011). This confirms the influence that cell cycle control has over the behaviour of pluripotent and differentiating cells.

1.1.5 G_2 phase

S phase is followed by the second gap phase: G_2 phase. Similar to G_1, this is also a period with high protein synthetic activity, and again contains a checkpoint through which the dividing cell must pass (Weinert and Hartwell, 1990, Khammanivong *et al.*, 2013). The purpose of this checkpoint is to ensure that nuclear DNA has been replicated properly before the onset of mitosis

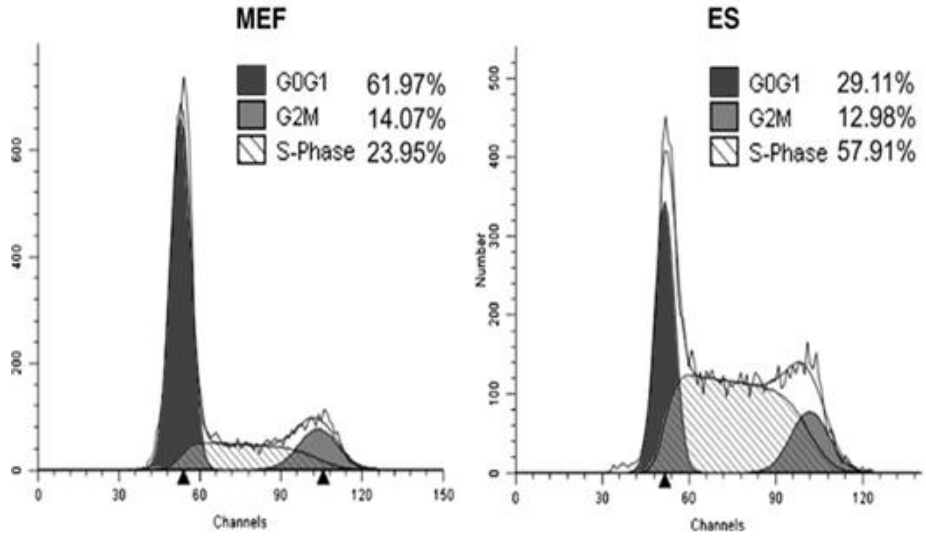


Figure 1.1 Cell cycle profiles of mouse embryo fibroblasts (MEF) and murine embryonic stem cells (ES). Asynchronous mESCs or early passage mEFs were grown to 60% confluency before harvest and fixation. Cells were treated with RNase A and stained with propidium iodide prior to analysis by flow cytometry for DNA content. Figure taken from Tichy, 2011.

(Plesca *et al.*, 2007). The G_2 checkpoint has a similar mechanism to the G_1 checkpoint, in that it again relies on cyclin-CDK interactions to control progression to the next cell cycle phase (Messier *et al.*, 2013).

1.1.6 Cell division during mitosis phase

Mitosis phase, or M phase, begins when the cell's chromosomes begin to condense, and ends when the two daughter cells have separated. M phase can be subdivided into another four phases, according to the appearance of chromosomes under a light microscope (Bishop and Young, 1977, De Arce *et al.*, 1982, Bolzer *et al.*, 2005). The first section, prophase, is characterised by the initial condensing of the chromosome into a more highly packaged structure. Centrioles start to form, the mitotic architecture of microtubules begin to align, and the nuclear membrane breaks down. Metaphase can be identified as the stage by which time chromosomes have condensed into distinct, separate units of DNA and align themselves in preparation for the separation of chromatids (Bishop and Young, 1977). Metaphase chromosomes are widely studied due to the ease of their microscopic identification and the reproducible banding patterns they produce when stained with various dyes. One such dye is Giemsa, which preferentially binds to AT rich regions of the chromosome to produce numerous bands along the chromosome that correspond to areas with relatively low gene density (De Arce *et al.*, 1982, Hurst and Eyre-Walker, 2000). Separation of chromatids occurs during anaphase, at which point chromosomes are pulled apart by the action of the microtubules of the mitotic spindle towards opposite poles of the cell (Royle, 2012). The final stage of mitosis is telophase, where nuclei of the daughter cells begin to form, cytokinesis separates the cells, and the final

cleavage of the two resultant cells occurs (Ashraf and Godward, 1980). The tightly packaged DNA/histone complex is relaxed and becomes de-condensed (Matsuoka *et al.*, 1994). The mitosis phase of the cell cycle demonstrates one essential facet of the DNA-histone interaction; that of proper packaging of DNA into condensed structures. However, there are other important roles for this nuclear complex, including its influence on transcription factor binding and gene expression.

1.2 PLURIPOTENCY AND STEM CELLS

1.2.1 Pluripotency in embryonic stem cells

Pluripotency refers to the state of embryonic cells at a very specific stage of development. The inner mass cells of the mammalian embryonic blastocyst have the potential to differentiate into any corporeal cell type (Martin, 1981). Embryonic stem cells (ESCs) can be isolated *in vitro* from the inner cell mass of the blastocyst. These cells were first extracted from mice in 1981, and were identified by their ability to differentiate into ectoderm, mesoderm and endoderm, but not extra-embryonic tissues, so are therefore described as pluripotent (Martin, 1981, Evans and Kaufman, 1981). These cells can divide symmetrically to produce two ES daughter cells, or they can differentiate to occupy all the niches vital to the development of the organism (Niwa, 2007).

Pluripotency is a robust state that is only disrupted by specific cues that instruct the cell to differentiate. A recent study involved knocking down many different genes in human ES cells; only a handful produced significant phenotype change.

One of these was the transcription factor Oct4, encoded by the *Pou5f1* gene (Nishiyama *et al.*, 2013). This transcription factor is also known to interact with a great many other components of the network of factors that maintain pluripotency, and is thought to be a key regulator of the undifferentiated state of ESCs (Gao *et al.*, 2013). Differentiation of the blastocyst into the various embryonic tissues is accompanied by a rapid loss of *Pou5f1* (a hallmark of stem cells) expression (Scholer *et al.*, 1989). This is followed by the development of the three primary germ layers: endoderm, mesoderm and ectoderm and eventually the entire cohort of cell types present in the adult (**figure 1.2**) (Meregalli, 2011). It is not fully known by what mechanism these cells differentiate, and why some cells become ectodermal and others endodermal, however the levels of several pluripotency factors seem to be critical in the early stages of ESC differentiation (Zhou *et al.*, 2007).

1.2.2 Nuclear factors involved in the maintenance of pluripotency

Control of the process of differentiation is also thought to lie with transcription factors such as Oct4, as well as Nanog and Sox2 (Boyer *et al.*, 2005, Abranches *et al.*, 2013). These proteins are known to be markers of pluripotency, and induce a pattern of gene expression that maintains the undifferentiated state of ESCs (Zhou *et al.*, 2007, Whyte *et al.*, 2013). Nanog is a key regulator of pluripotency, and its expression is greatly reduced as soon as cells differentiate from the blastocyst stage of development, and loss of Nanog activity mirrors loss of pluripotency very closely (Chambers *et al.*, 2003, Abranches *et al.*, 2013). This transcription factor further activates its own expression, in a positive feedback loop, and also activates other genes vital for the maintenance of pluripotency

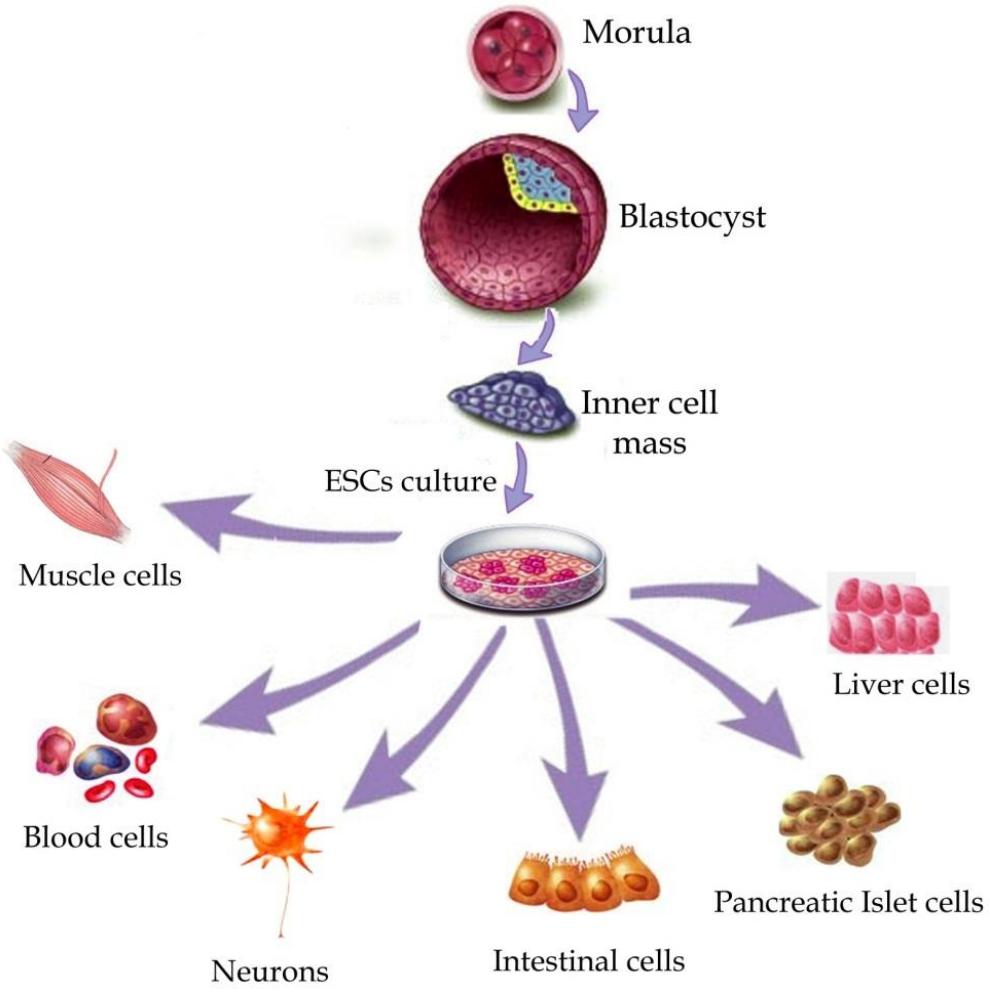


Figure 1.2 Diagram describing the differentiation process in mammals. Differentiation potentiality of human embryonic cell lines. ES cells are taken from the inner cell mass of the embryonic blastocyst and can be cultured *in vitro* for prolonged periods of time. Figure from Meregalli *et al.* 2011

(Boyer *et al.*, 2005). Nanog can promote expression of factors such as Oct4 and Sox2, which can interact via heterodimerisation to influence gene expression (Jauch *et al.*, 2011). All three pluripotency factors are associated closely with each other, targeting the same genes and promoting each other's production (**figure 1.3**) (Niwa, 2007). Data mining studies have indicated that 50% of Oct4 target regions were also bound by Sox2, and that of these regions, over 90% were also a target for Nanog binding (Boyer *et al.*, 2005, Ng and Lufkin, 2011). It is thought that these three transcription factors have a strong influence in the process of maintaining pluripotency by bringing about the particular pattern of gene expression that keeps embryonic stem cells from spontaneously differentiating. The loss of one or more of these transcription factors can be a cue, or a marker, for cell differentiation (Osorno *et al.*, 2012, Sterneckert *et al.*, 2012). In a normally developing embryo, pluripotent cells are only present transiently, for a matter of hours or days, depending on the species (Theunissen and Silva, 2011). However, in vitro and in artificially induced stem cells, pluripotent cells can be prolonged by maintaining an exogenous supply of Leukaemia Inhibitory Factor (LIF) (Cheng *et al.*, 2012). This cytokine binds to the receptor LIFR to act on the JAK/STAT pathway, among others, to maintain the pluripotent state of the cell (Dani *et al.*, 1998). The JAK/STAT pathway is a signal-transduction pathway that receives a stimulus (in this case the binding of LIF to the LIF-Receptor) and initiates protein-protein and protein-DNA interactions in order to bring about a change in the cell's gene transcription activity (Abranches *et al.*, 2013). This pathway is not only activated to regulate the maintenance of pluripotency, but is also involved in apoptosis and differentiation processes (Chen *et al.*, 2012). The importance of the proper regulation of this pathway can be shown by the increased susceptibility to

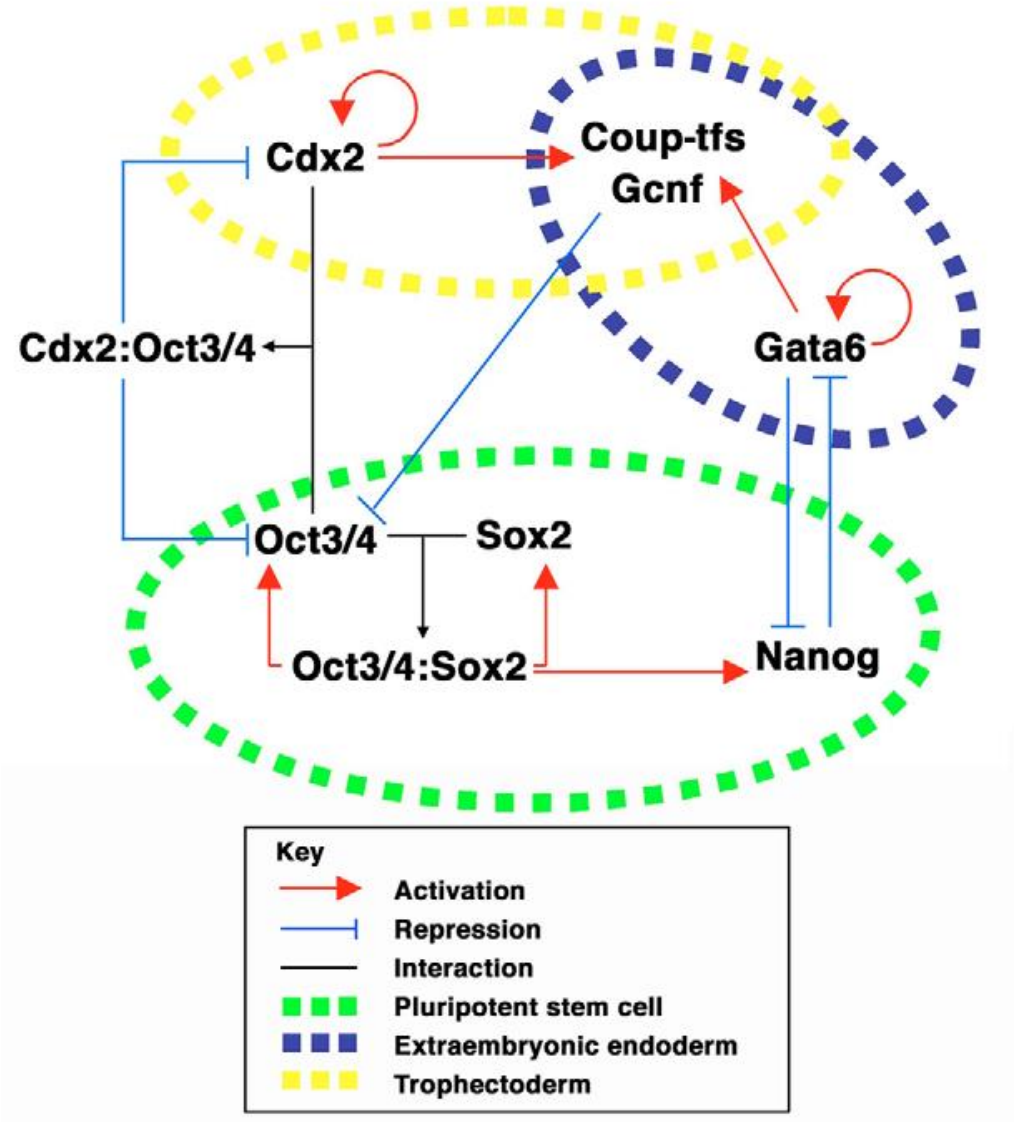


Figure 1.3 Diagram showing the interactions between Oct3/4, Sox2, Nanog and tissue-specific regulators in the embryo. Feedback loops show how Oct3/4 and Sox2 levels are maintained. The pluripotency factors promote their own transcription, but this is carefully controlled by the cell to prevent loss of pluripotency. Figure adapted from Niwa 2007.

cancer if the pathway is constitutively expressed, for example due to mutation or polymorphisms in the underlying gene sequences (Chen *et al.*, 2013b).

1.2.3 Differentiation of embryonic cells and the loss of pluripotency

Embryo development results in the loss of pluripotency, however some forms of stem cells are present throughout the life of an organism. Adult stem cells are differentiated cells, but still have the capacity to differentiate further in order to replace dead or damaged cells (Chicha *et al.*, 2011). These can usually only differentiate into one or a few types of terminally differentiated cells, and as such there are many classes of adult stem cells maintained at all times, often termed progenitor cells, which can differentiate into only a few different final cell types (Wang *et al.*, 2005, Yanagida *et al.*, 2013).

1.2.4 The cell cycle and control of pluripotency

In ESCs, the cell cycle is thought to be a key factor in the maintenance of the cell's pluripotency (Andang *et al.*, 2008, Chen *et al.*, 2013a). Oct4, also known as Oct3 or Oct3/4, is a transcription factor highly up-regulated in pluripotent cells and is closely linked to this process, as a loss of *Pou5f1* (the name for the Oct4 gene) expression is quickly followed by changes in the timings of the cell cycle phases, and subsequently differentiation of the cell (Becker *et al.*, 2007, Becker *et al.*, 2010). The increased length of the cell cycle then inhibits rapid proliferation of cells, driving cells further down the path to differentiation (Calder *et al.*, 2013). Conversely, in order to induce pluripotency in differentiated cells, the cell cycle must be reverted to an ES cell (short G_1) cycle for pluripotency to be stably maintained (Maherali *et al.*, 2007). In pluripotent cells, *Pou5f1* expression causes

a partial bypass of the G_1 checkpoint by repressing p21 (a G_1 block component), thus the loss of Oct4 results in a re-instatement of the complete gap 1 block (Hanna *et al.*, 2009). The role of micro-RNAs (miRNA) in the bypassing of the G_1 block has also been recently explored (Wang and Blelloch, 2009). These short RNA sequences, along with the RNA induced silencing complex (RISC) bind to specific mRNA transcripts to interrupt their translation into protein. In ES cells, there is a relative abundance of a specific class of miRNAs that promote G_1 to S transition(Wang and Blelloch, 2009). These are thought to act on pRB, p21, as well as other components of the G_1 checkpoint machinery, can act in a checkpoint-independent manner, and may be upregulated by Oct4 or other pluripotency factors (Wang and Blelloch, 2009, Wang *et al.*, 2013).

1.2.5 Induced stem cells: the potential to reverse the differentiation pathway

Until relatively recently, it was though that differentiation was a one-way pathway towards a fully committed cell state. Takahashi and Yamanaka discovered that the process of differentiation could be artificially reversed in mice by introducing extrinsic pluripotency factors to a population of adult fibroblasts (Takahashi and Yamanaka, 2006). A small proportion of these cells became pluripotent as denoted by the resumption of expression of ES cell marker genes. This was later found to be possible also for human cells (Takahashi *et al.*, 2007). These induced pluripotent stem cells (iPS cells) can maintain their pluripotency in vitro using LIF in a similar manner to ES cells (Cheng *et al.*, 2012). As well as foreshadowing important research into the nature of development and differentiation, these cells

can also act as tools to further many other fields of biology. The possibility to derive cells that are pluripotent and yet have the same genotype as the donor cells have important implications for the field of regenerative medicine and others. For example, cells that are difficult to isolate and culture in vivo can now be derived from induced pluripotent cells, given the right developmental cues (Yanagida *et al.*, 2013). A potential difficulty in the clinical use of transplanted iPS cells is their ability to form tumours (Marzi *et al.*, 2013). This is closely related to the ES cell-like profile of the iPS cells and their associated rapidly cycling nature. The G_1 checkpoint is reduced in pluripotent cells, as described earlier, which leaves pluripotent cells susceptible to tumorigenesis (Kawai *et al.*, 2010). The uncontrolled, but undifferentiated, cells go on to form teratomas, similar to those found in the earliest experiments with mice using ES cells (Martin, 1981, Lim *et al.*, 2013). A strong link has been found between the re-establishment of pluripotency and a genome-wide shift in post-translational histone modifications in induced pluripotent stem cells (Mattout *et al.*, 2011). This study found that iPS cells undergo an initial reorganisation of nuclear heterochromatin as an early event in the reprogramming of somatic cells. The heterochromatic signature for pluripotency foreshadows the upregulation of pluripotency genes and other epigenetic features necessary to maintain a stem cell phenotype.

1.3 CHROMATIN AND CHROMOSOMES

1.3.1 Chromosomes, DNA packaging and histones

Chromosomes are self-contained units of genetic material consisting of a length of DNA wound around histones, this DNA-histone complex being known as

chromatin. The basic unit of chromatin is the nucleosome, a structure of a single histone octamer wrapping around 147 base pairs of DNA (Luger *et al.*, 1997). Components of the octamer are two units each of histones H3 and H4, and two units each of histones H2A and H2B (Bakayev *et al.*, 1977, Blacketer *et al.*, 2010). The action of chromatin condensation can occur as neighbouring nucleosomes move closer together and organise themselves into fibres of 10nm diameter (Blacketer *et al.*, 2010, Fussner *et al.*, 2011). As a result of the compaction of the DNA-histone complex, transcription of genes by RNA polymerase (as well as other factors involved in the transcription machinery) is less frequent, due to the reduced accessibility of promoters and DNA binding sites to proteins (Mirsky, 1971, Belch *et al.*, 2010). Fibres of chromatin can be further compacted due to interactions between histones on different nucleosomes, thought to be the H4 tail and histone H2A, and compacted even further to form the thickly condensed chromosome that can be seen at metaphase under a light microscope (**figure 1.4**) (Lafontaine and Chouinard, 1963, Liu *et al.*, 2006, Sha, 2009, Yamagata *et al.*, 2009, Bian and Belmont, 2012). It is generally thought that the formation of the chromosome structure may be due to the association between chromatin and a protein-based nuclear matrix, or scaffold, which facilitates condensation of chromatin (Heng *et al.*, 2004, Bian and Belmont, 2012). This has been disputed, however, by studies in newts that imply that cross-links in chromatin, approximately 15kb apart and facilitated by linker proteins, are the basis of DNA condensation (Poirier and Marko, 2002). The mechanisms of chromatin compaction are not yet fully understood, but have roles in gene expression and cell division (Bian and Belmont, 2012). The compaction of chromatin has important consequences for the

cell, not only to fit the genetic material into the nucleus, but also to selectively inactivate regions of the genome into heterochromatin.

1.3.2 Heterochromatin: Dense and inactive chromatin

Heterochromatin is a highly condensed structure present throughout the cell cycle (except during replication) that represses activity of genes contained within its boundaries (Honda *et al.*, 2012). Heterochromatin is repeat-rich and contains few functional promoters, and can be constitutive or facultative (Trojer and Reinberg, 2007). Constitutive chromatin represents a permanent state of tight compaction, such as can be found at the centromeres of eukaryotic chromosomes (Fodor *et al.*, 2010). Conversely, facultative heterochromatin can be formed in a chromosomal region whose transcription is no longer required. The female inactive X chromosome is an example of facultative chromatin that has been formed from euchromatin during embryo development. As the embryo develops, a random copy of the X chromosome is inactivated, leaving a dense mass of chromatin termed a Barr body (Augui *et al.*, 2011). Both types of heterochromatin are associated with decreased histone acetylation and increased histone methylation at H3K9 and H4K20 (Trojer and Reinberg, 2007). These are examples of epigenetic carriers of information, which are crucial to the understanding of both the structure and function of chromatin, as well as having a critical role to play in the control of the cell cycle.

1.3.3 Epigenetic marks

Epigenetic information is defined as markers that can alter gene expression without changing the sequence of DNA (Reik *et al.*, 2001, Kouzarides, 2007).

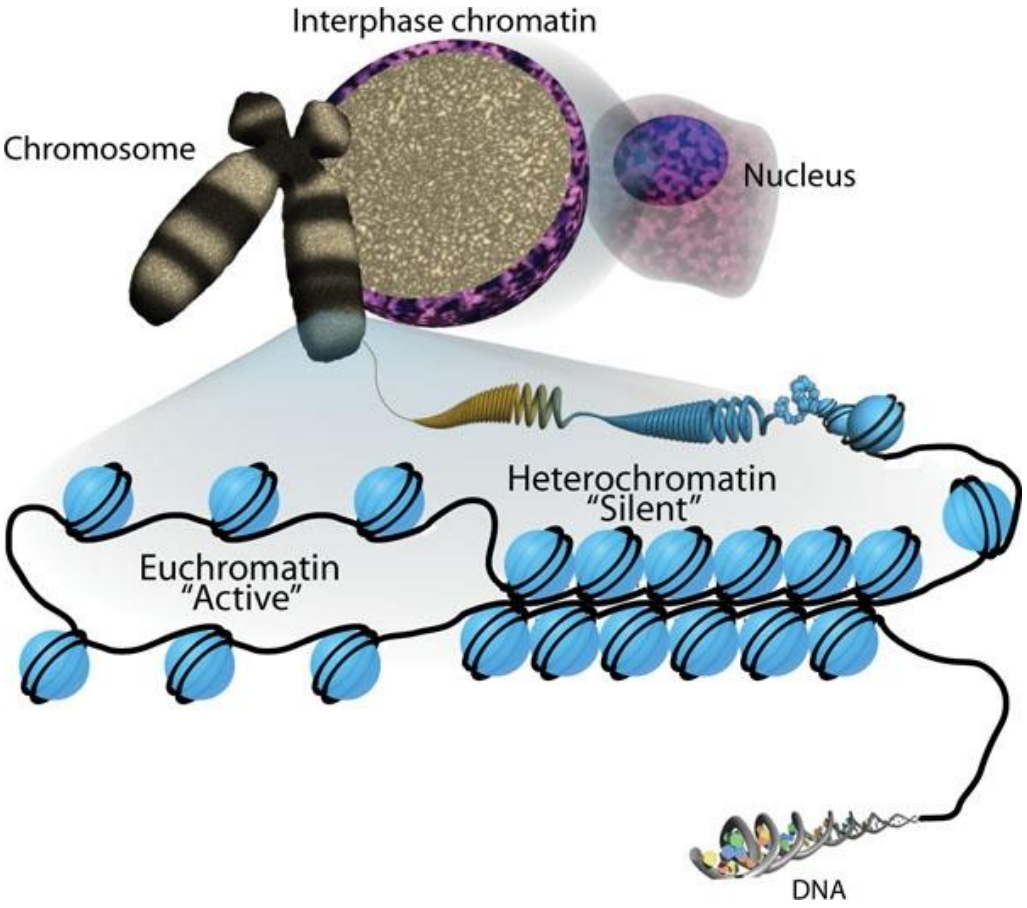


Figure 1.4 Diagram showing the various levels of chromatin packaging. Naked DNA (black line) is coiled around nucleosomes (blue spheres). This chromatin can be arranged into a fibre and then further compacted to form a metaphase chromosome as above. Figure taken from Sha and Boyer 2009.

These can be altered by, or respond to, changes in the cellular environment, and persist through multiple cell divisions to remain as a stable, heritable carrier of information (Buckley *et al.*, 2012, Huang *et al.*, 2013). However, this definition is somewhat misleading, as epigenetic motifs have a role in eukaryotes that is distinct from the relatively immutable and infinitely heritable genetic information encoded by DNA sequences. In several contexts, epigenetic information performs an analogous role to the DNA sequence, such as maintaining the cell identity of a terminally differentiated cell (Hanna *et al.*, 2009, Zheng and Pan, 2010). But an equally important aspect of the nature of epigenetic marks is their plasticity, i.e. the ability of these marks to change the phenotype of the cell; for example by being lost or gained through the course of differentiation to result in a distinct cell type, which then maintains a steady state level of epigenetic marks as above (Breiling *et al.*, 2004, Martino *et al.*, 2013). Currently, it is not known whether epigenetic marks fit the description of a heritable mechanism for carrying information across generations, but it has been discovered that there are several elaborate mechanisms for making sure that epigenetic markers are conveyed from one cell generation to the next (Breiling *et al.*, 2004, Francis, 2009). It may be the case that some epigenetic marks fulfil only functions to do with cell plasticity and may be easily and dynamically changed to meet the need for a switch in cell expression involved in cell differentiation. Other marks may display greater longevity through the cell cycle and mitosis to provide a steady and heritable mechanism to preserve epigenetic information through the generations. A molecular basis for these contrasting roles of epigenetic marks is evident in the varying half-lives of some of these modifications in mammalian cells. Acetyl- and phospho- groups display a much shorter longevity in cells than methyl- groups,

indicating a possible role in inheritance for methylated chromatin (Jackson *et al.*, 1975, Zee *et al.*, 2010).

1.3.4 DNA methylation: Epigenetic modification of the DNA molecule

There are several different forms and chromosomal locations of epigenetic marker, including DNA methylation and histone modifications. DNA methylation is facilitated by DNA methyltransferase enzymes (DNMTs) that methylate cytosine bases on the DNA double helix (Tucker *et al.*, 1996, Athanasiadou *et al.*, 2010). 70-80% of CpGs are methylated in the genome, indicating its abundance (Illingworth *et al.*, 2010). Regions free of methylated cytosines in stretches of CG-rich DNA are known as CpG islands (Bird, 1986, Jones, 2012). These are associated with gene-rich and highly conserved areas of the genome, probably because methylated cytosine residues can easily mutate to thymidine, and these mutations are often carried through the cell cycle due to the failure of DNA repair enzymes to recognise the mutation (Bird, 1986, Olinski *et al.*, 2010). This can also explain the depletion of CpGs relative to other dinucleotides. Genes vital for cell survival must be protected from such damage, and so remain free of CpG methylation, leading to the persistence of these islands of CpG integrity (Antequera *et al.*, 1990). DNA methylation is generally associated with repressing gene transcription, and is often aberrant or erroneously positioned during the progression of cancer cells, activating oncogenes or repressing tumour suppressors, highlighting its importance as a negative regulator of gene expression (Pai *et al.*, 2011, Cannuyer *et al.*, 2013, Rhee *et al.*, 2013).

CpG methylation is faithfully replicated onto newly replicated strands of DNA because DNMT1 enzyme has a high affinity for hemi-methylated CpG sites, methylates the unmodified strand, and completes the inheritance of epigenetic DNA methylation for that cell cycle (Pradhan *et al.*, 1999, Probst *et al.*, 2009). DNA methylation appears to be a stably inherited epigenetic mark that is carefully and faithfully maintained. Other DNMTs, such as DNMT3a and 3b, methylate previously unmodified CpGs at some loci, which implies a role for these enzymes in the establishment of a new epigenetic signature, possibly in response to external or developmental stimuli (Okano *et al.*, 1999, Athanasiadou *et al.*, 2010). Crucially, when all three DNMT enzymes are inducibly knocked out in mouse cells, it was found that DNA methylation levels reverted to almost zero (Tsumura *et al.*, 2006). This indicates that DNMT enzymes are indispensible for maintaining this chromatin modification, and that any previous DNA methylation patterns are lost upon removal of these DNA methyltransferases (Tsumura *et al.*, 2006, Yamanaka *et al.*, 2010). This mechanism was confirmed for the *Pou5f1* locus, which undergoes CpG methylation after differentiation of ES cells. However, upon knockout of the establishing DNMTs 3a and 3b, promoter and enhancer regions of *Pou5f1* showed a constant low level of DNA methylation (Athanasiadou *et al.*, 2010). These findings highlight the importance of enzymatic processes in maintaining the DNA methylation signature throughout cellular growth and development. These observations point to an active mechanism of stability, as opposed to the more passive mechanisms of genetic DNA inheritance.

1.4 EPIGENETICS AND HISTONE MODIFICATIONS

1.4.1 Post-translational histone modification

Facilitating chromosome compaction is not the only function of histones. A key structural feature of the histone molecule is the protrusion of the N-terminal histone tail. This is a primary structure amino acid sequence that sits on the outside of the nucleosome, and as such is accessible to enzymes and factors that can chemically modify individual amino acid residues (Luger *et al.*, 1997, Schwartz *et al.*, 2010, Margueron and Reinberg, 2011). There are over 80 histone tail sites which can be modified, and several different organic and inorganic chemical groups can be added. Methyl, acetyl and phosphate moieties are most commonly added, but there are other possible larger modifications including ubiquitylation, ADP ribosylation and citrullination (Kouzarides, 2007, Arnaudo and Garcia, 2013). **Figure 1.5** shows an array of possible post-translational histone modifications, which are deposited and removed by enzymes and cofactors such as Acetyl CoA (provides an acetyl group) and S-adenosyl-methionine (provides a methyl group) (Pestana *et al.*, 1971, Bhaumik *et al.*, 2007). In the case of acetylation, histone acetyltransferases (HATs) add acetate from Acetyl CoA to histone tail residues, whereas histone de-acetylases (HDACs) remove them (Ekwall *et al.*, 1997). The same is true of methylation, with histone methyltransferases (HMTs) and histone de-methylases (HDMs) adding and removing methyl groups (taken from S-adenosyl-methionine) from the N-terminus (Shi *et al.*, 2004). The opposing actions of these enzymes result in a dynamic equilibrium of histone modifications as they are constantly being added and removed (figure 1.6). Any heritable, epigenetic,

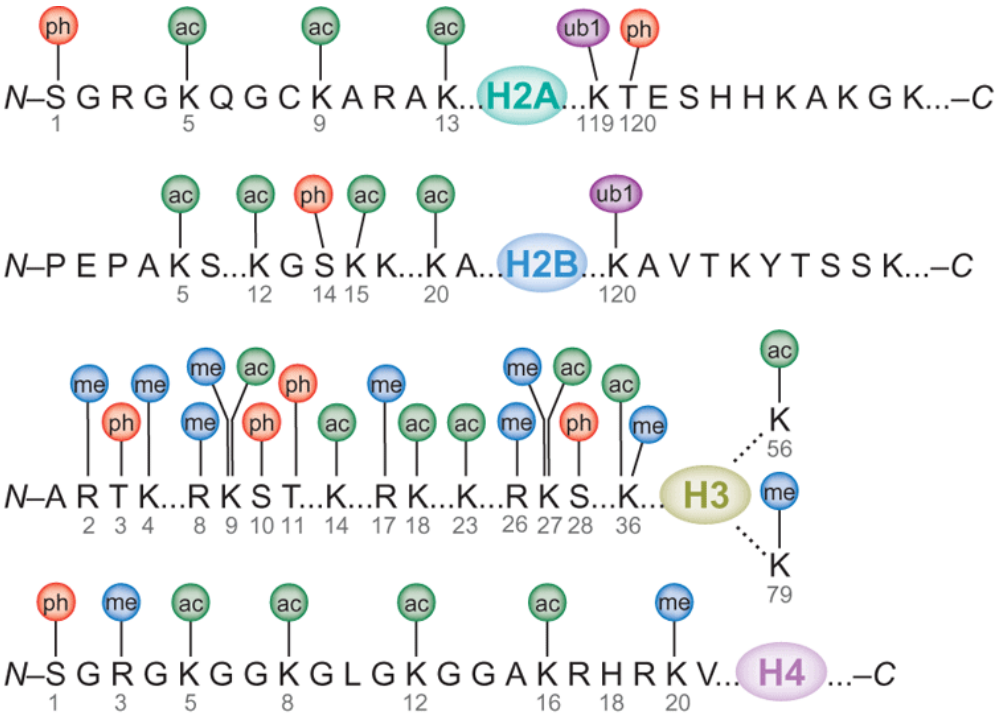


Figure 1.5 Histone modifications include acetylation (ac), methylation (me), phosphorylation (ph) and ubiquitination (ub1). Most of the known histone modifications occur on the N-terminal tails of histones, with some exceptions including ubiquitination of the C-terminal tails of H2A and H2B and acetylation and methylation of the globular domain of H3 at K56 and K79, respectively. Globular domains of each core histone are represented as coloured ovals. Figure taken from Bhaumik *et al.* 2007.

mark is stable through the cell cycle, and so the equilibrium in this case would favour the action of the transferases, ensuring maintenance of that mark.

1.4.2 Histone acetylation

Histone acetylation was the first to be recognised as a post translational histone modification (Allfrey *et al.*, 1964). It is associated with an open configuration of chromatin and in turn with actively transcribed regions of the genome (Bode *et al.*, 1983, Johnson *et al.*, 1998). It is mainly lysine residues that are acetylated, most of which are located on histones H3 and H4 (Turner and O' Neill, 1995, Kuo *et al.*, 1996, Arnaudo and Garcia, 2013). The structure of the lysine amino acid contains a positively charged ammonium species on its -R group, which can form ionic interactions with DNA and nearby histone molecules to form more tightly bound chromatin (Hong *et al.*, 1993). The disruption of this net positive charge by introducing the negatively charged acetate moiety causes the weakening of these (Nightingale *et al.*, 2006, Jorgensen *et al.*, 2013). intermolecular forces and leads to relaxation of the chromatin (Bannister and Kouzarides, 2011). Acetylation of N-terminal histone tails is a highly dynamic process and is quickly turned over by the opposing actions of HATs and HDACs. This can be observed in studies that employed inhibitors of HDACs, where a significant accumulation of histone acetylation is noticed (Candido *et al.*, 1978, Das *et al.*, 2009). However this is not the only action of acetyl marks on chromatin. Acetylated lysine residues can also recruit chromatin binding proteins that interact with other modified residues nearby, creating a form of communication between epigenetic marks (Forneris *et al.*, 2005, Bronner, 2011).

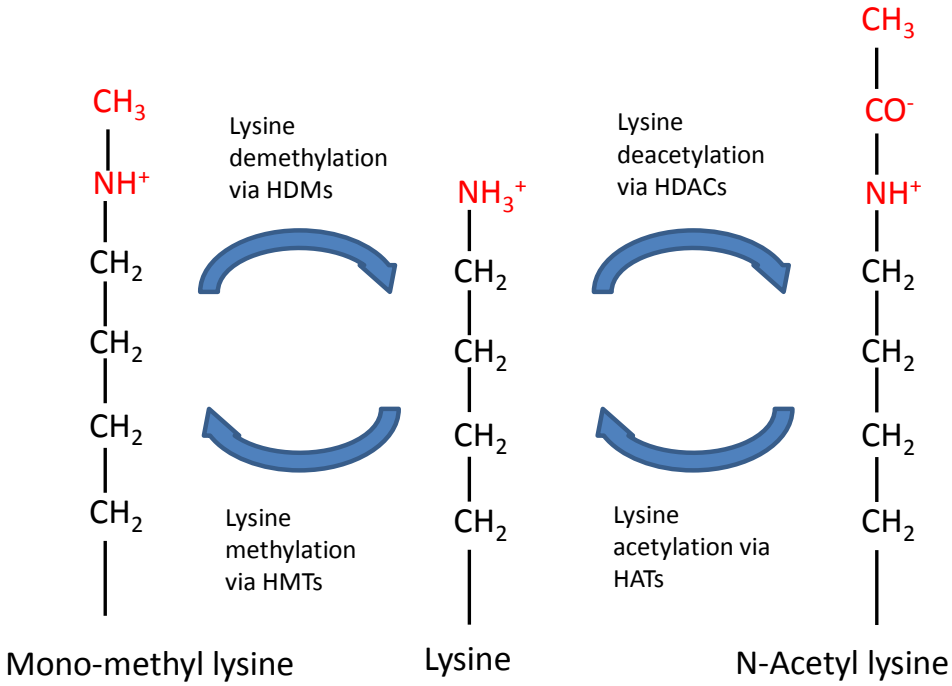


Figure 1.6 Representation of the chemical structures of the R groups of modified and unmodified lysine residues. The section in red denotes the end of the molecule that is modified by chromatin- modifying enzymes. For this representation, mono-methyl lysine has been used, however further methylation of the lysine molecule can result in di-methyl lysine (as above, but with CH_2-CH_3 instead of –CH_3) and tri-methyl lysine (CH_2-CH_2-CH_3 in place of the mono-methyl group). Enzymes involved in these modifications have been annotated above.

Lysine residues can be acetylated on histones H2B, H3 and H4, at several different sites (**figure 1.5**). This allows for many combinations of histone modifications, and therefore presents the opportunity to store a great deal of epigenetic information (Turner, 1998, Strahl and Allis, 2000). Several classes of enzymes have been found to interact specifically with modified chromatin (Agalioti *et al.*, 2002, Palacios *et al.*, 2008). These often have characteristic domains that recognise specific histone modifications; acetylated lysines are often recognised by proteins containing bromodomains, which are peptide motifs highly conserved between mammals, plants and yeast (Tamkun *et al.*, 1992, Bannister and Kouzarides, 2011).

1.4.3 Histone methylation

As well as being acetylated, histone tail residues can also be methylated at lysine and at other residues. Firstly, unlike lysine acetylation, addition of a neutral methyl group does not affect the net charge of the lysine molecule (Kubicek and Jenuwein, 2004). Also, multiple (1, 2 or 3) methylations can occur on the same amino acid residue that can interact with different chromatin-binding enzymes, as shown in **figure 1.5** (Martin and Zhang, 2005). Until recently, it was thought that histone methylation was a stable and persistent modification, similar to DNA methylation (Rice and Allis, 2001, Peters and Schubeler, 2005). However, after the discovery of histone demethylases, beginning with Lsd1, it became clear that methylation is also controlled through equilibrium between histone methyltransferases and demethylases (Shi *et al.*, 2004, Forneris *et al.*, 2005). Now it is clear that many different classes of chromatin-binding proteins interact

with methylated residues (Matthews *et al.*, 2007, Adams-Cioaba and Min, 2009). These proteins can bind to methylated histones via specific domains such as chromodomains and PHD fingers (Palacios *et al.*, 2008, Oliver *et al.*, 2012). The possibility of four different states of methylation at each lysine residue, and the possible acetylation of the same residue, provides a highly nuanced and flexible system for the changing of chromatin states and signalling between histone modifications and chromatin-binding enzymes.

1.4.4 Histone phosphorylation

Phosphorylation of N-terminal histone tails occurs mostly on serine, threonine and tyrosine residues (Berger, 2010). Similar to acetylation, this adds a negatively charged chemical species to the side chain of the amino acid, although phosphate is more highly charged (PO_4^{3-}) than acetate (H_3CCOO^-). This is thought to have a repressive effect on chromatin by attracting more ionic bonding to the nucleosome and so to decrease the potential for the underlying DNA to be read by RNA polymerase (Bannister and Kouzarides, 2011, Ribeiro-Mason *et al.*, 2012). The level of phosphorylation is controlled by the opposing actions of kinases and phosphatases. One such enzyme is Aurora B kinase, which phosphorylates serines 10 (a mark associated with mitotic chromatin) and 28 on histone H3 (Sugiyama *et al.*, 2002, Hirota *et al.*, 2005). This enzyme is also an important link between the action of histone modification and the progression of the cell cycle, as the action of phosphorylating H3S10 disrupts the binding of HP1 protein to heterochromatic DNA, allowing the nuclear chromatin to undergo mitosis properly (Murzina *et al.*, 1999, Hirota *et al.*, 2005).

1.4.5 Histone variants

Another feature in the changeable epigenetic landscape is the existence of variant histones. These are histone molecules such as H2A.X, H2A.Z, Cenp-A H3.2 and H3.3 that contain differences in the amino acid sequences of the protein (Sullivan *et al.*, 1994, Meneghini *et al.*, 2003, Shaw *et al.*, 2009). These variant molecules are recognised as such by nuclear factors and are distributed within certain specific regions of the genome (Shaw *et al.*, 2009, Van Attikum and Gasser, 2009, Fachinetti *et al.*, 2013). For example, the histone variant H2A.Z is highly conserved across species, and is found in S phase as a small fraction of total H2A (Santenard and Torres-Padilla, 2009). It also displays only 60% amino acid sequence homology with H2A, although the crystal structures appear more similar (Svotelis *et al.*, 2009). H2A.Z is located at boundary elements situated at the interface between repressed chromatin and active chromatin (Meneghini *et al.*, 2003). The function of the boundary element is to prevent the silencing of active chromatin (euchromatin) by neighbouring repressive chromatin (heterochromatin) (Li and Zhou, 2013). This is a good example of a mechanism that prevents the over-writing of a previously established epigenetic state. In this case, the importance of the boundary element structure is highlighted by the highly conserved nature of the H2A.Z protein, indicating that this mechanism is evolutionarily vital for the continued stability and proper function of the genome (Meneghini *et al.*, 2003). Although H2A.Z is expressed at S phase in dividing mammalian cells, it has also been found to promote the progression of the cell cycle through the G_1 block. It has also been found that the prevalence of H2A.Z at

gene promoters falls rapidly after the start of S phase and is not immediately reinstated (Nekrasov *et al.*, 2012). This highlights the potential for epigenetic marks to be erased after the replication phase of the cell cycle.

Another histone variant, Cenp-A, is substituted for histone H3 in centromeric chromatin (Sullivan *et al.*, 1994, Hake and Allis, 2006). Cenp-A shows a degree of sequence homology with H3.1 (~60%), but is incorporated very specifically into centromeric regions by recognising DNA elements located within the centromeres (Sullivan *et al.*, 1994). It is now thought that Cenp-A, independent of the DNA sequence, is the defining epigenetic mark for centromere identity, the maintenance of which is essential for the proper location and functioning of the centromere (Fachinetti *et al.*, 2013). The complexity of the epigenetic state of cells can provide for an elaborate and sensitive mechanism for the control of gene expression. This potential can be further increased as combinations of epigenetic features communicate, regulate and influence not only each other, but also the transcription of genes.

1.4.6 Chromatin remodelling

As well as the several different types of possible modifications that enzymes can deposit or remove from chromatin, another epigenetic mechanism that can affect gene expression is ATP-dependent chromatin remodelling. This is a process by which nucleosomes are repositioned or removed from discrete loci so as to affect the transcription of the underlying gene (Kwon *et al.*, 1994, De La Serna *et al.*, 2006, Belch *et al.*, 2010, Euskirchen *et al.*, 2012). This is done to allow transcription factors and polymerases to access the DNA and maintains the plasticity of the chromatin structure. The SWI/SNF complex in yeast was one of

the first remodelling enzymes to be described, and is conserved in mammals (Kwon *et al.*, 1994). This protein complex interacts with nucleosomes at their ATPase domains to disrupt the contacts between DNA and histones, in preparation for the removal of the nucleosome from the target region of DNA (Belch *et al.*, 2010, Euskirchen *et al.*, 2012). There is also a role for modified histones in this process, as recent studies have shown that acetylation of histones at specific residues may be marked for eviction and so contribute to a more open chromatin state (Elliott *et al.*, 2013). Crucially, these remodelling factor complexes can also contain histone acetyl- and methyltransferases, highlighting the potential and importance of communication between chromatin-modifying enzymes, histone modifications and the transcription machinery (Xu *et al.*, 2004, Nightingale *et al.*, 2007).

1.5 INTERACTIONS BETWEEN EPIGENETIC MARKS

1.5.1 Epigenetic communication

Communication between epigenetic marks can occur among histone modifications, histone variants and methylated DNA through the concerted actions of chromatin-binding and modifying enzymes (Zippo *et al.*, 2009, Xu *et al.*, 2012). Marks can promote or antagonise each other's actions and maintenance, and can produce carefully controlled feedback regulation mechanisms that provide an extra level of gene regulation (as well as epigenome regulation) that DNA sequence is too rigid to accomplish (Kowenz-Leutz *et al.*, 2010). This example is focused on the C/EBPβ transcription factor, which is methylated at its N terminus by the PRMT4/CARM1 arginine methyltransferase, reducing its binding affinity with the SWI/SNF chromatin remodelling complex and downregulating downstream

processes. Phosphorylation of the C/EBPβ enzyme reduces its interaction with PRMT4/CARM1, reducing arginine methylation and leading to adipogenic differentiation led by the enhanced recruitment of the SWI/SNF complex.

Often the mechanisms involved with the careful regulation of transcription involve many elements of epigenetic information, such as the double strand break repair pathway, which requires H2A.Z, chromatin remodelling enzymes and proper histone methylation and acetylation to function properly (Rossetto *et al.*, 2010, Xu *et al.*, 2012).

1.5.2 Mechanisms of epigenetic crosstalk

Crosstalk between epigenetic marks is vital for the proper function of the eukaryotic genome. The varied array of modifications to chromatin must be controlled carefully and be appropriate to the context of the cell in order to fulfil their functions. UHRF1 is involved in the methylation of DNA by DNMT1, and can also interact with several different histone modifications and histone-modifying enzymes (Bostick *et al.*, 2007, Unoki *et al.*, 2009, Bronner, 2011). UHRF1 is a target of E2F, a transcription factor that promotes progression of the cell cycle, and has a critical role in the maintenance of DNA methylation by acting as a chaperone for DNMT1 to bind to hemimethylated sites (Unoki *et al.*, 2004, Unoki *et al.*, 2009, Oda *et al.*, 2013). This observation confirms the existence of enzymes that can read complex epigenetic states and alter cell behaviour accordingly. UHRF1 also influences heterochromatin formation and ubiquitylation of histone H3, as well as binding to di- and tri-methylated H3K9 and HDAC, to form a

network of links between many different epigenetic marks, linking chromatin structure with histone modifications (Unoki *et al.*, 2009).

Another example of epigenetic communication is cross talk between modified histone tail residues. This can be illustrated by the interaction between tri-methylated lysine 9 of histone H3 (H3K9me3) and the adjacent phosphorylated serine 10 (H3S10P) (Rea *et al.*, 2000, Nightingale *et al.*, 2006). H3K9me3 is associated with heterochromatin, and is equated with being a marker for silent areas of DNA (Puschendorf *et al.*, 2008). HP1 protein, which is involved with the formation and maintenance of a heterochromatic state, is recruited by H3K9me3, causing repression of the underlying sequence (Fanti and Pimpinelli, 2008). However, during mitosis, the adjacent H3S10 residue is phosphorylated. This disrupts the interaction between H3K9me3 and HP1, by interfering with the binding site of HP1 to the H3 histone tail and thereby reversing its inhibitory effect in order for the DNA to properly segregate during mitosis (Fischle *et al.*, 2005, Papamokos *et al.*, 2012). This phospho switch underlines the way in which histone modifications can influence each other, and how the functions of certain modifications can be altered or nullified due to subtle variations in the epigenetic landscape.

The phenomenon of histone modifications interacting with each other is not limited to adjacent residues, however. Forneris *et al.* discovered that binding affinities of the histone demethylase LSD1 was greatly affected by histone modification (Forneris *et al.*, 2005). LSD1 demethylates H3K4 to have a negative effect on underlying gene transcription. However, in this case, acetylation of H3K9 reduces the activity of the LSD1 enzyme, and phosphorylation of H3S10 inhibits the demethylase action entirely (Forneris *et al.*, 2005). Other histone demethylase

enzymes are similarly affected by nearby non-target post-translational histone modifications (Lohse *et al.*, 2013). These can be seen as examples of an enzyme reading the 'histone code'.

Chromatin associated proteins can also have their affinities for modified histones altered by the presence or absence of methylated DNA. Bartke *et al*. identified a group of proteins that are up or down-regulated, as a result of CpG island DNA methylation (Bartke *et al.*, 2010). A number of proteins were identified that bind to histones only when underlying CpGs are methylated. Others are prevented from binding to their target sites in chromatin by the presence of methyl-CpG. In this study, several genes were affected by changes in CpG DNA methylation, which indicates that this form of epigenetic communication could be widespread and offer control of a large part of the epigenome (Bartke *et al.*, 2010).

Histone modifications are also intimately linked with the process of transcription initiation, elongation and termination (Li *et al.*, 2011). It has been found that a carefully controlled concert of histone methylation and acetylation is associated with the progress of the transcription machinery; histone marks such as H4K20me1 and H3K36me3 are present at the transcription elongation stage and may have a function to do with exon splicing, so can generally be found in the body of genes (Li *et al.*, 2011, Zhu *et al.*, 2013). These combinations are not definitive, however, as genes with CpG island-containing promoters tend to display H3K36 monomethylation in their gene bodies, much closer to the transcription start sites (TSS) than for trimethylation (Vavouri and Lehner, 2012). Transcription initiation is marked by promoter or enhancer associated histone marks such as H3K4me1, 2, 3 and H3K27me3 (Birney *et al.*, 2007). The relationship between histone modifications is a reciprocal one, as the same marks

that direct the polymerase's elongation are disrupted once the enzyme is stalled artificially (Li *et al.*, 2011). It may be the case that the chromatin landscape is disrupted by RNA polymerase, just as the DNA double helix is disrupted, by the rapid changing of histone methylation states caused by the proximity and action of the polymerase enzyme (Le Martelot *et al.*, 2012).

1.5.3 The epigenetic/histone code hypothesis

Some better-studied marks have well established effects on the transcriptome and other cellular functions. For example, it is known that H3K4me3 has a broadly activating effect on underlying genes, and H3K27me3 and H3K9me2 have a repressive effect (Pauler *et al.*, 2009, Lloret-Llinares *et al.*, 2012). It is also known that H3K9me3 and H4K20me3 are associated with heterochromatin (Kouzarides, 2007, Bannister and Kouzarides, 2011, Honda *et al.*, 2012). However, it is becoming clear that histone modifications should no longer be thought of as simply promoting transcription, marking heterochromatin or other singular functions, but as part of a landscape of modified chromatin that is uniquely tailored to the specific context of the locus, the need for its transcription, its packaging status and so on (Mattout *et al.*, 2011, Le Martelot *et al.*, 2012, Abraham *et al.*, 2013). A much more holistic view is required, and the complex array of chromatin binding enzymes that can distinguish subtle differences in the epigenetic landscape of a genomic region indicate that the histone and epigenetic codes are capable of controlling many aspects of the cell's genome and epigenome (Kouzarides, 2007, Bartke *et al.*, 2010, Bannister and Kouzarides, 2011).

What is not currently known is the combined effect of a set of histone modifications, histone variants and/or DNA methylation on gene transcription. This concerted action on chromatin is discussed in the histone code, and the epigenetic code, hypothesis (Strahl and Allis, 2000, Turner, 2000, Turner, 2007). According to this theory, the epigenetic state of a cell, however complex, predicts the gene transcription of the associated DNA (Nightingale *et al.*, 2006, Turner, 2007). Like the genetic code, it is thought to be heritable, passing its information down through cell generations, such as that observed for the inheritance of heterochromatic marks (Hathaway *et al.*, 2012). However, in order to stably pass epigenetic information on, it must pass through the greatest challenge to its longevity; the eukaryotic cell cycle, and more specifically the replication fork of S phase.

1.6 EPIGENETIC INHERITANCE

1.6.1 Inheritance of epigenetic marks

Translation of histone proteins from their respective mRNAs results in an unmodified single histone protein (Leffak *et al.*, 1977). These are then assembled into dimers (H3-H4, H2A-H2B), then these dimers bind to each other and are integrated into a new strand of chromatin at S phase (Gruss *et al.*, 1993, Probst *et al.*, 2009). It may be that newly synthesised H2A-H2B dimers bind to old H3-H4 dimers in a semi-conservative manner, or entirely new nucleosomes could be produced to package the newly synthesised DNA strand (Jansen *et al.*, 2007, Xu and Zhu, 2010). Epigenetic information is added and removed from newly synthesised histones throughout this process, but according to the epigenetic

code hypothesis, the marks on old chromatin should be copied and transferred onto the newly synthesised strand (Turner, 2000). Several mechanisms have already been described that confirm this theory, one of which is the faithful reproduction of methylated cytosines onto the new DNA double helix (Bostick *et al.*, 2007, Athanasiadou *et al.*, 2010).

1.6.2 Inheritance of CpG methylation

Semi-conservative replication of DNA causes the two new strands of DNA to bind not to each other, but to the old strands (Meselson and Stahl, 1958). This results in methylated cytosine residues being methylated on one strand of a CpG sequence, but not on the opposite side: a hemimethylated site (Pradhan *et al.*, 1999). This site is recognised by UHRF1, a protein that recognises methylated DNA and also binds DNMT1, which proceeds to modify the unmethylated DNA so that the mark is copied to the new strand of DNA (Bostick *et al.*, 2007). DNMT1 has an approximately 40-fold affinity preference for hemimethylated sites, and so as long as DNMTs are properly expressed, these enzymes can efficiently maintain this epigenetic mark across generations (Reik *et al.*, 2001, Goyal *et al.*, 2006).

1.6.3 Inheritance of histone modifications

The mechanism for the re-establishment of histone modifications has remained somewhat cryptic. It is known that half of histones on newly formed chromatin are taken from pre-replication chromatin, and half are drawn from newly synthesised

histone proteins, but it is not fully known how this is done (Probst *et al.*, 2009, Jasencakova *et al.*, 2010). Newly synthesised histones are very quickly acetylated at lysines 5 and 12 whilst still being processed in the cytoplasm (Loyola *et al.*, 2006). These are removed after incorporation into chromatin, however, so as not to interfere with the previous epigenetic signature (Xu and Zhu, 2010). How previously expressed histone modifications are faithfully replicated on new chromatin is not clear. It has been discovered that molecules such as PCNA and DNMT1 can recruit chromatin modifying enzymes to newly replicated DNA, implying that there is a mechanism for making sure replicated chromatin modifications match the old, but so far there is a paucity of knowledge as to the association of histone modifications with chromatin through the entire cell cycle (Lehnertz *et al.*, 2003, Probst *et al.*, 2009, Bartke *et al.*, 2010). Some of the specific examples of epigenetic inheritance discovered thus far are described below.

H3K9me3 is known to be maintained due to a positive feedback loop between its methyltransferase (SuVAR39) and HP1, where HP1 recruited by pre-replication SuVAR39 then contributes to the re-establishment of methylated H3K9 after replication (Fodor *et al.*, 2010, Haldar *et al.*, 2011). An analogous system is thought to exist for H3K27me3, where its methyltransferase complex (PRC2) persists throughout DNA replication and facilitates the deposition of H3K27me3 on daughter strands of chromatin (Hansen *et al.*, 2008, Santos *et al.*, 2010). These observations provide evidence of repressive methylation maintenance in mammalian cells through epigenetic means.

There are some clues as to the inheritance of histone modifications; specifically, in terms of the localisation of histone marks across the genome. Genome

sequencing studies have been performed that show certain histone modifications are reliably and reproducibly found only in certain genomic locations, aligning themselves with features such as gene transcription start sites (TSSs) and heterochromatic regions including the centromere (Birney *et al.*, 2007, Marks *et al.*, 2009). For example, H3K4me3 is enriched at the TSS of genes throughout the genome, and is less densely distributed at intergenic loci (Guenther *et al.*, 2007, Tian *et al.*, 2011). This mark is inherited at gene promoters for developmental loci through the differentiation process, as a constituent part of the bivalent motif (Radman-Livaja *et al.*, 2010, Sachs *et al.*, 2013).

1.6.4 Polycomb and trithorax motifs

The methylation of H3K27 is affected by the action of the polycomb repressor complex PRC2 (Hansen *et al.*, 2008, Pasini *et al.*, 2010). This generally has the effect of compressing chromatin, along with its partner complex PRC1, which ubiquitylates histone H2A (Landeira *et al.*, 2010). In mammals, this complex has antagonistic qualities to the trithorax group of proteins responsible for H3K4 methylation (i.e. de-compression of chromatin) (Schuettengruber *et al.*, 2009, Voigt *et al.*, 2013). Despite the fact that these two modifications have antagonistic properties, they are often found together at gene loci, rather than within intergenic DNA (Bernstein *et al.*, 2006, Shin *et al.*, 2012). This observation is theorised to be the basis of the bivalent mark. Bivalent loci are marked with both H3K4me3 (in some studies H3K4me2) and H3K27me3 in embryonic stem cells and in developing blastocysts (Bernstein *et al.*, 2006, Sachs *et al.*, 2013, Shin *et al.*, 2012). The genes that these modifications cover are associated with embryo development and differentiation, and as long as both marks are present,

transcription of these developmental genes is repressed (Bernstein *et al.*, 2006, Mikkelsen *et al.*, 2007). However, upon certain cues, these loci lose the H3K27me3 mark, and their expression is up-regulated (**figure 1.7**) (Sha, 2009). As a result, the cell is driven towards a differentiation pathway, and pluripotency is lost (Azuara *et al.*, 2006, Jia *et al.*, 2012). This describes a clear mechanism by which histone modifications affect gene transcription to bring about a change in cell phenotype. Although the tri-methyl K27 modification is lost, the H3K4me3 mark could be said in this example to be a heritable epigenetic mark that also has an important functional role in regulating the development of cells (Azuara *et al.*, 2006).

1.6.5 Inheritance of epigenetic information through differentiation

Differentiation is an ideal model for the study of epigenetics, as it results in cells that are genetically identical to their predecessors, but with contrasting phenotypes. It also involves the dual nature of epigenetic marks, at the same time being heritable predictors of gene expression, and being influenced by external and environmental changes in order to alter gene expression to adapt. The bivalency paradigm highlights the possible resolution of this, by showing how some marks can remain constant as a heritable mark, and how others can be added or lost to confer a degree of transcription control to the epigenome (Bernstein *et al.*, 2006, Mikkelsen *et al.*, 2007, Sachs *et al.*, 2013). This view of such a system is necessarily simplistic, although it does shed some light on how an epigenome-wide system of heritability and transcription control can occur within the same framework. Differentiation involves wide-ranging changes in the profile of gene expression, as not only do tissue-specific genes become up-regulated, but

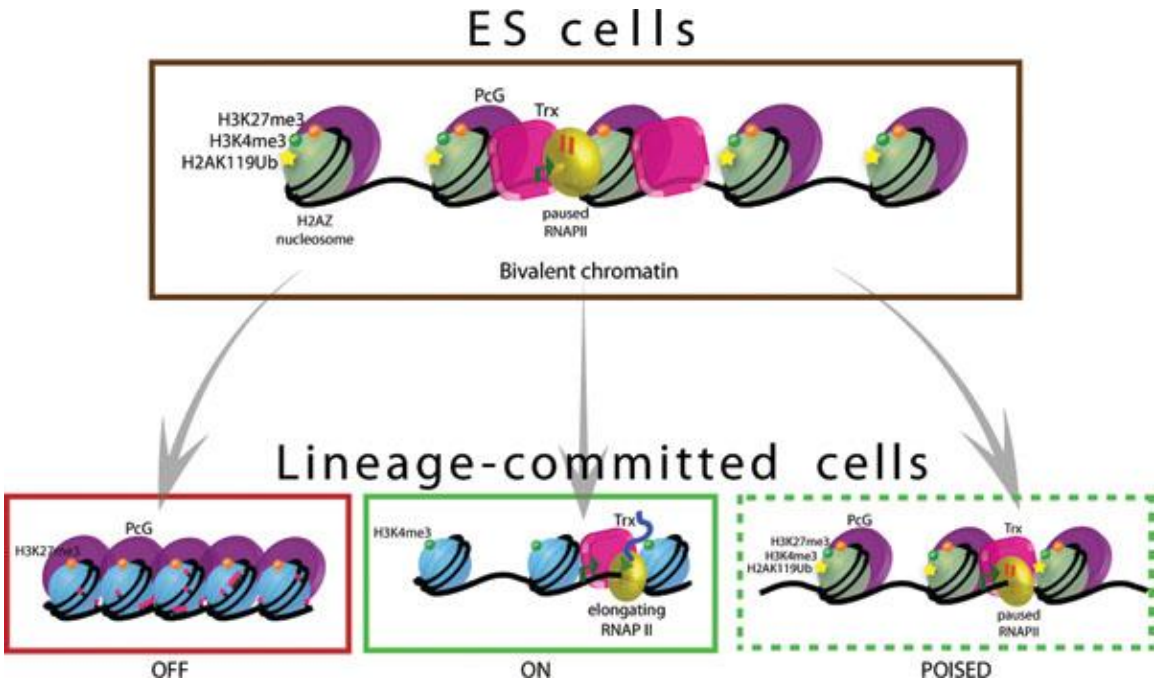


Figure 1.7 Bivalent chromatin domains mark the promoters of developmentally important genes in pluripotent ES cells. PRC2 and TrxG proteins catalyze the tri-methylation of histone H3 on lysine 27 and 4, respectively. PRC1 is also recruited to many of these genes and can mono-ubiquitinylate histone H2A on lysine 119, a modification that is also thought to be important for gene silencing, possibly through blocking RNAPII elongation. As such, bivalent genes are said to be silent, yet poised for activation. H2AZ is highly enriched in a manner that is remarkably similar to PRC2 and may also be an important regulatory component at bivalent genes. Upon differentiation, the bivalent histone marks can be resolved to monovalent modifications in which the gene is “ON” or “OFF”. Bivalent domains can also be maintained or newly established in lineage-committed cells. Taken from Sha and Boyer, 2009.

those genes responsible for maintaining pluripotency are concurrently silenced (Lengner *et al.*, 2007). Profound changes through differentiation can also be observed in other vital processes such as the cell cycle (Andang *et al.*, 2008, Chen *et al.*, 2013a).

Evidence for heritability of histone modifications has been shown in embryonic cells from as early as the 8-cell stage (Vermilyea *et al.*, 2009). It has been shown that increasing the level of acetylation in the 8-cell stage embryo by use of HDAC inhibitors can lead to increased acetylation of histones for several cell divisions up to the blastocyst (ES cell) stage, confirming that altered epigenetic states can be heritable, even in early embryonic cells that are rapidly differentiating (Vermilyea *et al.*, 2009).

ESCs classically have an abbreviated cell cycle, particularly in G_1 (Fujii-Yamamoto *et al.*, 2005). After differentiation, this changes to include a much more obstinate G_1/S checkpoint block (Chen *et al.*, 2013a). The mechanism for bypassing part of the restriction point in ESCs is lost, and the up-regulation of cell cycle related genes have the effect of stalling cells for longer at G_1 (Chen *et al.*, 2013a). What is less well known is how this process is controlled by epigenetic modifications. Bivalency offers one explanation, and as stated before, it is generally found at developmental loci (Azuara *et al.*, 2006, Bernstein *et al.*, 2006). There are other studies that have provided insight into certain epigenetic marks that have been observed to change in concert with each other upon differentiation (Serandour *et al.*, 2011). However, there is less information on how the crucial processes of epigenetic modification and changes in the cell cycle are linked throughout differentiation. As the cell cycle is so central to the maintenance of pluripotency,

epigenetic processes are critical in contributing to this, and in particular to the G_1/S block.

The factors that maintain pluripotency are well-known and discussed above. Both *Nanog* and *Pou5f1* genes are known to be involved with the maintenance of bivalent histone modifications (Bernstein *et al.*, 2006, Adamo *et al.*, 2011). This is critical to the control of pluripotency, as once the bivalent mark is lost; the cell will begin to differentiate. A mechanism has been elucidated whereby the levels of the bivalent histone modifications are kept in check by the action of an enzyme called Utf1 (Jia *et al.*, 2012). This factor limits methylation of H3K27 and concurrently labels the underlying mRNAs for destruction, ensuring a balanced, or 'poised', state that maintains the pluripotency of the cell, and is down-regulated at the onset of differentiation (Jia *et al.*, 2012).

1.7 GENOMIC DISTRIBUTIONS OF HISTONE MODIFICATIONS

The vast array of histone modifications present on eukaryotic chromatin can present a wealth of information to any chromatin-binding proteins that can read the histone/epigenetic code (Strahl and Allis, 2000, Turner, 2007, Turner, 2000). One of the important questions remaining in the field today is whether we can reach the breadth of knowledge required to be able to read the code ourselves, and even to manipulate it to prevent or hinder damaging conditions such as cancer. In order to properly answer this question, we must know the distributions of each chromatin modification in the epigenome.

Histone modifications are not deposited blindly across the entire genome; they are specifically targeted to various regions of chromatin depending on the function of the mark (Bloushtain-Qimron *et al.*, 2009). For example, the H3K4me3 modification can be found within gene-coding regions, particularly around the TSS of the gene (Guenther *et al.*, 2007, Lloret-Llinares *et al.*, 2012). The distributions of these histone modifications reveal much about their activities and interactions. As well as the example above, there are other histone modifications that are restricted only to densely packaged and transcriptionally inactive chromatin (Schotta *et al.*, 2004, Gonzalo *et al.*, 2005, McManus *et al.*, 2006, Puschendorf *et al.*, 2008). H3K9me3 can be found in abundance at the centromeres of mammalian cells, and also at non-coding and silenced sites within the euchromatic areas of the genome (Puschendorf *et al.*, 2008, Alder *et al.*, 2010).

The changes in these distributions throughout differentiation can determine the lineage or fate of the cell (Breiling *et al.*, 2004, Hathaway *et al.*, 2012, Hattori *et al.*, 2013). The upregulation of genes involved in lineage-specification, and the downregulation of genes involved in pluripotency maintenance are controlled by epigenetic processes, and so are the distributions of histone modifications changes. Some developmental genes are known to be regulated in clusters, for example the *Hox* clusters (Lewis, 1978, Durston *et al.*, 2011). This makes it easier for genes in the cluster to be regulated all at once, so that a generally permissive or restrictive chromatin environment can be enforced upon the entire region. When development proceeds, the repressive histone modifications are removed and transcription and differentiation proceeds (Larson and Yuan, 2010, Kashyap *et al.*, 2011). This process has also been linked to control by bivalent motifs as described above (Kashyap *et al.*, 2011). The dynamic and intricately controlled nature of

histone modifications throughout differentiation highlights the importance of the understanding of the distributions of critical histone marks throughout the genome and epigenome.

1.8 AIMS

In order to understand the effects and activities of histone modifications, we must first determine where they are, and how they change over time. In biological terms, time can be defined in terms of the number of cell divisions, and so histone modifications either remain stable throughout several cell divisions, or are re-distributed as a response to environmental stimuli. The progress of differentiation involves wide-reaching changes to the expression profile of the cell, and so must involve some changes to the epigenetic landscape of the nucleus. In order to test these principles, I aim:

- To determine the distributions of several key histone modifications in metaphase mouse embryonic stem cells, and crucially to find out whether these distributions change after differentiation, exploring one aspect of epigenetic heritability.

- To produce a full epigenetic karyotype of a few of these histone modifications to correlate their distributions with specific chromosomes, and to compare and contrast activating and repressive histone modifications and their similarity to data for asynchronous cells. This will provide insight into another aspect of histone modification inheritance, the transition from interphase to metaphase.

- To go further into the exploration of histone modification distributions by examining a more focused area of the genome, namely the *Hox* clusters. This will

allow the discrimination of histone marks at the single gene scale, leading to a comprehensive picture of chromatin environment at a discrete chromatin region. Cells will be separated into asynchronous and mitotic populations to probe the link between epigenetic modifications and the cell cycle.

- To determine what epigenetic changes occur at the developmental gene cluster *Hoxa* during the differentiation process and whether these are also affected by the progression of the cell cycle.

2. MATERIALS AND METHODS

2.1 TISSUE CULTURE

2.1.1 Tissue culture of embryonic stem cells

CCE/R (Wild-type) and OS25 (Billon *et al.*, 2002) murine embryonic stem cells were grown in gelatinised (0.1%, Sigma) T25 and T75 vented flasks (Sarstedt) at 37^{o}C in 5% CO_2 in DMEM supplemented with the following:

20% foetal bovine serum; 2mM L-glutamine; penicillin (100units/ml) /streptomycin (100µg/ml); 100x non-essential amino acids at 1% v/v, 125µM 2-mercaptoethanol (all Invitrogen) and leukaemia inhibitory factor LIF (ESGRO 1000 units/ml, Millipore).

Cells were passaged once confluent, usually in a 1:6 ratio, every two days, and were supplemented with 50% new medium on the non-passage days. To passage, undifferentiated cells were trypsinised for 5 mins using 0.5% Trypsin/EDTA (Invitrogen) and centrifuged at 1200rpm for 5mins (MSE Mistral 2) in 15ml or 50ml centrifuge tubes (Falcon). The pellet was then resuspended in 10ml of the above ES cell medium. Cells were then put into fresh gelatinised flasks. Harvesting of cells was accomplished by trypsinising as above and resuspending cells in 1xPBS (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 2mM KH_2PO_4, 5mM Sodium Butyrate, pH adjusted to 7.4 (all Sigma)). The cell pellet was then washed twice in PBS by centrifugation (1200rpm/10min) and resuspended in PBS.

2.1.2 Tissue culture of differentiating cells and fibroblasts

OS25 and CCE/R cells were encouraged to differentiate by trypsinising as above, and then plating them into 7cm non-adherent dishes (Sterilin) in the absence of LIF. Plates were incubated at 37^{o}C in 5% CO_2 and 50% media was replaced every day. Two days after plating, 1µM retinoic acid was added to further promote differentiation. Harvesting of cells was done by trypsinising for 3mins, then transferring to centrifuge tubes (Falcon), centrifuging at 1200rpm for 5 minutes and washing in PBS as above.

Mouse embryonic fibroblasts were grown in gelatinised (0.1%, Sigma) vented flasks (Sarstedt) at 37^{o}C in 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) containing:

10% foetal bovine serum; 2mM L-glutamine; penicillin (100units/ml) /streptomycin (100µg/ml); 100x non-essential amino acids at 1% v/v and 125µM 2-mercaptoethanol (all Invitrogen).

2.1.3 Cryopreservation of cells

Undifferentiated cells to be cryopreserved were harvested as above. After washing twice with ES cell medium, pellet was resuspended in FBS (Invitrogen) containing 10% Dimethyl sulphoxide (DMSO, Sigma). This was then divided into 1ml aliquots in cryo vials (Sigma) and the vials placed in a freezing container (Mr. Frosty, Sigma) at -80^{o}C overnight. The vials were then transferred into a Dewar containing liquid nitrogen at -196^{o}C for storage. Cells brought up from liquid nitrogen storage were thawed at room temperature and then resuspended in 10ml

ES cell medium. This solution was centrifuged at 1200rpm for 10min (MSE Mistral 2). The pellet was washed twice in ES cell or MEF medium and finally resuspended in 5ml. This was then added to a gelatinised T25 vented flask (Sarstedt) and kept in a CO_2 incubator as above (37^oC, 5% CO_2).

2.2 FLOW CYTOMETRIC ANALYSIS OF CELL POPULATIONS

Cells used for flow cytometry were trypsinised and harvested as above. Cells were fixed in 5ml ice cold ethanol and centrifuged at 800rpm for 10min (MSE 2000R). The pellet was washed three times in PBS and resuspended in 2ml PBS in polypropylene tubes (Falcon). An equal volume of 1xPBS/1% Triton X-100 (Sigma) was added to disaggregate the cells, 20 µl RNase A and 20µl Propidium iodide (10mg/ml, Sigma) were added to stain the nuclei. The sample was analysed using a flow cytometer and its associated software (Beckman Coulter FACSCalibur). Gates were determined specifying the populations of 2n (G_1 phase cells), 4n (G_2/M phase cells) and 2-4n (S phase cells). Graphs were produced by plotting cell counts (y-axis) against fluorescence (x-axis). For an example of the resulting plot, see **figure 2.1**.

2.3 CHROMOSOME IMMUNOFLUORESCENCE

2.3.1 Isolation of nuclei and chromosome spreading

5µl colcemid (10µg/ml, Invitrogen) per ml of medium was added to cultured mouse cells (in 25cm^3 tissue culture flasks). Cells were incubated at 37^oC for 2½ hours

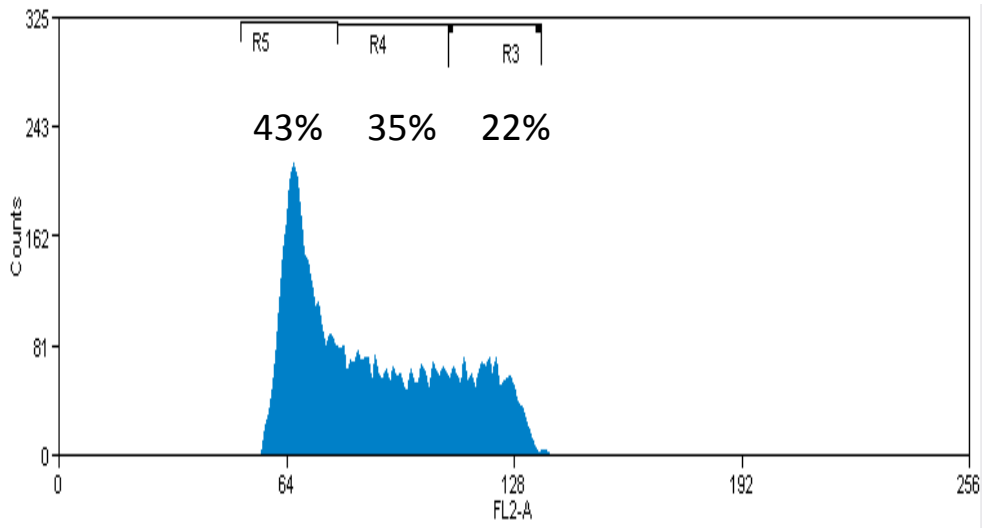


Figure 2.1 FACS results from asynchronous differentiating (d4) OS25 cells treated with Propidium iodide. The graph shows gating for three distinct cell cycle populations. R5 refers to the G_1 population, R4 to the S phase population, and R3 contains cells in G_2 and mitosis phase, as shown by the fluorescence (X-axis) being nearly double that of the G_1 cells (cell count on Y-axis). Shown above the graph are the proportions of each population relative to the total cell count.

and shaken off. Cells were centrifuged at 1200rpm for 10 minutes at 4^{o}C (MSE 2000R), and then washed twice in ice-cold 1xPBS/ Sodium Butyrate pH 7.4. Cells were counted and then resuspended at 10^5 cells per ml in 0.1M KCl (at least 2 biological replicate experiments). The suspension was left at room temperature for 10 minutes, and then 2x10^4 cells were pipetted into a cytospin chamber attached to an ethanol-washed glass slide (VWR). The chamber was centrifuged at 1800rpm (Thermo CytoSpin 4) for 10minutes at room temperature. The chamber was removed and the glass slide was immersed in KCM buffer (120mM KCl, 20mM NaCl, 10mM Tris pH 8.0, 0.5mM EDTA, 0.1% v/v Triton X-100, all Sigma) for 10 minutes.

2.3.2 Chromosome indirect immunofluorescence

Primary antibodies (see **table 2.1**) raised in rabbit against specific histone modifications was diluted in 1% bovine serum albumin (BSA, Sigma) in KCM buffer, and 50μl was then added onto the slide. Commercial antibody specificity data is shown in figures **2.2** and **2.3**. In-house antibodies have been tested previously and demonstrated to be specific to target modified residues. The slide was humidified in the dark at 4^{o}C for 1 hour. The slide was then washed twice in KCM for 10 minutes each. 50μl goat anti rabbit FITC (diluted x50 in 1% BSA/KCM) was added, the slide was then humidified again for 1 hour. The slide was then washed twice in KCM and then immersed in 4% Formaldehyde in KCM for 10 minutes exactly. The slide was then rinsed in distilled water and mounted on an ethanol-washed cover slip with 1μg/ml DAPI (Invitrogen) in Vectashield mounting medium (Vector Labs). The cover slip was sealed using nail varnish and the slide

Antibodies used in immunofluorescence experiments

Antibody	Dilution factor
H3K4me1 (in-house SAS cut)	100x
H3K4me2 (in-house SAS cut)	100x
H3K4me3 (in-house SAS cut)	100x
H3K9me1 (Abcam)	300x
H3K9me2 (Millipore)	400x
H3K9me3 (Millipore)	600x
H3K9ac (in-house SAS cut)	100x
H3K27me3 (Millipore)	600x
H3K27ac (Abcam)	300x
H4K12ac (in-house SAS cut)	400x
H4K16ac (in-house SAS cut)	200x
H4K20me3 (Abcam)	400x

Table 2.1 Sources and dilutions of antibody added to each reaction in the immunofluorescence procedure. Antibodies were diluted in 1% BSA in KCM buffer.

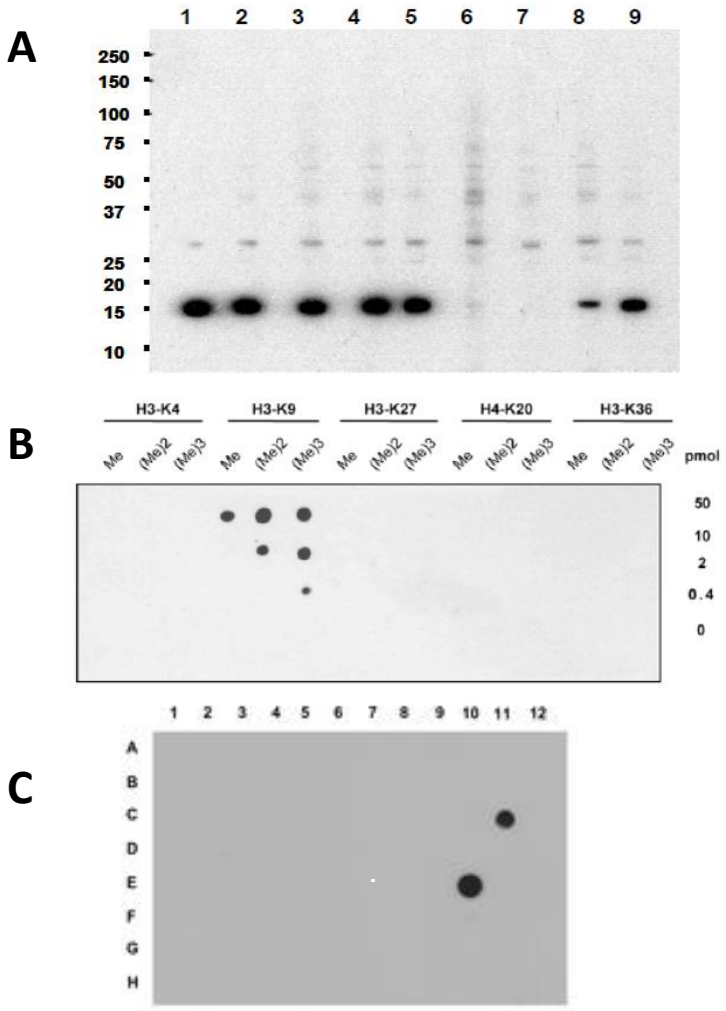


Figure 2.2 Commercially acquired antibody specificity (Millipore):
A. Western Blot Analysis and Peptide Inhibition: Representative blot HeLa Acid extract were resolved by electrophoresis, transferred to nitrocellulose and probed with anti-dimethyl-Histone H3 (Lys9) (1:500, Lane 1) or preincubated with 0.4uM Histone H3 peptide with following modifications: Lane 2: Linear non-modified, Lane 3: Branched non-modified, Lane 4: Branched trimethyl, Lane 5: Linear trimethyl, Lane 6: Branched dimethyl, Lane 7: Linear dimethyl, Lane 8: Branched monomethyl, Lane 9: Linear monomethyl Proteins were visualized using a goat anti–rabbit secondary antibody conjugated to HRP and a chemiluminescence detection system.
B. Dot Blot Analysis: Dot-blot analysis demonstrating specificity of anti-H3K9me3 for trimethyl Lys9 of histone H3. Synthetic peptides containing various histone modifications (0-50 pmol) were spotted onto PVDF. The membrane was incubated with a 1:100 dilution of a previous lot of anti-trimethyl-Histone H3 (Lys9), clone 6F12-H4, and processed for Western blotting.
C. Histone peptides with various modifications were probed with Anti-trimethyl Histone H3 (Lys27) (1:500 dilution). Proteins were visualized using a Donkey Anti-Rabbit IgG secondary antibody conjugated to HRP and a chemiluminescence detection system. C11: KAAR[trimethyl-K]SAPA-C E10: KAAR[trimethyl-K]SAPSTGGVKKC.
Figures and data taken from Millipore website: www.emdmillipore.com.

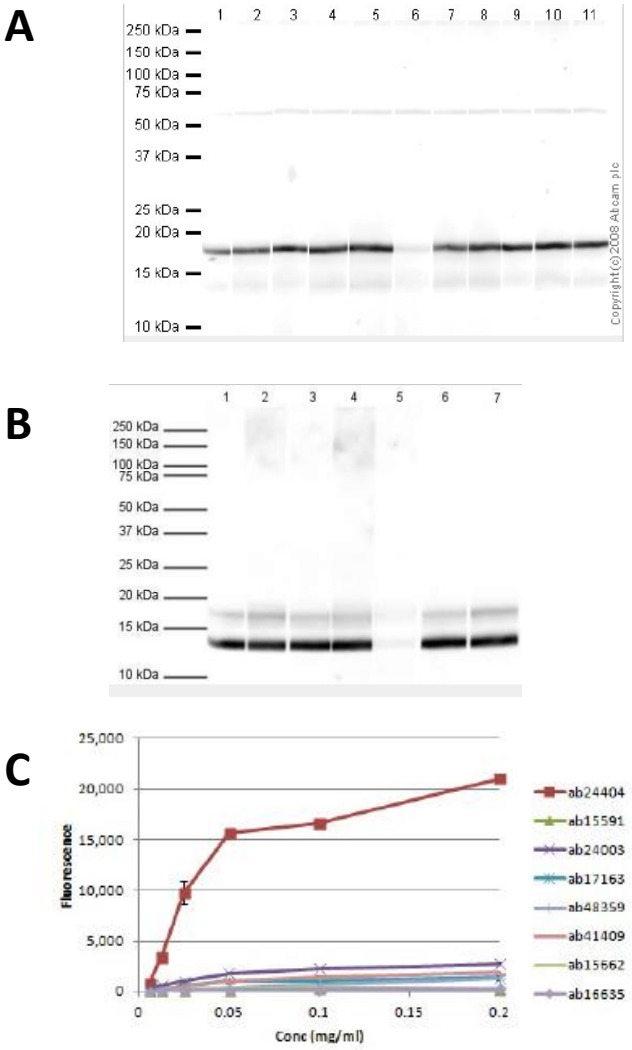


Figure 2.3 Commercially acquired antibody specificity (Abcam): Western blot blocking experiments -

A. *All lanes* **:** Anti-Histone H3 (mono methyl K9) antibody - **Lane 1 :** Calf Thymus Histone Preparation Nuclear Lysate **Lane 2 :** Human Histone H3 (unmodified) **Lane 3 :** Human Histone H3 (mono methyl K4) **Lane 4 :** Histone H3 (di methyl K4) **Lane 5 :** Human Histone H3 (tri methyl K4) **Lane 6 :** Human Histone H3 (mono methyl K9) **Lane 7 :** Human Histone H3 (di methyl K9) **Lane 8 :** Human Histone H3 (tri methyl K9) **Lane 9 :** Histone H3 (mono methyl K27) **Lane 10 :** Human Histone H3 (di methyl K27) **Lane 11 :** Human Histone H3 (tri methyl K27) Lysates/proteins at 0.5 µg per lane.

B. *All lanes* **:** Anti-Histone H4 (tri methyl K20) antibody at 1 µg/ml. **Lane 1 :** Human Histone H4 (unmodified) peptide **2 :** Human Histone H4 (unmodified) peptide **3 :** Human Histone H4 (mono methyl K20) peptide **4 :** Human Histone H4 (di methyl K20) peptide **5 :** Human Histone H4 (tri methyl K20) peptide
6 : Human Histone H3 (tri methyl K4) peptide **7 :** Human Histone H3 (tri methyl K9) peptide Blocking peptides at 1 µg/ml per lane.

C. Peptide Array against peptides to different Histone H3 modifications. Six dilutions of each peptide are printed on to the Peptide Array in triplicate and results are averaged before being plotted on to a graph. Results show strong binding to Histone H3 acetyl K27 peptide (ab24404), indicating that this antibody specifically recognises the Histone H3 acetyl K27 modification. **ab24404** - H3 acetyl K27 **ab15591** - H3 acetyl K14 **ab24003** - H3 acetyl K18 **ab17163-** H3 unmodified **ab48359-** H3 acetyl K23 **ab41409** - H3 acetyl K36 **ab15662** - H4 acetyl K12 **ab16635** - H3 acetyl K9. Figures and data taken from Abcam website www.abcam.com

was stored in a dark box. Slides were photographed using an Axioplan2 microscope (Zeiss) at x100 magnification in Smartcapture software (Digital Scientific UK). Images were analysed using ImageJ (NIH) and SmartType (Digital Scientific UK) software.

2.3.3 Fluorescence In Situ Hybridisation slide preparation

5µl Fluorescence In Situ Hybridisation (FISH) mouse chromosome paint DNA probe (Cambio StarFISH) was added to 5µl StarFISH hybridisation buffer (Cambio) and denatured at 65^{o}C for 10min, then incubated at 37^{o}C for 2.5 hours. Cover slips were removed from previously immunostained slides, and were immersed in 70% formamide/SSC (0.3M NaCl, 0.03M sodium citrate, pH 8) at 75^{o}C for 10min. Slides were progressively dehydrated for 2min each in 75%, 90% and 100% methanol, and dried on a heating block. 10µl FISH probe was added to the slide and the cover slip was attached using rubber cement. The slide was humidified at 37^{o}C overnight. The cover slip was carefully removed and the slide immersed in 1xSSC at 75^{o}C for 30min. The slides were washed twice at room temperature in PN buffer [100 mM Na_2HPO_4, 50 mM NaH_2PO_4, 0.1% Triton X-100 (v/v)] for 10min each, then 1µg/ml DAPI in Vectashield medium (Vector Labs) was added, and then a cover slip was applied and sealed. Images were obtained as above. Two biological replicates were performed, each slide containing many chromosome spreads so that n≥8 as reported above.

2.3.4 Computational analysis of chromosome immunofluorescence

To produce numerical results from the images obtained from the immunofluorescence procedure, ImageJ software was used (NIH). The freeform

line tool was used to trace a path along the selected chromosome arm, and the line was widened to fit as closely as possible the width of the chromosome, whilst still staying inside its boundaries. The software calculated the average FITC fluorescence at each pixel along the length of the chromatid. Another line was drawn in an empty section of the image in order to provide an average background reading. This value was subtracted from the other results, and then the results were normalised as a percentage of the maximum observed fluorescence for that image. Using this method, it was possible to gauge the changes in fluorescence across chromosome arms. This was repeated for each chromosome and chromosome arm under consideration (n= at least 4 chromosomes, or 8 chromosome arms). The relative lengths of the chromosomes were normalised by using the NCBI Map Viewer to obtain the length of each chromosome in terms of base pairs. The data was then segmented into 10 megabase tranches, so that values obtained for each pixel were allocated to a discrete data bin. This way, images of chromosomes displaying different physical lengths were normalised to the same genetic length. A histogram chart was produced from this data consisting of the mean relative fluorescence of a particular 10 megabase section of a chromosome represented by a bar, and the standard deviation between replicates represented by the error bars. This was then compared to publicly available NCBI sequence data, namely gene density, CpG density and repeat density. These were obtained by counting the number of genes, CpGs and repeats in each 10 megabase window and representing them as bars on a histogram in a similar way to my own data. Error bars are not present on these graphs, as they are taken from a single source (NCBI website, chromosome viewer). These sets of data cannot be compared statistically as there are

insufficient replicated to perform parametric statistical tests, and the use of non-parametric tests would not yield statistically significant results.

2.4 NATIVE CHROMATIN IMMUNOPRECIPITATION

2.4.1 Chromatin preparation and purification for N-ChIP

Cultured CCE/R cells were trypsinised (Invitrogen) and washed 3 times in ice cold 1xPBS/Sodium butyrate. Cells were resuspended at $2x10^7$ cells per ml PBS. The final cell pellet was resuspended in 1 x TBS (1mM Tris-HCl, 15mM NaCl, 3mM $CaCl_2$, 2mM $MgCl_2$ and 5mM sodium butyrate, all Sigma) at a concentration of 2 x 10^7 cells per ml. An equal volume of 1% Tween40 (Sigma) diluted in TBS and $1/200^{th}$ of the volume 0.1M PMSF (Sigma) was added. The cell suspension was stirred on ice for 1 hour. The sample was homogenised using a hand operated, all-glass homogeniser with a tight pestle (Dounce). After 7 smooth strokes the homogeniser was placed on ice until the suspension cleared (usually 1-3 minutes). During the cooling/settling period, release of nuclei was checked by examining a small aliquot under the microscope in a standard counting chamber. Homogenisation continued until 75% of cells yielded intact nuclei. Following homogenisation, nuclei were centrifuged at 2000rpm for 10 minutes at 4^oC (MSE 2000R). Nuclei were resuspended in 1ml of 5% sucrose (Sigma)/TBS pH 7.4 with 0.1mM PMSF and centrifuged at 2000rpm 10 minutes at 4^oC (MSE 2000R). Nuclei were resuspended in 1ml Digestion buffer pH7.4 (0.32 M Sucrose, 50 mM Tris-HCl (pH7.4), 4 mM $MgCl_2$, 1 mM $CaCl_2$, and 0.1 mM PMSF). The amount of chromatin was checked by measuring the A_{260} of an aliquot diluted 20-fold in 0.1% SDS (Sigma), using a light spectrometer (Amersham). Samples were centrifuged

at 2000rpm for 10 minutes at 4^{o}C (MSE 2000R) and pellets resuspended to a chromatin DNA concentration of 0.5mg/ml and divided into 1ml aliquots in Eppendorf tubes. Chromatin was digested with micrococcal nuclease (Sigma) at 50U per 500 µg chromatin for 5 minutes at 37^{o}C. Digestion was stopped by addition of 20µl 0.5M EDTA (Sigma) (to a final concentration of 10 mM) and chromatin placed on ice for 5 minutes. Sample was centrifuged at 4000rpm for 5 minutes (Fresco), supernatant (S1) was removed and kept on ice. Pellets were resuspended and combined in 500µl lysis buffer at pH7.4 (2 mM Tris-HCl (pH7.4), 0.2 mM EDTA, 5 mM Na butyrate, 0.2 mM PMSF) per 1ml solution (if multiple micrococcal nuclease digestions had been performed; then in 1 ml final volume). Pellets were dialysed overnight at 4^{o}C in dialysis tubing (Sigma) against 2 litres of lysis buffer, and the S1 fraction was kept at 4^{o}C. Dialysed chromatin was centrifuged at 2000rpm for 10 minutes at 4^{o}C (MSE 2000R). The supernatant was removed and place on ice (fraction S2). The pellet was resuspended in an equal volume of lysis buffer (fraction P). A_{260} of all samples was checked and distribution of chromatin between the three fractions was calculated. All samples were analysed by 1.2% agarose (Sigma) gel electrophoresis, to ensure that a nucleosomal ladder had been precipitated in the S1 and S2 fractions (**figure 2.4**). These fractions were the combined and the final chromatin concentration assayed by A_{260} measurement.

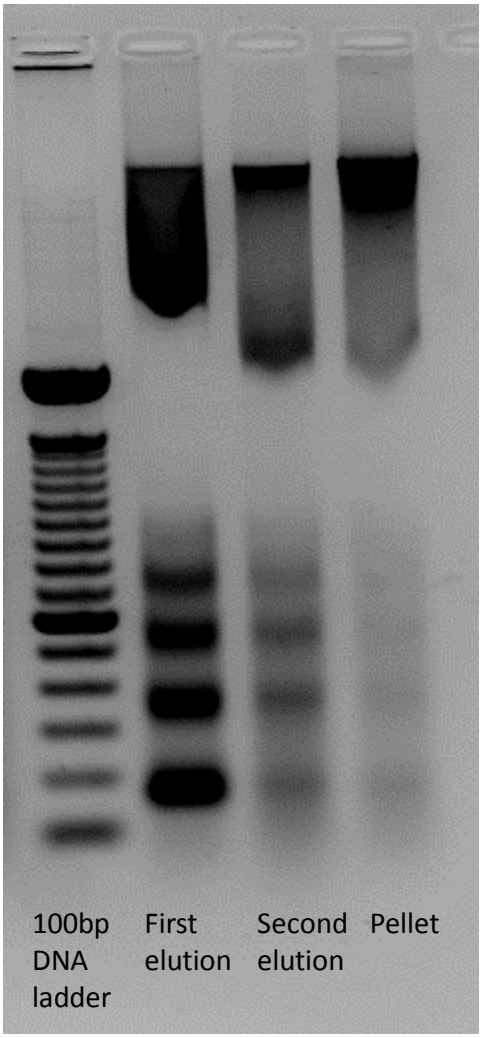


Figure 2.4 Inverted UV photograph showing chromatin resulting from the N-ChIP chromatin extraction procedure. First elution (S1) and second elution (S2) display bands equivalent to multiples of nucleosome lengths (~146-200bp), material remaining in the pellet contains very few oligonucleosomes. Chromatin was run out on a 1.2% agarose gel, stained with ethidium bromide and viewed using a transilluminator.

2.4.2 Native chromatin immunoprecipitation: N-ChIP

Antibodies were added to S1+S2 combined fractions in the quantities below (**see table 2.2**). Antibody solutions were added to 50µg unfixed chromatin and incubation buffer (50 mM NaCl, 20 mM Tris–HCl, pH 7.5, 20 mM Na butyrate, 5 mM Na_2EDTA, 0.1 mM PMSF) was added up to a final volume of 800µl. Half that amount of chromatin was added to 10µl pre-immune serum, as a non-specific control and made up to 400µl with incubation buffer. After overnight incubation on a slowly rotating platform (TAAB) at 4 °C, 200 µl (100µl for PI) pre-swollen protein A–Sepharose (50% v/v slurry, GE Healthcare) was added and the incubation was continued for a further 3 h at room temperature on a fast turntable (Bibby). The antibody–chromatin mixture was centrifuged at 13000rpm for 10 min (Thermo Fresco). Supernatant was removed using siliconised glass Pasteur pipettes (Appleton Woods) and added to 333µl phenol chloroform (25 parts phenol: 24 chloroform: 1 methylbutanol) to constitute the unbound fraction. The protein A–Sepharose pellet was resuspended in 1 ml wash buffer (50 mM Tris–HCl, pH 7.5, 10 mM EDTA, 5 mM Na butyrate) containing 50 mM NaCl, transferred to siliconised 15ml CTs and layered onto 9 ml of the same buffer. After centrifugation at 2000rpm for 10 min at 4 °C (MSE 2000R), supernatant was removed and the pellet washed in 10 ml wash buffer containing 100 mM NaCl and again in 10 ml of wash buffer containing 150 mM NaCl. Samples were centrifuged at 2000rpm for 10 minutes (MSE 2000R) and resuspended in 1ml wash buffer (with 150mM NaCl). Samples were transferred to siliconised Eppendorfs and centrifuged at 13000rpm for 10 minutes (Thermo Fresco). Bound material was eluted from the protein A–Sepharose by addition of 250µl 1% SDS in incubation buffer and incubated for 15 min at room temperature on a fast turntable (Bibby). After

Antibodies used in N-ChIP experiments

Antibody	Amount to add to 50µg chromatin
H3K4me3 (in-house SAS cut)	100µl
H3K9me2 (Millipore)	10µl
H3K9ac (in-house SAS cut)	100µl
H3K27me3 (Millipore)	10µl
H3K27ac (Millipore)	10µl
H4K20me3 (Millipore)	10µl
Pre-Immune Control (in-house from rabbit)	10µl

Table 2.2 Sources and amounts of antibody added to each reaction in the N-ChIP procedure.

centrifugation at 13000rpm for 10 min (Thermo Fresco) the supernatant was taken and added to 333µl phenol chloroform. 250µl 1% SDS in incubation buffer was again added to the pellet After centrifugation at 13000rpm for 10 min (Thermo Fresco) the supernatant was taken and added to the bound fraction. An equal volume of incubation buffer (500µl) was added to the bound fraction to reduce the concentration of SDS to 0.5%. Unbound and bound fractions were vortexed and centrifuged at 2000rpm for 10 min (MSE 2000R) to separate the phases. Supernatant was removed and an equal volume of phenol chloroform added. Samples were vortexed and centrifuged at 2000rpm for 10 min (MSE 2000R). An equal volume of chloroform (24 parts chloroform: 1 methylbutanol) was added, vortexed/centrifuged as before, and the supernatant transferred to a 6ml polypropylene tube (Falcon). Finally the DNA was precipitated at −20 °C using 100µl 4M LiCl (Sigma), 25µl glycogen (2mg/ml, Roche) and ice-cold ethanol (Fisher) up to 5ml.

2.4.3 Verification of DNA following ChIP

Samples were centrifuged at 3000rpm for 20 minutes (MSE 2000R). The pellet was air dried and resuspended in 200µl water, transferred to Eppendorfs and stored at -20^{o}C. PicoGreen analysis of the DNA concentration was performed by adding 2µl DNA to 98µl PicoGreen (Invitrogen) diluted 200-fold in TE buffer: 1mM Tris-HCl pH 8.0, 0.1mM EDTA; in duplicate, and analysed using a spectrophotometer (Victor3 – Perkin Elmer). The percentage pull-down for each antibody was calculated for bound fractions based on a standard curve of known quantities of mouse genomic DNA, and unbound samples were diluted to the same DNA concentration as the bound fraction.

2.4.4 Analysis of N-ChIP samples by Real-time PCR

PCR reactions were performed in triplicate on 384 well plates (ABgene), with mouse genomic DNA control and no template control. Primers (Invitrogen) were designed to specific regions of the mouse genome using Primer3 (http://bioinfo.ut.ee/primer3 and http://www.ncbi.nlm.nih.gov/*tools*/***primer***-***blast***). Primers were designed to be specific to a single target sequence, produce a single dissociation curve peak, have an efficiency of 90-110%, and to have a melting temperature of 60 +/-2 ^{o}C (see **table 2.3**). Primers for the *Hox* cluster were targeted towards the promoters of all 10 *Hoxa* genes, the promoters of the 5' and 3' flanking genes and upstream and downstream intergenic sequences (**figure 2.5**). 0.5µl forward primer, 0.5µl reverse primer, 2µl sterile water and 5µl 2x Quantitect SYBR mastermix (Qiagen) were added to 2µl template DNA for each reaction. The plates were centrifuged and loaded onto the ABIPrism 7900 PCR machine (Applied Biosystems). PCR program was as follows: 95^{o}C for 15mins; then 40 cycles of 94^{o}C for 15 secs, 60^{o}C for 30 secs and 72^{o}C for 30 secs. A melt analysis consisting of 95^{o}C for 15secs, 60^{o}C for 15secs, then a 95^{o}C melt gradient was performed to verify a single product. Analysis of the data was performed using the double delta Ct (cycle threshold) method of relative DNA quantification to compare bound with unbound fractions. This uses the time (in cycles) at which the fluorescence of the product bound to SYBR green reaches a threshold (that is set above the background fluorescence). As SYBR green increases in fluorescence by a factor of 20 when bound to double-stranded DNA, proper amplification of a target will result in a fluorescence signal higher than that of the background. Means of the three replicates are taken, and the unbound fraction

Mouse genomic DNA primer sets

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Tm (°C)	Product length (bp)
Skap2	CGGGCAACGGTACAGGTGGC	CGGAGCGTTCTGATTGGCGC	60	148
Evx1	GGAACTCGCTGCGCGCCTTA	AGGTGTGGGCCTCGCTCGAA	60	143
Hoxa1	GCGCGTCACCTACACTGAG	CGTGACTCTACCAGCCAATG	60	117
Hoxa2	TGAGGCGTTCCTTTCTGACT	TGGCTGTCACATGATTGCTT	60	103
Hoxa3	CAGTTCCTGGAGCAGCCGCC	CTGTCGCTTCGGCAGCCCAA	60	111
Hoxa4	TGGAGGGCCTCGAACCCTCG	GCCCTGTTCCTGCCCACGTC	60	116
Hoxa5	GATCGGCAGCTGACGGCCTC	CGCTTCCGACCTCGGGCTTC	60	126
Hoxa6	ATTTGCTGCTGTCGCTTTTG	CCCTCTGCAGGACTGTGATT	60	113
Hoxa7	GGCCTGAGCAGTTTATGAGG	CTTACCTTGCTGGGCTTCAG	60	129
Hoxa9	AGCTGCGCGATCCCTTTGCA	GTGTCCGCCCGGCAGAACAA	60	122
Hoxa10	TCTTCTGGCCCATCAATACA	CACTCCCAGTTTGGTTTCGT	60	101
Hoxa11	TCAATTTCAACATCGGGTCA	GAAGCGGATCCGTGAAGTAG	60	105
Hoxa13	GCCAGTCCAAGGCCAAGGCA	CGGAGGTTGGGAGGGGCAGT	60	118
Hoxa upstream (intergenic)	TGGAGCTCAGTGTCGAGAGTGTGA	ACTCCATGCTGTTGCCATATCTGC	58	160
Hoxa downstream (intergenic)	AACCCAAGTTGGCTGCCCCATC	ACCCAGGCAAACGTTCGGAGTAA	60	137
Gapdh	TCCTGGGAACCATCACCCGGTC	TGTGCACGCACCAAGCGTGT	60	94
Nanog	CCAATGTGAAGAGCAAGCAA	CCCCCAAAAAGAGGCTTTAC	60	155
Pou5f1	CAACAGGCTTTGTGGTGCGA	CAACCCTTAGGACGGGACCC	60	129

Table 2.3 Sequences, lengths and melting temperatures of primers used during qPCR of material derived from ChIP experiments

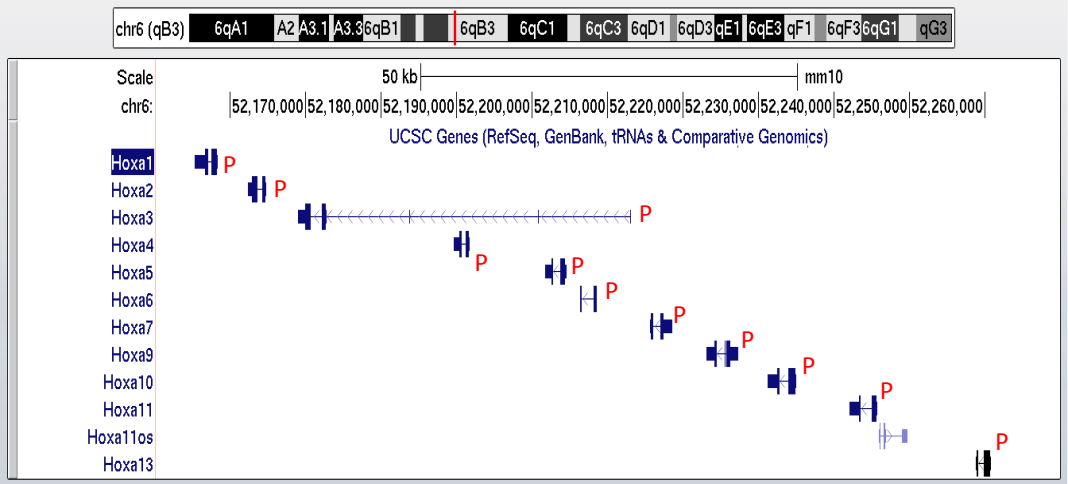


Figure 2.5 Diagram of UCSC genome browser data showing the positions of Hoxa genes within the Hoxa cluster on mouse chromosome 6. Chromosome bands and base position are shown above. Approximate primer positions are marked with a red P, and are targeted to the 5' promoter region for each gene.

value is subtracted from the bound fraction value. The negative of this value is then taken as a power of 2 to produce a relative enrichment value for that sample. A value above 1 indicates that the Ct value for the bound fraction is lower (i.e. earlier in the PCR, denoting higher initial copy number) than the unbound fraction, or that the bound fraction is enriched. In formula terms:

$$\Delta\Delta Ct = 2^{-(\text{Bound Ct} - \text{Unbound Ct})}$$

This allows the comparison of the amplification of DNA from bound to unbound fractions. Three biological replicates were performed, each with three technical replicates (triplicates).

2.5 RNA EXPRESSION ANALYSIS

2.5.1 Extraction of RNA from cultured cells

CCE/R cells were cultured in Sterilin 7cm dishes and harvested every day for 7 days. Cells were shaken off, washed x3 in PBS and the pellets frozen at -20^{o}C. Cells were then thawed and resuspended in 600µl RLT buffer (Qiagen RNeasy Kit) and passed through a 21G needle 7 times. An equal volume of 70% v/v ethanol was added, and the suspension transferred to a centrifuge column (Qiagen). Samples were centrifuged at 13000rpm (Eppendorf MiniSpin) for 30 sec, flow through was discarded, and 700µl RW1 buffer (Qiagen) was added. Samples were centrifuged again at the same speed for 30s, flow through was discarded and 500µl RPE buffer (Qiagen) was added to the column. Samples were centrifuged again for 30s and flow through discarded. Another 500µl RPE was added and samples centrifuged for 2 minutes. Then, columns were placed in a

new collection tube and the samples centrifuged again for 1min. Columns were again placed in a new collection tube and 30µl RNase free water (Qiagen) was added. Tubes were centrifuged for 1 min, another 30µl of water was added, samples were centrifuged again for 1 min, and eluate was transferred to Eppendorf tubes. Product was run out on 1.2% agarose gel to check for the characteristic double band of RNA-associated ribosomes.

2.5.2 Synthesis of cDNA and subsequent analysis by qPCR

1µg of RNA was added to a PCR tube with 1µl oligo-dT (Invitrogen Superscript III kit), 1µl dNTP mix and made up to 13µl with water. Samples were heated at 65^{o}C

in a PCR block (MJ research) for 5 minutes, then cooled at 0^{o}C for 1 min. 1µl DTT, 1µl RNase inhibitor, 1µl reverse transcriptase and 4µl first-strand buffer were added, and the tube was heated at 50^{o}C for 60min, then 70^{o}C for 15min. 1µl RNase H was added, and the sample was heated at 37^{o}C for 20min. Samples were diluted up to 100µl and stored at -20^{o}C for later analysis using real-time PCR described above. Primers designed for cDNA analysis span exon junctions so that any residual genomic DNA is not amplified (see **table 2.4**).

Mouse cDNA (expression) primer sets

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Tm (°C)	Product length (bp)
Skap2	GGCCCCTCTGATGGAAACCAGT	GCTGAGCGCACACCACCGTT	60	141
Evx1	GGGGAGGCCAGGAGGAGGAG	TCGGAGCCTGGCGCTTTGC	60	256
Hoxa1	AGTTGTGGTCCAAGCTATGGC	ACTCTCCAACTTTCCCTGTTTTG	58	388
Hoxa2	CGACGCTTTCACACTCGACA	TTTCCAGGGATTCTTTGTGGCC	60	273
Hoxa3	CCGCGGTCTGAAGGCTACGT	CGTGACACTCCGCCGCCAAT	60	235
Hoxa4	CCACGTGAGCGCCGTCAACTCC	GTGTGGGCGATCTCGATGCGG	60	150
Hoxa5	CCACCACGGCGGGAAAAACT	TCTGGGCCACCTATATTGTCGTG	58	240
Hoxa6	GTCCAACTCGGTCTTGGCCTGC	TACACGGCACCCGCACAGGAATT	60	270
Hoxa7	CTCGCAGAGAAGCGGCTACGG	TCCTGTCGGGTCCTGAACTGCG	60	287
Hoxa9	CGCTCTCCTTCGCGGGCTTAC	GGCAGCCGGGTTATTGGGATCGA	60	242
Hoxa10	TCCGGCACCCCTTCGTATTCA	CGCCTTTGGAACTGCCTTGACT	58	180
Hoxa11	GGTGGCTCCGGTGGCCAACG	ATGCGGGACAGTTGCAGACGC	60	131
Hoxa13	TCGGGGGTCCTGGCGAGTCG	GGGAGACGACGTCGGGCAGAGT	60	162
Actb	CCACACCCGCCACCAGTTCG	TACAGCCCGGGGAGCATCGT	60	112
Nanog	TCCTGGTCCCCACAGTTTGCC	AGGCAGGTCTTCAGAGGAAGGGC	60	169
Pou5f1	TGGGCTAGAGAAGGATGTGG	TGGGAAAGGTGTCCCTGTAG	60	118

Table 2.4 Sequences, lengths and melting temperatures of primers used during qPCR of cDNA derived from expression analysis experiments

3. RESULTS – CHROMOSOME IMMUNOFLUORESCENCE

3.1 INTRODUCTION TO IMMUNOFLUORESCENCE RESULTS

3.1.1 Inheritance of histone modifications

One of the central assumptions of the theory of epigenetics is that epigenetic modifications must be replicated faithfully upon cell division; otherwise they would be transient bystanders in the mechanism of inheritance (Turner, 2000, Masumoto *et al.*, 2011). However, the process and action of epigenetic inheritance and its link to the cell cycle have not yet been fully elucidated. The cell cycle is a crucial paradigm in the concept of inheritance as it is the basic unit of change for passing on cellular information to future generations. DNA sequences are replicated accurately through a single cell cycle, and are maintained throughout the lifetime of the organism and passed down vertically to survive for generations. It is thought that epigenetic marks, such as histone modifications, could behave in a similar way to provide a deeper layer of heritable information that can complement and interact with genetic sequences (Strahl and Allis, 2000, Turner, 2000, Guerrero-Bosagna *et al.*, 2010). I hope to probe the nature of histone modifications at the transition between cell cycles in order to better understand the nature of the inheritance of epigenetic information.

3.1.2 Histone modifications at the interphase – mitosis phase transition

The cell cycle model I am using in this study is the transition from interphase to mitosis phase in murine embryonic stem cells. Mitosis, and more specifically metaphase, is the period of the cell cycle characterised by highly condensed and

packaged DNA in chromosomes, the lack of a nuclear envelope, the formation of spindle fibres in preparation for cell splitting, and very low transcriptional activity (Konrad, 1963, Parsons and Spencer, 1997). This is useful for the study of epigenetics because it has been alleged that histone modifications are not truly epigenetic and are in fact merely indicators of gene transcription, controlled entirely by underlying DNA sequences (Singh *et al.*, 2008, Yi and Richards, 2009). Therefore, breaking the link between transcription and histone modifications by performing experiments involving cells that are not transcribing genes is particularly useful. I used mouse ES cells, which display an unusually short cell cycle profile, characterised by short gap phases (Becker *et al.*, 2007, Tarasov *et al.*, 2008). I also used these cells as they can be differentiated into any intra-embryonic cell type, providing a model for the behaviour of histone modifications through differentiation, another key insight into the wider question of epigenetic inheritance (Evans and Kaufman, 1981, Martin, 1981).

3.1.3 Chromosomes under the microscope

Metaphase is an ideal point to analyse the association of histone modifications with DNA through the cell cycle, as it is transcriptionally inert, a very short, discrete and specific phase of the cell cycle and one entirely concerned with passing on genetic information to daughter cells (Cohen and Deane, 1976, Parsons and Spencer, 1997). The chosen method to arrest cells in metaphase is by using a mitotic spindle microtubule poison, colcemid. Then, chromosome immunostaining is used to fluorescently label specific histone modifications on the surface of metaphase chromosomes (Terrenoire *et al.*, 2010, Ding *et al.*, 2012). This is analysed using a fluorescence microscope, which can distinguish regions of

higher and lower intensities of antibody staining. This method allows for a whole-genome, single-cell analysis of the histone modification landscape at metaphase in mouse cells. The advantages of a single cell analysis method are that the effect of inter-cell noise giving unclear results is eliminated and that the entire epigenome of any given modification can be viewed at once (Terrenoire *et al.*, 2010). Another advantage is that using immunomicroscopy it is possible to gauge the progression of the cell cycle with remarkable precision. It is immediately obvious, for example, whether or not cells viewed using this technique are going through the initial stages of metaphase, a very discrete and short window in the cell cycle that has important implications as I have explained above. For the purposes of this study, the microscope was appropriate to discover the links between histone modifications and chromosomal regions.

Previous studies using the immunofluorescence techniques probed the distributions of histone modifications in human cells (Terrenoire *et al.*, 2010). This method relied on the technique of reverse-DAPI banding in order to identify chromosomes in the resultant karyotypes. This method is somewhat subjective and prone to error, as it relies on the operator to distinguish G-banding patterns that are very similar to each other and can be indistinct. The technique employed in this study involves attaching a chromosome-specific FISH (Fluorescence In Situ Hybridisation) tag to DNA in order to unequivocally identify each chromosome (Evans *et al.*, 1991, Nagamachi *et al.*, 2013). This can greatly improve the precision of the karyotype identification process and can therefore improve the confidence with which the epigenetic profile of that chromosome can be called.

3.1.4 Chromosome immunofluorescence was used to indicate specific histone modification distributions

The chromosome immunofluorescence procedure involves purifying nuclei from cultured mouse cells and centrifuging them to release the chromosomes. Fluorescent-tagged antibodies indirectly bind to histone modifications H3K4me1, -me2 and -me3; H3K9me1, 2, 3 and H3K9ac; H3K27me3 and H3K27ac; H4K12ac; H4K16ac and H4K20me3 were attached to the chromatin to distinguish modified from unmodified sections of the chromosomes. These marks were chosen in order to obtain results from different classes of modifications, including acetyl and methyl marks on both histones H3 and H4, and residues with multiple possible modifications, such as H3K4, H3K9 and H3K27. Another reason that these specific modifications were chosen is that they are associated with different activities within the nucleus. H3K4me3 and H3K27me3 are constituent parts of the bivalent mark, with transcription activating and repressing properties respectively (Bernstein *et al.*, 2006, Schwartz *et al.*, 2010). The bivalent mark is present on certain genes or gene clusters that change their expression patterns upon differentiation, where the repressive component of the mark is lost in order to promote expression of the associated gene (Bernstein *et al.*, 2006, Shin *et al.*, 2012). Several other marks have reciprocal, concomitant or complementary actions, and so probing a wide range of histone modifications was necessary in order to form a more comprehensive picture of what the state of the epigenome is in eukaryotic cells.

H3K4me1, 2 and 3 are closely associated with genes, and are densely distributed within gene-rich regions of the genome (Pekowska *et al.*, 2011). However, each of

these marks aligns with a different part of the gene region. H3K4me1 was found to associate closely with enhancers (Pekowska *et al.*, 2011), H3K4me2 is enriched across the bodies of genes (Lindahl Allen *et al.*, 2009, Roudier *et al.*, 2011) and H3K4me3 commonly associates with gene promoters (Guenther *et al.*, 2007, Jiang *et al.*, 2011, Lloret-Llinares *et al.*, 2012). Marks such as H3K9me3 and H4K20me3, however, are often found in repeat-dense areas of the genome such as the centromere (Schotta *et al.*, 2004, McManus *et al.*, 2006, Puschendorf *et al.*, 2008). Acetyl marks on both histones are generally associated with relaxed chromatin and active transcription, due to their ionic charge-cancelling properties (Allfrey *et al.*, 1964, Grunstein, 1997). There is a paucity of information as to precisely how these marks are maintained or if their genomic locations are altered after differentiation, or between cells of the same population.

These studies were performed using mouse ES cell lines CCE/R and OS25, mouse embryonic fibroblasts (MEF), and OS25 cells during the first stages of differentiation. Two different cell lines were used to determine the variation of histone modification distributions within two different populations drawn from the same species. One of these cell lines (OS25) was differentiated for 7 days in order to assay the changes in epigenetic signature associated with loss of pluripotency, and concomitant lengthening of the cell cycle. MEFs were used to extend this study of developing cells into even more developed embryonic tissues, and again to provide another comparison between two different cell lines, this time for the differentiated state of embryonic cells. All cells used for these experiments were male, allowing the analysis of the almost entirely heterochromatic Y chromosome.

3.2 CHROMOSOME IMMUNOFLUORESCENCE AT METAPHASE

3.2.1 Results of the immunostaining procedure for H3K4me1 and H3K4me2

Chromosome spreads in **figure 3.1** show that H3K4me1 is distributed across undifferentiated (d0) CCE/R chromosome arms in a non-uniform pattern, as denoted by the bands of green fluorescence contrasted against the red background depicting areas low in H3K4me1. The heterochromatic centromere and Y chromosome are unlabelled and so appear red. The X chromosome also appears paler than autosomes. This pattern of K4 methylation is also found on metaphase chromosomes from a different cell population; OS25 d0 cells. This is a separate transgenic cell line, as opposed to the wild-type CCE/R line, and yet displays the banding pattern of H3K4me1 just as clearly (**figure 3.1**). The centromeres, and the X and Y chromosomes are also distinguishable in spreads for the OS25 ES cell line. Chromosome spreads labelled with H3K4me2 show a very similar picture to those for H3K4me1. Antibody-stained bands show the areas of high K4me2 concentration, contrasted against areas with less H3K4me2. Both undifferentiated cell lines show a similar banding pattern and characteristic unstained centromeres. However, the X chromosome appears indistinguishable from other chromosomes in H3K4me2 spreads. The Y chromosome, on the other hand, is clearly visible in these spreads as an unstained small chromosome in both H3K4me1 and H3K4me2 experiments. This confirms the negative correlation between mono and di-methylated H3K4 and heterochromatic regions of the karyotype (Rudolph *et al.*, 2007).

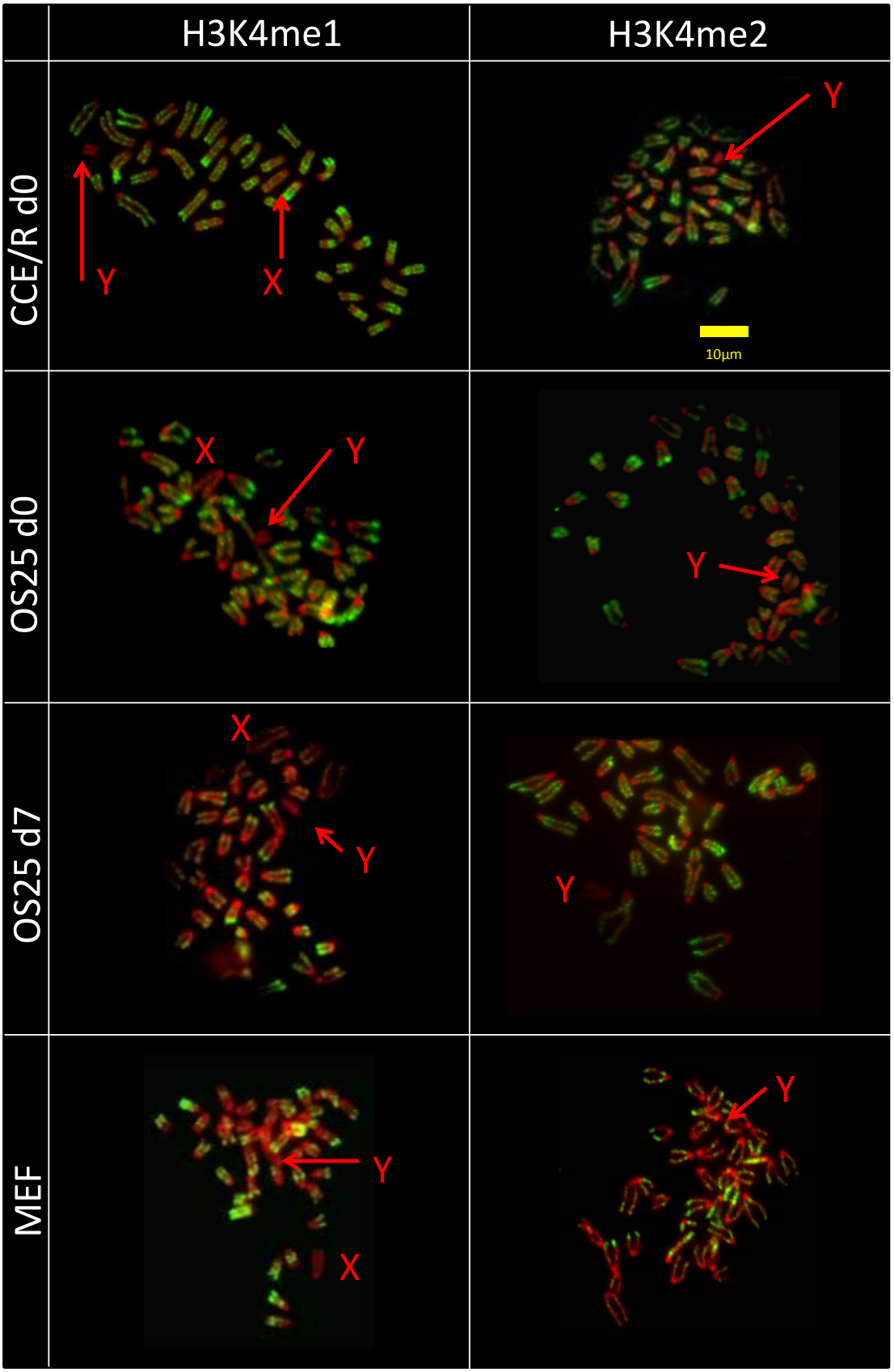

H3K4me1
H3K4me2
CCE/R d0
OS25 d0
OS25 d7
MEF
Y
X
10µm

Figure 3.1 Photographs of chromosome spreads prepared using the immunofluorescence protocol. Left panel shows chromosomes labelled with FITC (pseudo-coloured green) attached to antibodies specific to H3K4me1; right panel shows chromosomes labelled with H3K4me2. Chromosomes were counterstained with the nuclear stain DAPI (red). The cell types used were, in order from top to bottom; undifferentiated CCE/R wild type mouse ES cells, undifferentiated OS25 ES cells, OS25 cells differentiated for 5 days, and mouse embryonic fibroblast cells (MEF). The X and Y chromosomes, where clearly distinguishable, have been highlighted with red arrows. Scale bar is shown in yellow.

OS25 cells were then differentiated for 7 days and then chromosomes were analysed as before. **Figure 3.1** shows that H3K4me1 has a reduced level of enrichment compared to that seen in d0 cells. H3K4me2 shows no noticeable changes in localisation after differentiation. In mouse embryonic fibroblast cells, however, both marks show a change from d0 OS25 cells to MEFs, with reduced FITC fluorescence for MEF cells indicating a lower general level of methylation. H3K4me2-rich bands appear less frequently and are less wide than those for the other three cell populations, whereas the H3K4me1 distribution shows that some chromosomes have remained strongly stained with FITC, and others have lost H3K4me1 as they display low FITC fluorescence (**figure 3.1**). Sex chromosomes from this study of XY mouse cells did not show any noticeable change from undifferentiated to differentiated cells for H3K4-me1 and –me2, but instead displayed maintenance of these histone modifications throughout the first 7 days of differentiation (**figure 3.1**).

The two main chromatin structures, euchromatin and heterochromatin, can be distinguished by looking at -me1 and -me2 methylated H3K4 distributions. Constitutive heterochromatin, such as that found at the centromere or on a large proportion of the Y chromosome, is clearly highlighted in red in **figure 3.1**. Mono and dimethylated H3K4, as marks associated with active chromatin, do not colocalise with centric heterochromatin or the Y chromosome (Zhang *et al.*, 2009). The absence of any flecks of green fluorescence at these regions indicates they are mainly free of these modifications. Interestingly, the X chromosome also appears to be poorly stained for H3K4me1, despite the presence of over 2000 genes on the chromosome arm. This may represent epigenetic labelling of the sex chromosomes in order to distinguish them from autosomes, and possibly to

provide a mechanism of control for dosage compensation in mammals (Keohane *et al.*, 1996, O' Neill *et al.*, 2003, Rens *et al.*, 2010, Dunlap *et al.*, 2012).

3.2.2 H3K4me3 chromosomal distribution at metaphase

Although all levels of H3K4 methylation are associated with gene-rich regions, H3K4me3 is particularly closely aligned with gene transcription start sites (TSS) (Guenther *et al.*, 2007, Lloret-Llinares *et al.*, 2012) H3K4me3 shows a similar distribution pattern to H3K4me1/me2 on undifferentiated metaphase chromosomes in both cell lines, however this pattern is interrupted by the progress of differentiation, displaying a reduced level of fluorescence in d7 embryoid body cells, but then reverting to the d0 enrichment of the mark in MEF cells (**figure 3.2**). Images for d7 OS25 cells show a different distribution of H3K4me3, as the fluorescence appears less distinct and less easily recognisable as being banded. The Y chromosome is clearly unlabelled in all cell types as before, but the X chromosome appears not to be distinguishable from autosomes as in spreads for the H3K4me1 modification. Earlier studies analysed epigenetic modifications in differentiating asynchronous cells using molecular techniques such as ChIP on chip (ChIP coupled with microarray analysis) and ChIP-sequencing (Marks *et al.*, 2009, Furey, 2012). Results from these experiments show that there were few differences in H3K4 trimethylation between d0 and d10 female mouse embryonic stem cells. Autosomes from these two samples were particularly similar, and major differences were only found between the X chromosomes, as after differentiation one of these had become inactivated (Marks *et al.*, 2009). This may be an indication of the differences between a microscopic and a molecular

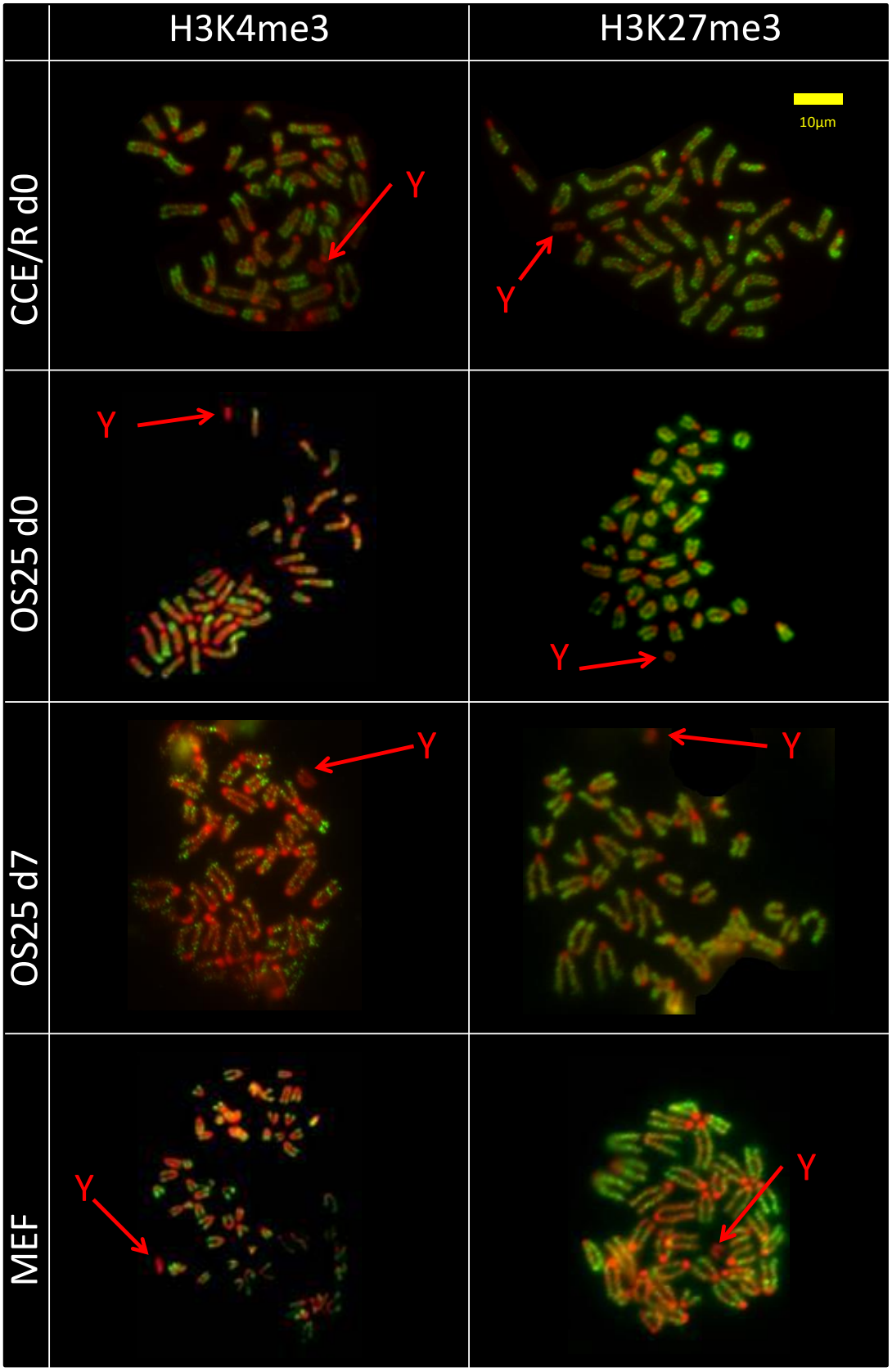

H3K4me3
H3K27me3
CCE/R d0
OS25 d0
OS25 d7
MEF
Y
10µm

Figure 3.2 Photographs of chromosome spreads prepared using the immunofluorescence protocol. Left panel shows chromosomes labelled with FITC (pseudo-coloured green) attached to antibodies specific to H3K4me3; right panel shows chromosomes labelled with H3K27me3. Chromosomes were counterstained with the nuclear stain DAPI (red). The cell types used were, in order from top to bottom; undifferentiated CCE/R wild type mouse ES cells, undifferentiated OS25 ES cells, OS25 cells differentiated for 5 days, and mouse embryonic fibroblast cells (MEF). The Y chromosome, where clearly distinguishable, has been highlighted with a red arrow.

method, as the molecular method considers each gene or stretch of DNA individually, whereas the microscopic method considers the epigenetic state of a large region of DNA at once. As such, small changes in methylation of histones may appear insignificant in the analysis of single genes, but when these alterations are combined over a range of 10 megabases, the effect can be more easily picked up by use of immunofluorescence.

It is known that all three K4 methyl marks are present in several different cell types throughout differentiation and development, although the levels of these marks can be regulated in order to produce different cell fate outcomes (Jiang *et al.*, 2011). These marks are often found in close proximity to each other, but can also have different distributions across the genome, as they are associated with separate components of genes and perform different functions. For example, in haematopoietic stem cells, it was shown that H3K4me1 occupied mainly internal regions of genes and intergenic spaces, as opposed to H3K4me3, which correlated strongly with gene promoter regions, with only a small proportion of the tri-methyl mark being present in intergenic or intra-gene regions (Cui *et al.*, 2009, Jiang *et al.*, 2011). This may explain the differences in the maintenance of these marks through differentiation, since the H3K4me3 mark may be actively maintained at many loci due to its tight association with transcription start sites, whereas H3K4me1 may be allowed to be lost at intergenic regions without affecting its function, the stability of the surrounding chromatin or resulting transcription. H3K4me2 is often distributed within coding regions of genes, as well as at the promoter (Morillon *et al.*, 2005, Zhang *et al.*, 2012). These data show that, at the chromosomal level, H3K4me2 is reduced in a similar fashion to K4me1 on differentiation to embryonic fibroblasts (**figure 3.1)**.

Despite the differences in the locations of the K4 methyl marks in genes, all three levels of methylation show banding at the undifferentiated stage (**figures 3.1 and 3.2**). This reveals that these modifications are to some degree correlated and could reflect the association of methyl K4 with gene-rich regions, therefore sections of chromatin containing a relative abundance of genes would be expected to have higher levels of methylated K4.

H3K4me3 also forms a bivalent mark with H3K27me3, which is assumed to influence the expression of genes involved in differentiation and development. This is thought to constitute a system of epigenetic control, and is a dynamic motif where one of the marks is lost at some point in order to alter gene transcription (Bernstein *et al.*, 2006, Vastenhouw and Schier, 2012). It therefore provides an ideal model for the study of epigenetic change after differentiation.

3.2.3 Distribution of H3K27me3 on mitotic chromatin

Immunofluorescence data shows that trimethylation of lysine 27 on histone H3 is more widespread across the chromosome than for H3K4me3. **Figure 3.2** shows that K27me3 is generally present across the entire chromosome arm, but still displays separate regions of higher and lower methyl K27 intensity. These bands are more subtly distinguished from the rest of the chromosome, but still reveal differences between genomic regions in the chromosome-wide distribution of H3K27me3. These differences persist up to at least the MEF stage of differentiation, as seen in **Fig 3.2**, showing that both the bivalent markers H3K4me3 and K27me3 display a variegated distribution across chromosome arms, and that they persist through metaphase and through differentiation. This observation indicates that the bivalent mark is continued at the chromosomal level

up to at least the embryonic fibroblast stage, and is evidence against a large-scale reorganisation of these two histone marks as a result of very early embryonic differentiation.

These bivalent modifications are both distributed across the arms of autosomes. They are also distributed along the X chromosome in a pattern that is not easily distinguishable from the other chromosomes. This is supported by previous immunofluorescence studies which show that the female active X, as could be compared to the male active X, is indeed marked by both H3K4me3 and H3K27me3 in undifferentiated cells and after differentiation (O' Neill *et al.*, 2008, Terrenoire *et al.*, 2010).

3.2.4 *Analysis of mouse chromosome spreads immunostained for methylated H3K9*

Methylation of histone tails on residues H3K4 and H3K27 are markedly enriched at and around genes (Barski *et al.*, 2007, Schwartz *et al.*, 2010). It is worthwhile, therefore, to analyse other residues that do not display this association, in order to probe the actions of modifications that have different localisations with respect to gene positions. Lysine 9 on histone H3 can, like K4, have up to three posttranslational methyl groups attached. These are generally linked to transcription repression, particularly di- and tri- methyl K9, and are spread throughout the genome, with the exception of H3K9me3, which shows an affinity for centric heterochromatin (McManus *et al.*, 2006, Puschendorf *et al.*, 2008). The functions of H3K9me1 are not fully known; however this mark, like H3K9me2, is distributed mainly within heterochromatin in plants (Naumann *et al.*, 2005), but has

a more euchromatic distribution in animal cells and has been implicated in controlling transcription repression (Qi *et al.*, 2010).

Figure 3.3 shows that H3K9me1 and me2 show similar distributions, localising across the chromosome arms and not at the centromere. This pattern is maintained throughout differentiation, with no obvious regions of lower or higher staining intensity. As opposed to the K4 marks, these images reveal that at the chromosomal level, mono and di-methyl K9 are distributed evenly and consistently across the chromosome arms, suggesting that they are present both at gene-rich regions and in intergenic sequences. These results show that metaphase chromosomes show strong immunostaining for both H3K9me1 and H3K9me2. This is contrary to earlier findings in human HeLa cells, which suggested that these marks are almost undetectable at metaphase (Duan *et al.*, 2008).

As is the case for the repressive mark trimethyl H3K27, the X chromosome is indistinguishable from the autosomes for mono and dimethyl H3K9. This could indicate that repressive histone modifications are distributed on autosomes and X without special prejudice, therefore contrasting with the lack of activating methyl H3K4me3 on X. The Y chromosome is also stained with H3K9me1 and H3K9me2 (**figure 3.3**), whereas the centromere is left unstained. This indicates that unlike the other marks under investigation, mono and dimethyl H3K9 distributions can discriminate between centric heterochromatin and Y chromosome heterochromatin. The distributions of these two marks remains unchanged through differentiation up to the MEF stage, as chromosome arms remain stained, whereas centromeres are still clearly distinguishable. This reflects the results from an earlier study that found that H3K9me2 levels did not change appreciably at the

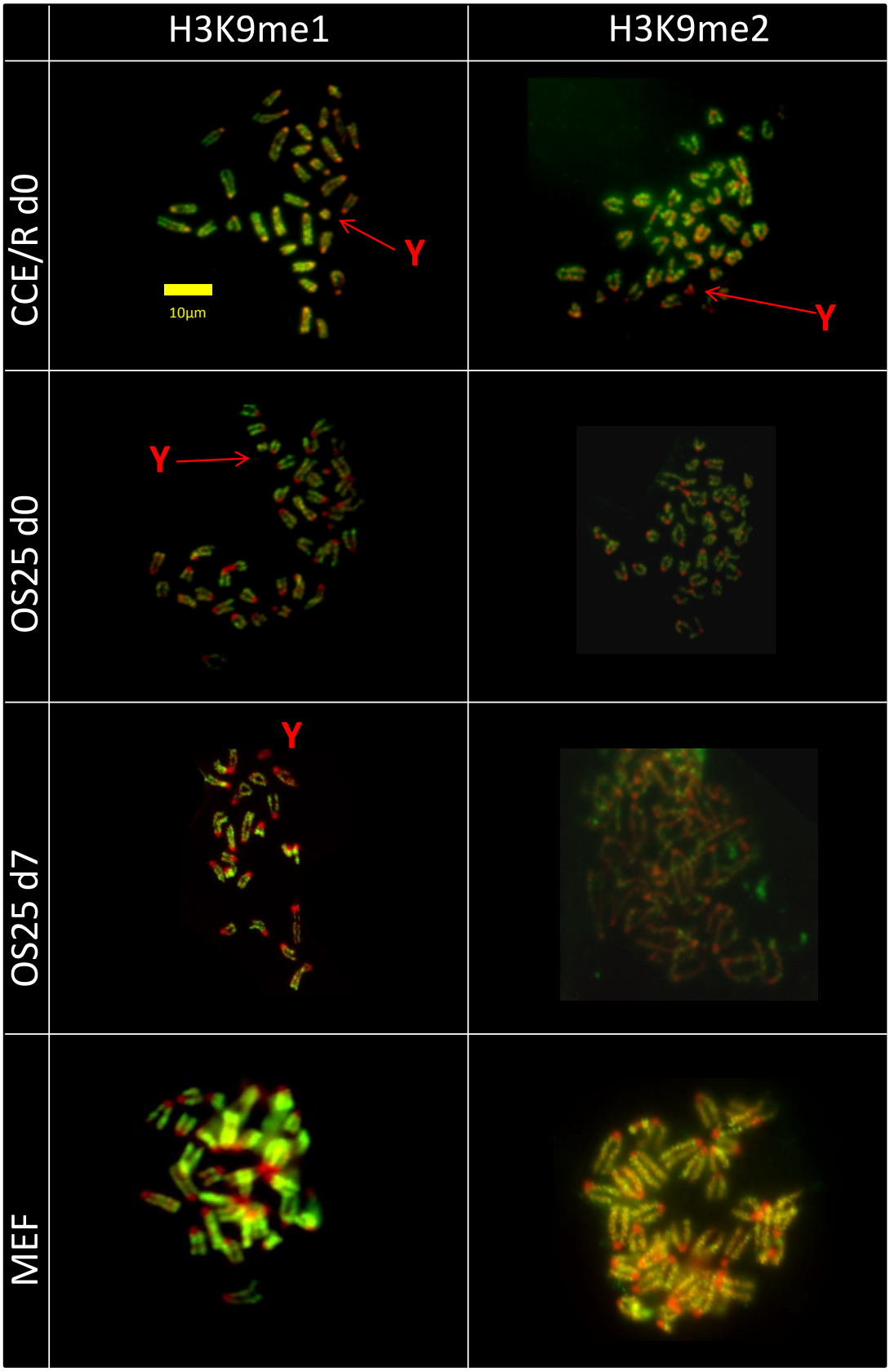

H3K9me1
H3K9me2
CCE/R d0
OS25 d0
OS25 d7
MEF
Y
Y
Y
Y
10µm

Figure 3.3 Photographs of chromosome spreads prepared using the immunofluorescence protocol. Left panel shows chromosomes labelled with FITC (pseudo-coloured green) attached to antibodies specific to H3K9me1; right panel shows chromosomes labelled with H3K9me2. Chromosomes were counterstained with the nuclear stain DAPI (red). The cell types used were, in order from top to bottom; undifferentiated CCE/R wild type mouse ES cells, undifferentiated OS25 ES cells, OS25 cells differentiated for 5 days, and mouse embryonic fibroblast cells (MEF).

global level from stem cells to differentiated cells in humans (Lienert *et al.*, 2011).

H3K9me3 has previously been very closely linked with heterochromatin, and has a repressive effect on underlying DNA sequences (McManus *et al.*, 2006, Alder *et al.*, 2010). It is known that H3K9me3 is a marker of centric heterochromatin in differentiated cells, but recently it has been discovered that K9me3 is also associated with genic regions in ES cells and may have a role in the bivalent system of epigenetic gene regulation (Alder *et al.*, 2010, Fodor *et al.*, 2010). Although **figure 3.4** shows that the intensity of the H3K9me3 signal is highest at the centromeres, there is also a general distribution across the arms of all chromosomes, confirming that H3K9me3 is widespread throughout the genome at the ES cell stage of development.

H3K9me3 localises to centric heterochromatin, but also to the Y chromosome, only one of two modifications in this sample set to do so (**figure 3.4**). In undifferentiated CCE/R cells, the entire Y chromosome is stained with the anti-K9me3 antibody, indicating that not only is this histone modification present in heterochromatic regions, but also is accessible to the antibodies used in the immunofluorescence technique. A similar pattern is evident for d0 OS25 cells (**figure 3.4**). The Y chromosome in this instance provides a good model for the link between histone modifications and underlying DNA sequence. The Y chromosome contains almost no coding genetic information and is largely a construct of condensed heterochromatin (Baumann 2008). Differentiated OS25 cells display a different H3K9me3 distribution, as the mark is almost entirely confined to centromeric chromatin at this stage of differentiation. MEF cells, on the other hand, show H3K9me3 present on both centromeres and chromosome arms in a similar pattern to undifferentiated cells (**figure 3.4**).

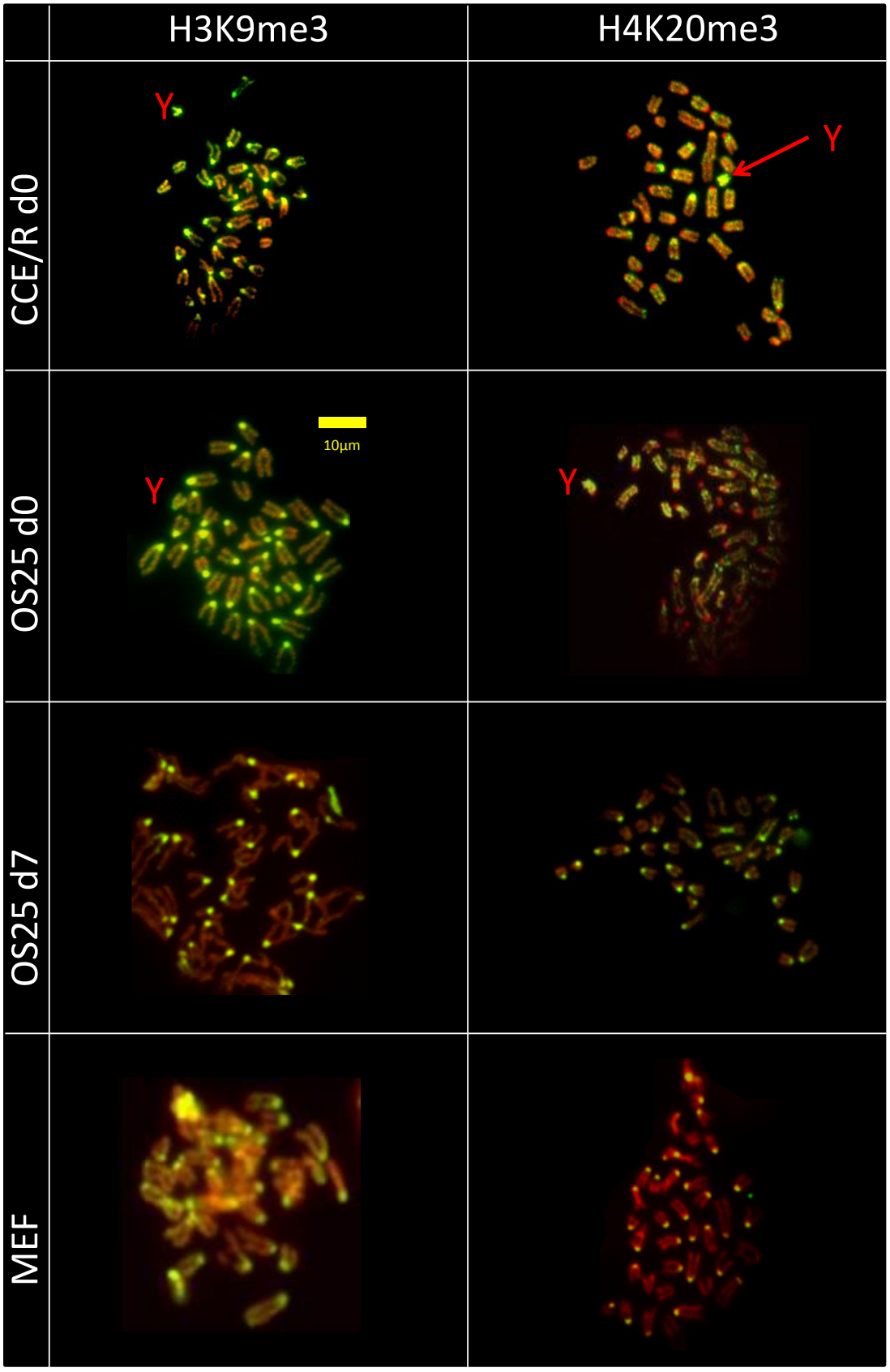

H3K9me3
H4K20me3
CCE/R d0
OS25 d0
OS25 d7
MEF
Y
Y
Y
Y
10µm

Figure 3.4 Photographs of chromosome spreads prepared using the immunofluorescence protocol. Left panel shows chromosomes labelled with FITC (pseudo-coloured green) attached to antibodies specific to H3K9me3; right panel shows chromosomes labelled with H4K20me3. Chromosomes were counterstained with the nuclear stain DAPI (red). The cell types used were, in order from top to bottom; undifferentiated CCE/R wild type mouse ES cells, undifferentiated OS25 ES cells, OS25 cells differentiated for 5 days, and mouse embryonic fibroblast cells (MEF). The Y chromosome, where clearly distinguishable, has been highlighted with a red arrow.

3.2.5 Prevalence of tri-methylation at H4K20 during metaphase

These results have highlighted the different distributions of methyl marks on modified histone H3 residues and their associations with heterochromatin. To extend the investigation further, I chose another histone modification, this time on histone H4 to analyse the similarities and differences between epigenetic modifications at metaphase on different histone molecules. Like lysine 9 of histone H3, H4K20 methylation has also been linked with heterochromatic repression of chromatin (Schotta *et al.*, 2004, Gonzalo *et al.*, 2005), and both have been implicated as markers or effectors in the maintenance of heterochromatin. However, as shown in **figure 3.4**, the difference in H4K20me3 distribution between ES cells and differentiated cells is pronounced. Differentiated MEF cells, like for H3K9me3, show only centromeric staining on metaphase chromosomes. However, images for d0 ES cell spreads show that trimethyl K20 is present along the chromosome arms, and not at the centromere. Therefore, at some point during differentiation, the localisation of K20me3 within centric DNA has been reversed.

3.2.6 Immunofluorescence of acetylated lysine residues on metaphase chromosomes

Histone methylation is known to be a complex process involved in several nuclear processes, and is thought to be involved in many and diverse biochemical pathways and not simply the regulation of gene expression. Histone acetylation is thought to have a more direct, physical role in the relaxation of chromatin and so amplifying underlying gene transcription (Allfrey *et al.*, 1964, Kouzarides, 2007).

Acetylation is also thought to be more dynamic, being added and removed from chromatin more rapidly than methyl marks (Allfrey *et al.*, 1964, Waterborg, 1998). Therefore, the persistence of acetyl marks can be assayed by the technique of chromosome immunofluorescence. This can show whether or not acetylation can be maintained in the transcriptionally (and metabolically) inert metaphase. Results from **figure 3.5** show that H3K9ac is distributed evenly across all chromosome arms for d0 CCE/Rs, but is absent from the centromere. This distribution holds true for d0 OS25 cells and through differentiation in d7 OS25s. MEF cells, however, show a subtle banding pattern for H3K9ac that was not detected during earlier developmental stages. This pattern resembles those found for methyl modifications such as H3K4me1, and may reflect a more targeted and specific pattern of H3K9 acetylation, among the background of a genome-wide deacetylation of lysine 9 after differentiation. **Figure 3.5** shows that H3K27ac does display a banding pattern across chromosome arms in d0 CCE/R cells, and to a less pronounced extent in d0 OS25s as well. This can also be seen in differentiated MEF cells, although this acetyl mark also undergoes genome-wide deacetylation at this stage as denoted by the lack of widespread FITC staining. Two acetylated residues on histone H4, H4K12ac and H4K16ac show similar distributions in all cell types assayed. Acetylated H4 is ubiquitous on chromosome arms but not centromeres in undifferentiated CCE/Rs and OS25s, and in differentiated OS25s. In a similar fashion to H3K9ac, K12ac and K16ac appear to display variegated banding in MEF cells, and a generally lower level of FITC staining indicating less widespread acetylation at these residues.

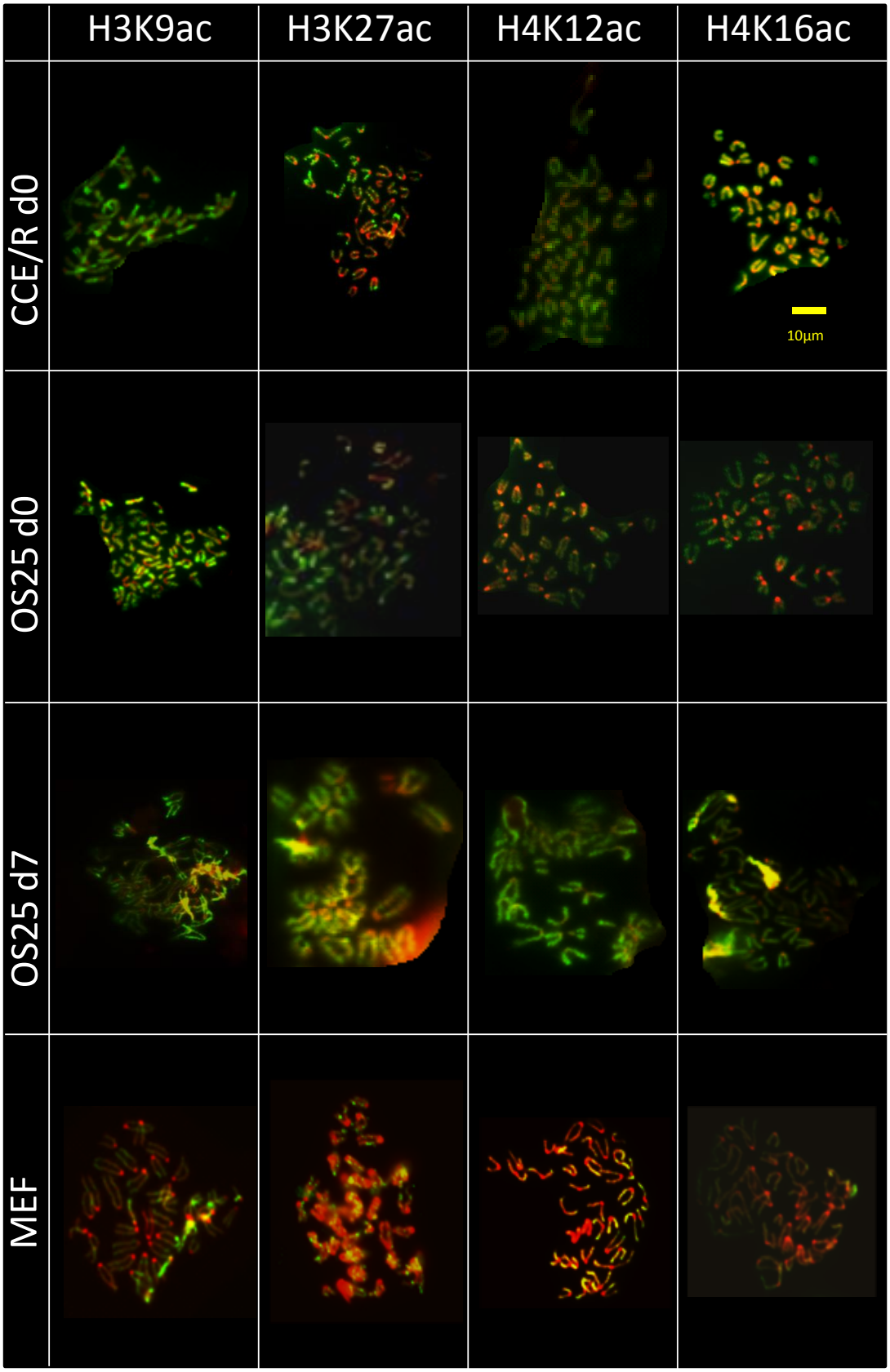

H3K9ac
H3K27ac
H4K12ac
H4K16ac
CCE/R d0
OS25 d0
OS25 d7
MEF
10µm

Figure 3.5 Photographs of chromosome spreads prepared using the immunofluorescence protocol. Far left panel shows chromosomes labelled with FITC (pseudo-coloured green) attached to antibodies specific to H3K9ac; centre-left panel shows chromosomes labelled with H3K27ac; centre-right panel shows H4K12ac immunofluorescence; and the far right panel displays chromosomes stained with anti-H4K16ac bound to FITC. Chromosomes were counterstained with the nuclear stain DAPI (red). The cell types used were, in order from top to bottom; undifferentiated CCE/R wild type mouse ES cells, undifferentiated OS25 ES cells, OS25 cells differentiated for 5 days, and mouse embryonic fibroblast cells (MEF).

3.2.7 Histone modifications present and absent at the Y chromosome

The Y chromosome in mice is specific to males and contains very little gene coding information. It is also much smaller than the other chromosomes, and the only chromosome in mice that is not telocentric. It is mainly composed of heterochromatin and therefore has very little gene activity. **Figures 3.1, 3.2 and 3.4** show that the Y chromosome can sometimes be distinguished by the histone modifications that are either present or absent on the Y chromosome. The bivalent modifications H3K4me3 and H3K27me3 are not present on the Y chromosome, as shown in **figure 3.2**, neither are H3K4me1 and H3K4me2 (**figure 3.1**). All these modifications are associated with genomic features present in gene-rich areas of the genome and are not thought to be widespread in heterochromatic DNA. **Figure 3.4** shows two more methylated chromatin residues, H3K9me3 and H4K20me3. Both are known to mark heterochromatic areas, which is reflected in results for undifferentiated cells in **figure 3.4**, showing fluorescence on the Y chromosome (Kouzarides, 2007; Puschendorf *et al.,* 2008). However, for both of these heterochromatic marks, their clear marking of the Y chromosome has been lost upon differentiation. This means that the transcriptionally inert Y chromosome is subjected to dynamic changes in histone modifications through differentiation. The epigenetic state of Y in undifferentiated cells as determined by **figures 3.1-3.5** is that it is depleted in H3K4 methylation of all kinds, and in H3K27me3. The Y chromosome is rich in H3K9me3 and H4K20me3, and is marked with acetyl modifications for H3K9, H3K27, H4K12 and H4K16 so that the Y is indistinguishable from other chromosomes. Y is visible as being depleted in some spreads for H3K9me2, but not in others, whereas the Y chromosome appears to be marked with H3K9me1 in d0 cells, but not in d7 differentiated cells, as shown in

figure 3.2. This could highlight a role for H3K9me1 in the marking of the Y chromosome specifically during the pluripotent phase of the embryo.

3.3 FISH IDENTIFICATION AND KARYOTYPING OF CHROMOSOMES

3.3.1 Building epigenetic karyotypes using FISH

Having analysed several chromatin modifications, the bivalent markers of H3K4me3 and H3K27me3 both display quite different distributions on metaphase chromosomes and could be said to each have their own unique banding pattern on each chromosome. The bivalent marker is also a key player in the switching on and off of some genes during the process of differentiation, and so applying the chromosome immunofluorescence technique to cells at various stages of development allows insight into the changes in epigenetic markers upon differentiation, which according to these results are not visible at the chromosomal scale. To investigate these observations further, I undertook a novel epigenome-wide study of the immuno-stained chromosome slides using FISH to identify each stained chromosome and so to align histone modifications with DNA sequence.

Having undertaken to obtain snapshots of the epigenome for different histone modifications, I then endeavoured to analyse the banding patterns more precisely. This can only be done by classifying and identifying each of the immunostained chromosomes to align epigenetic with genetic information. Previously, reverse-DAPI banding has been used in humans to identify each chromosome in a karyotype by selectively staining AT rich chromosome regions (Liu *et al.*, 2006). However this is not a suitable technique under these circumstances as mouse chromosomes are physically shorter than in humans, and mouse chromosomes are not morphologically distinct enough to be classified by this method as

indicated by previous work in our laboratory. Therefore I employed a molecular technique, FISH, to unmistakeably identify each chromosome based on unique DNA sequences (Nagamachi *et al.*, 2013). Whole-chromosome FISH probes were used for each chromosome individually in order to build up a full karyotype of 40 chromosomes for both H3K4me3 and H3K27me3 immunolabels. Histone modification fluorescence patterns can then be aligned with NCBI mouse genome data on gene density, repeats and CpG islands to determine the associations of these two marks with well-studied chromosomal features. By analysing the interface between genetic and epigenetic features, I will understand more about the positions of histone marks at the transcriptionally silent period of metaphase, and in turn about the nature of these marks with regard to future gene transcription and the cell cycle.

Once a chromosome spread similar to those already shown has been produced, the slides are then washed to remove any residual fluorescence and then whole-chromosome FISH DNA probes were used to hybridise specifically to a single chromosome (Cartwright *et al.*, 2013). The only dyes remaining after the FISH procedure was complete were the DNA stain DAPI and the FITC label attached to the chromosome-specific FISH probe. Key to this process is that the chromosomes remain in the same position on the slide after FISH. **Figures 3.6-3.10** show that OS25 cells have highlighted chromosome pairs due to FISH fluorescence, which can identify chromosomes at the same locations as those already imaged using immunofluorescence. Using this method, it is possible to produce an epigenome karyotype of histone modifications by immunofluorescence. This was previously difficult, as traditional methods of karyotyping rely on linking banding patterns to the physical size of chromosomes (Liu *et al.*, 2006, Terrenoire *et al.*, 2010).

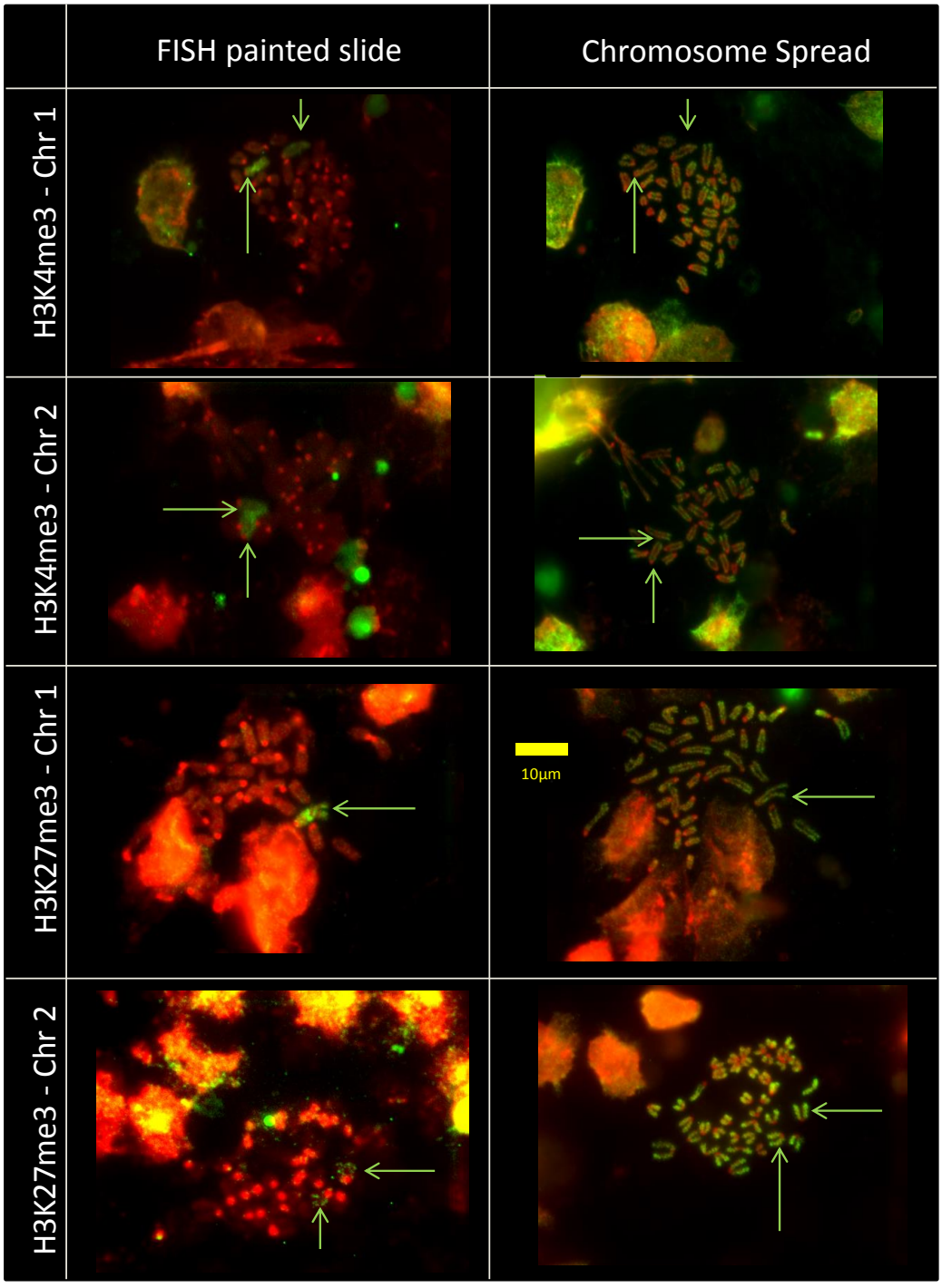
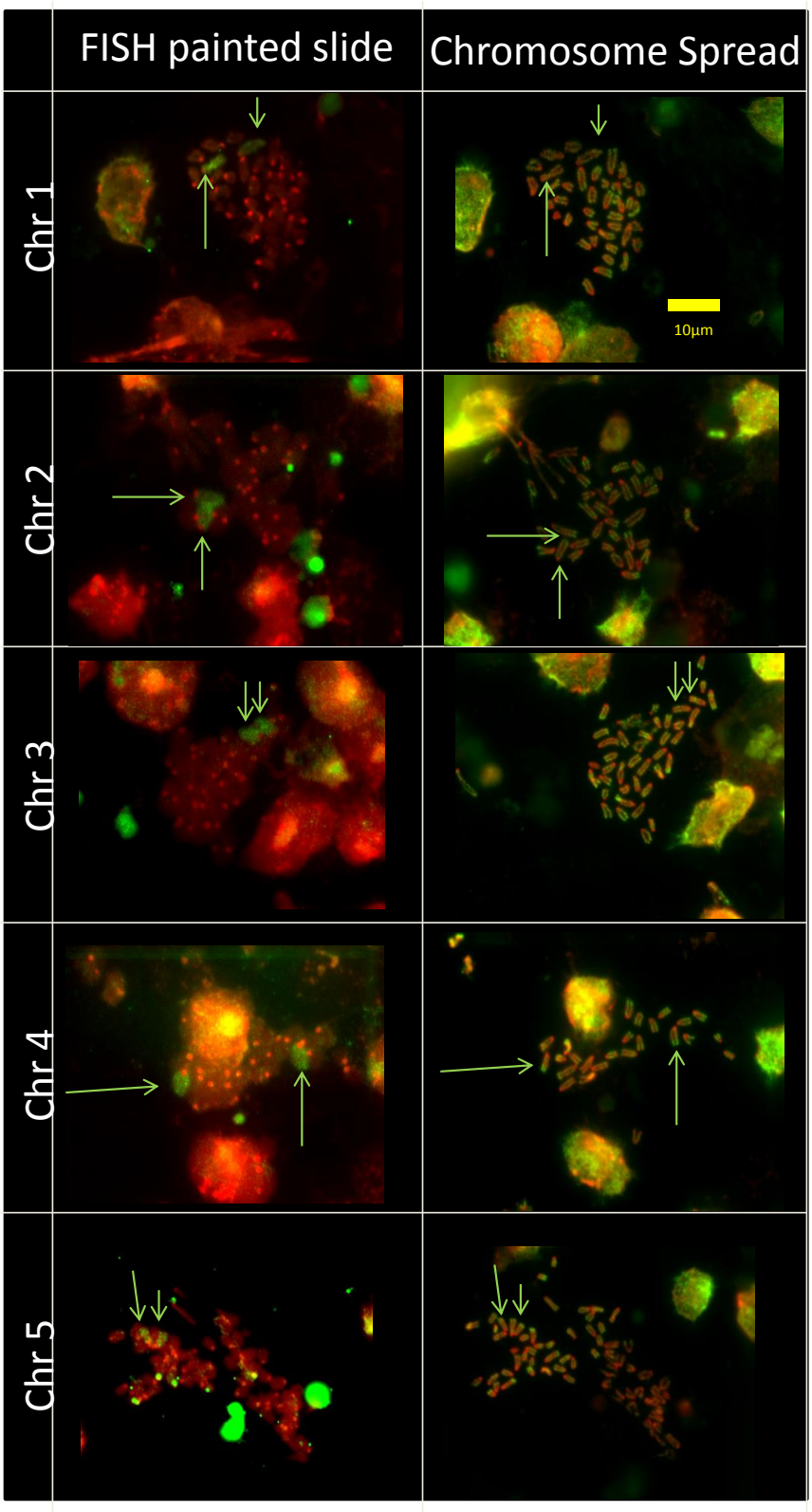
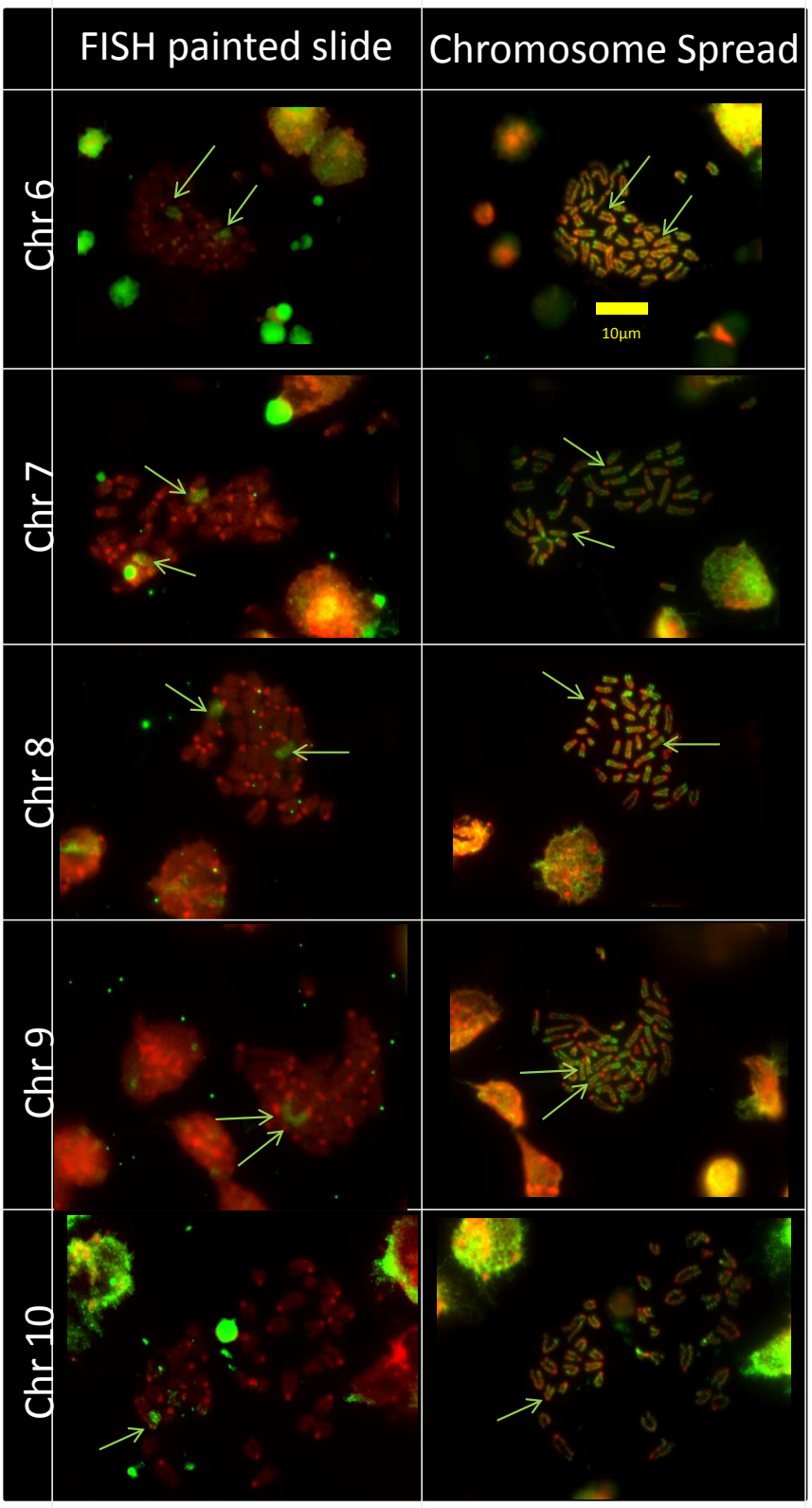
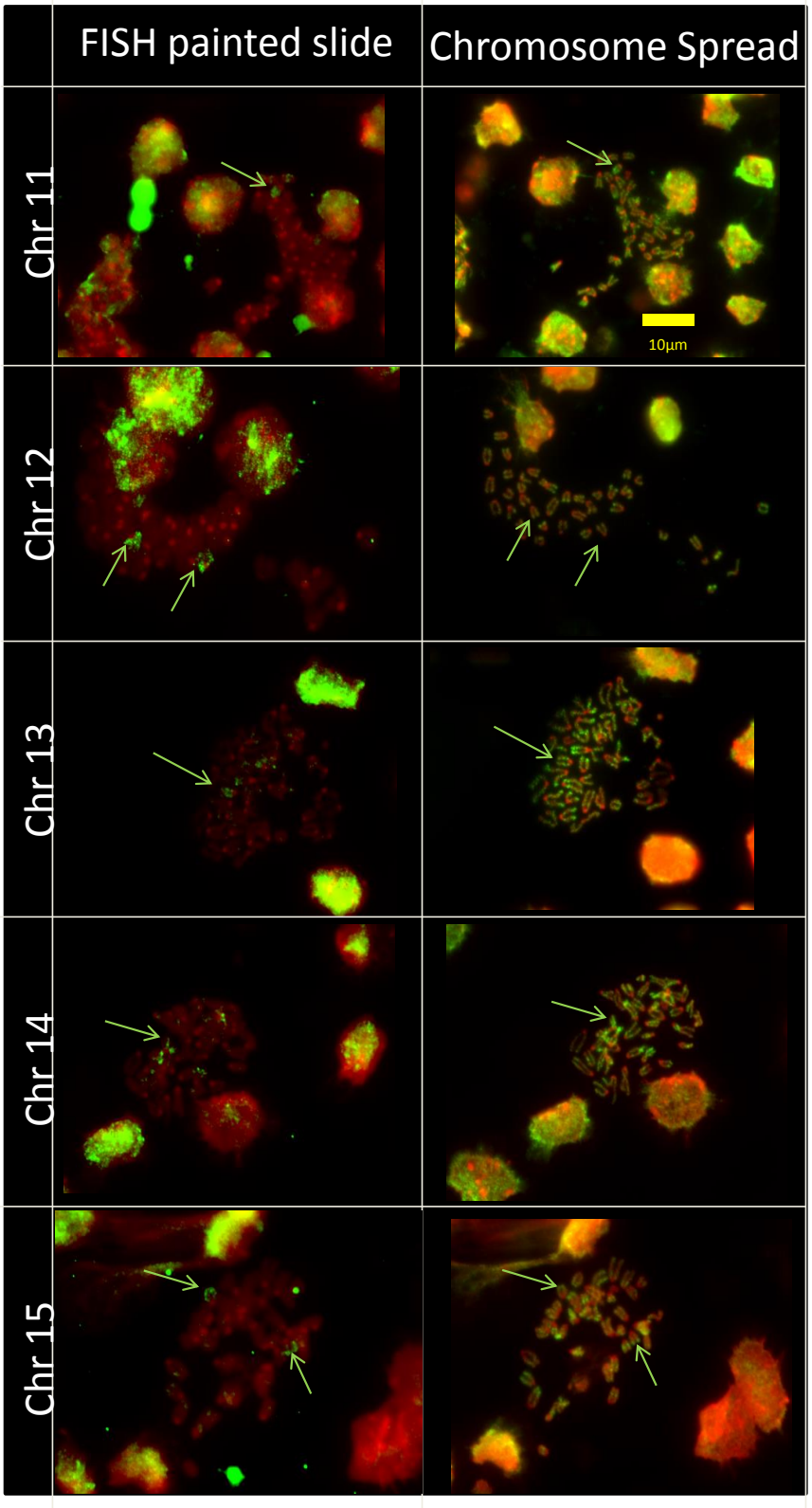
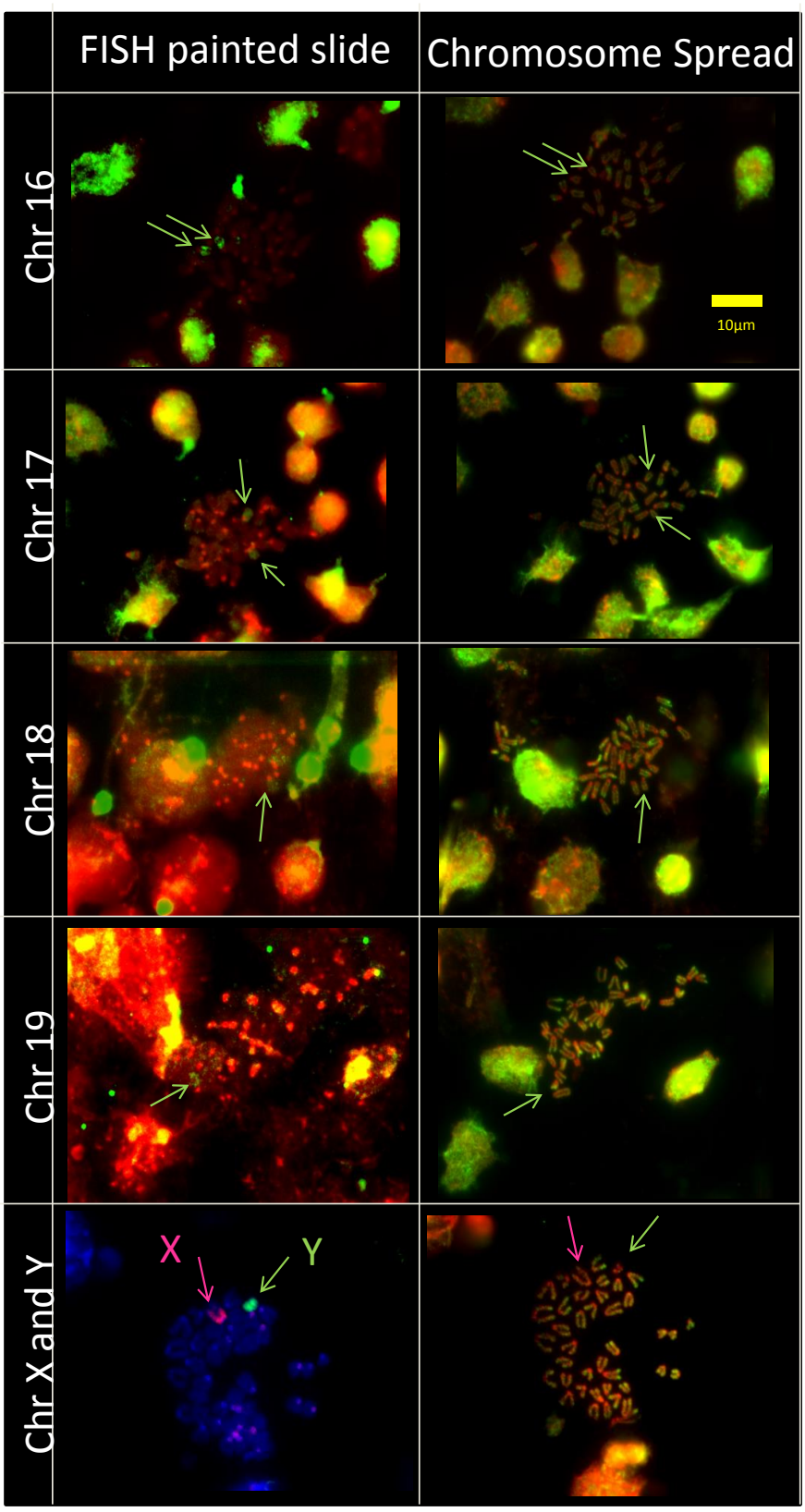


Figure 3.6 Left panel shows previously immunostained chromosome spread having been treated with FISH probes for chromosomes 1 and 2 (green) counterstained with DAPI (red). Right panel shows the original spread showing FITC staining for either H3K4me3 or H3K27me3 antibodies (green) also counterstained with DAPI. Green arrows indicate the locations of the target chromosome within the spread. Scale bar indicates 10µm.


FISH painted slide
Chromosome Spread
Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
10µm


FISH painted slide
Chromosome Spread
Chr 6
Chr 7
Chr 8
Chr 9
Chr 10
10µm


FISH painted slide
Chromosome Spread
Chr 11
Chr 12
Chr 13
Chr 14
Chr 15
10µm


FISH painted slide
Chromosome Spread
Chr 16
Chr 17
Chr 18
Chr 19
Chr X and Y
X
Y
10µm

Figures 3.7-3.10 – H3K4me3 FISH Chromosome paints. Left panel shows previously immunostained chromosome spread having been treated with FISH probes for chromosomes 1-19, X and Y (autosomes and Y chromosome in green, X chromosome shown in red) counterstained with DAPI (red). Right panel shows the original spread showing FITC staining for antibodies (green) also counterstained with DAPI. Green arrows indicate the locations of the target chromosome within the spread. Scale bar indicates 10μm.

3.3.2 FISH identification of chromosomes already assayed for H3K4me3 association

The combined results from individual FISH probes of each chromosome stained with H3K4me3 are displayed in **figure 3.11**, showing a complete mouse epigenome. The entire panel of chromosomes is shown, each one next to an ideogram showing approximate relative fluorescence across the chromosome, and an ideogram showing the G-banding pattern of each chromosome showing AT rich areas in darker colours. G-banding is a long-standing method for distinguishing human chromosomes, and is analogous to the reverse-DAPI method used in some studies (De Arce *et al.*, 1982, Terrenoire *et al.*, 2010). Almost every chromosome has a unique pattern of K4me3 staining that distinguishes it from the others, and each replicate of a chromosome shows a similar pattern to its counterparts (**figures 3.12-3.14**). Approximate patterns are shown beside the photos of chromosomes and are compared to G-banding patterns on the corresponding ideogram. The Y chromosome is also unstained, but is easily distinguished by its small size and unique morphology. The global snapshot of the H3K4me3 distribution in mouse shows a clear preference for certain chromosomal areas over others in most chromosomes. This also confirms that each chromosome has a specific and unique pattern that can identify it in terms of its H3K4me3 landscape.

3.3.3 Relationship between the H3K4me3 epigenome and genomic features

H3K4me3 banding does not closely correspond to G- banding, as seen on **figure 3.11**. These banding techniques highlight areas of DNA that are AT-rich, and do

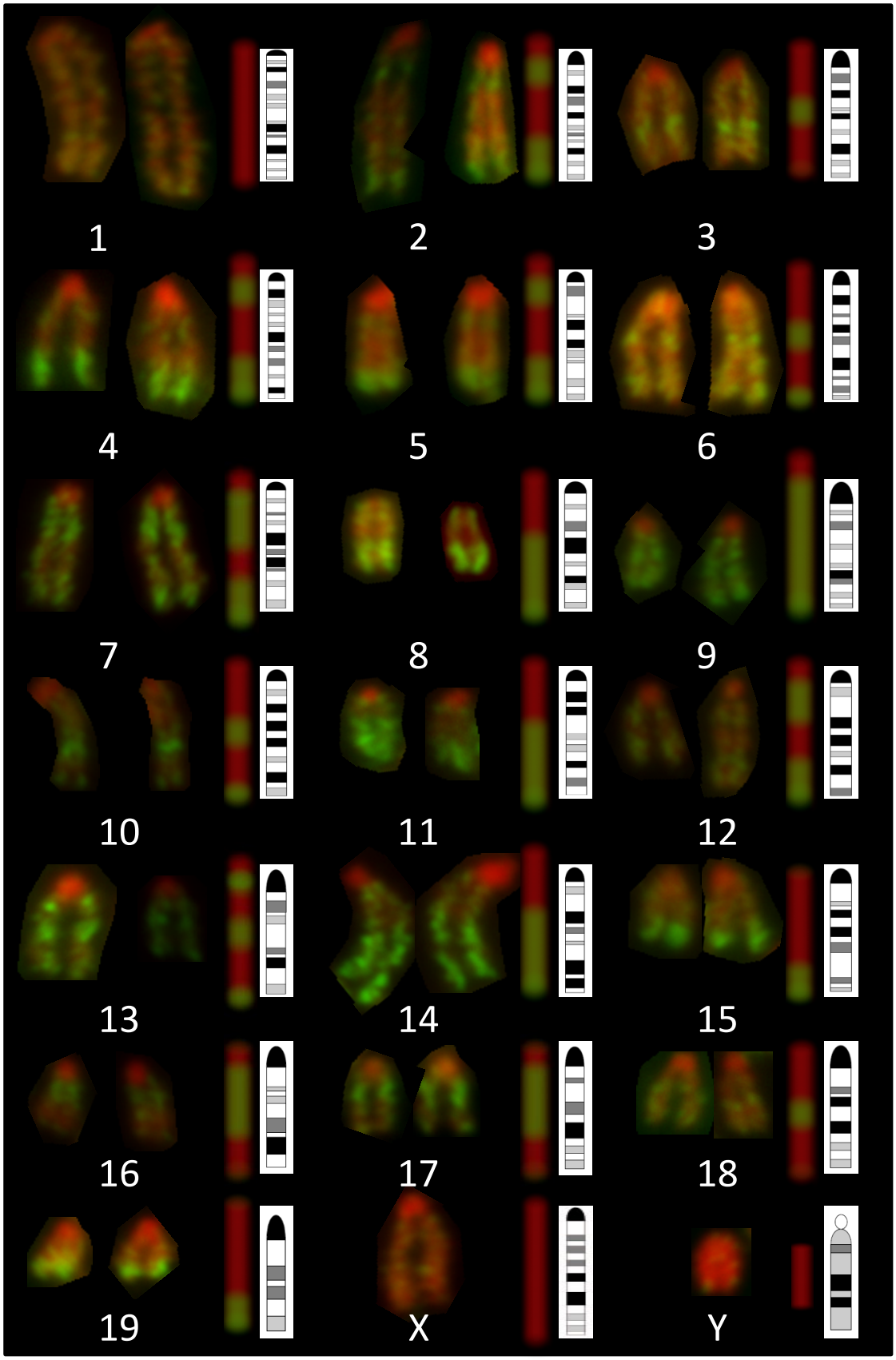

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Y

Figure 3.11 Karyotype of OS25 ES cells showing the H3K4me3 epigenome. Chromosomes were first immunostained with FITC (green) labelled antibodies to H3K4me3 and counterstained with DAPI (red), then washed and labelled with FITC-DNA probes specific for each chromosome. These were captured separately under a fluorescence microscope and aligned with their respective ideograms (university of Washington).

Chr1
Chr2
Chr3
Chr4
Chr5
Chr6

Chr13
Chr12
Chr11
Chr10
Chr9
Chr8
Chr7

Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr X
Chr Y

Figures 3.12-3.14 Replicates of chromosome staining by immunofluorescence. Chromosomes 1-19, X and Y were identified by FISH, cut out from the chromosome spread and aligned together on a single page. Four replicates for each separate chromosome indicate the reproducibility of the technique.

not always correspond to gene-poor or gene-rich areas, and so it can be said that this mark does not correlate closely with AT density on metaphase chromosomes. The exception to this is chromosome 13, where H3K4me3 is depleted at regions of heavy Giemsa staining (**figure 3.11**). To more closely analyse the significance of the bands of histone modifications on metaphase chromosomes, I analysed sequence data directly and aligned features of the genome with the epigenetic karyotype obtained from this study. Data include gene density, CpG island density and repeat density. H3K4me3 has been found to exist mainly at promoter regions, and so these histone modifications may be aligned with gene-rich regions. Similarly, H3K4me3 has been found to colocalise to sequences containing a high proportion of CpG islands, which are also associated with genes and are largely absent from intergenic loci (Bird, 1986, Cross and Bird, 1995, Illingworth *et al.*, 2010). Repeat sequences are found throughout the genome in both intergenic and genic sequences, and even centric heterochromatin and telomeres. However, some stretches of DNA contain more repeats than others, and the longer satellite repeats found in intergenic sequences and heterochromatin mean that these regions contain comparatively fewer repeats than gene-rich regions, which have shorter but more numerous repeats (Singer, 1982).

Figure 3.15-3.17 shows results obtained from the karyotype data, aligned with NCBI sequence data. The fluorescence has been converted into a numerical value that is relative to the maximum and minimum fluorescence seen in that chromosome spread. Values for FITC emittance (Y-axis) for FISH labelled chromosomes (at least 4 chromosomes per value) were obtained by analysing the relative brightnesses of pixels along each chromosome using computer software. These were normalised to the background brightness and displayed as a

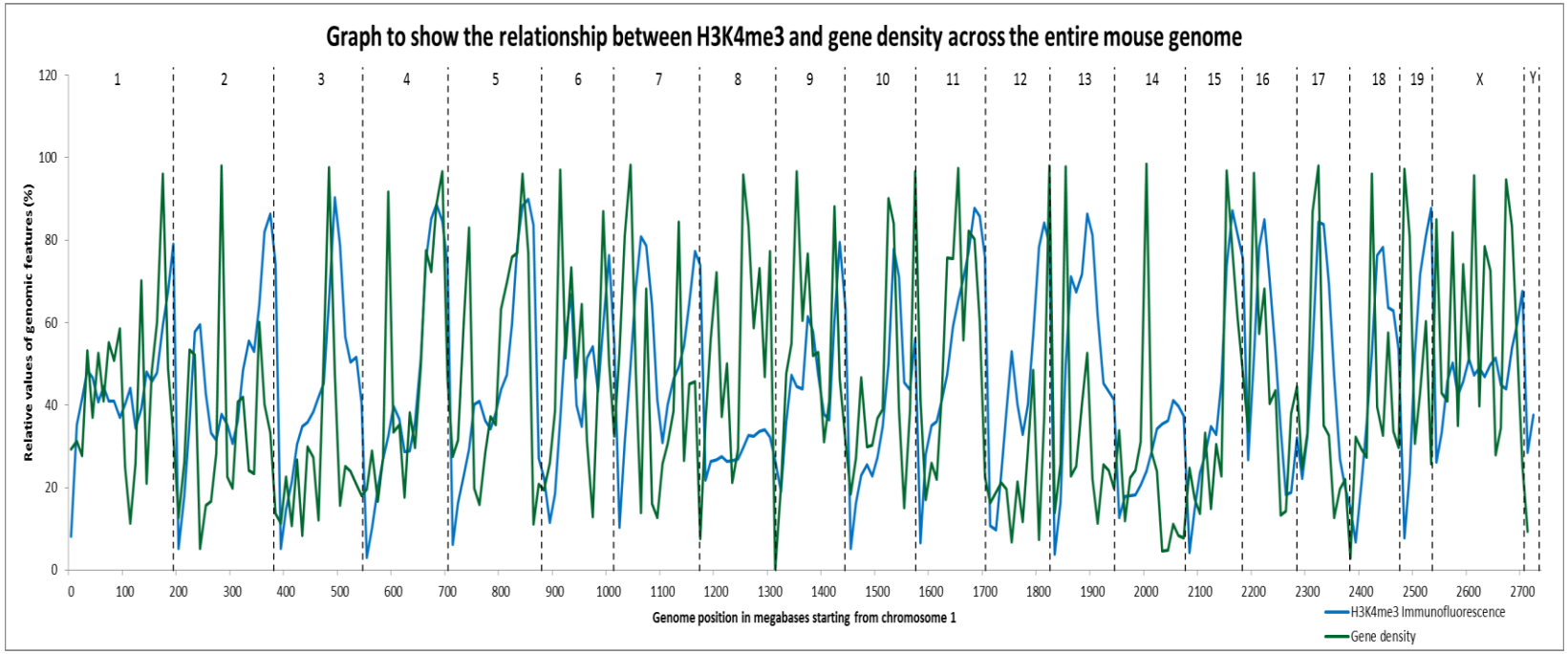


Figure 3.15 Line chart representing the density of H3K4me3 immunofluorescence across the entire mouse genome in undifferentiated OS25 ES cells (blue). Shown in green as a comparison is the NCBI data for gene density. Chromosomes have been separated by dashed lines and the appropriate chromosome label has been included at the top of the graph.

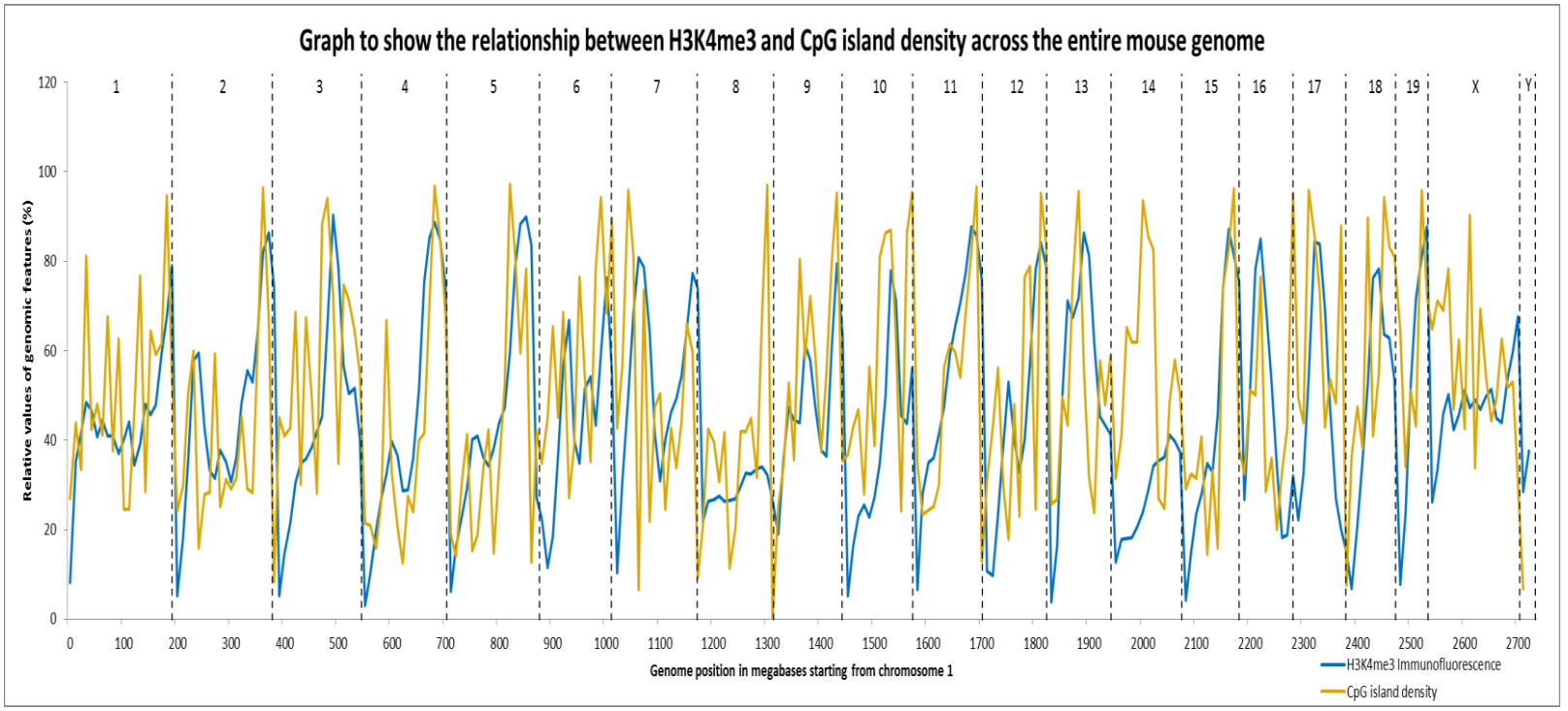


Figure 3.16 Line chart representing the density of H3K4me3 immunofluorescence across the entire mouse genome in undifferentiated OS25 ES cells (blue). Shown in yellow as a comparison is the NCBI data for CpG island density. Chromosomes have been separated by dashed lines and the appropriate chromosome label has been included at the top of the graph.

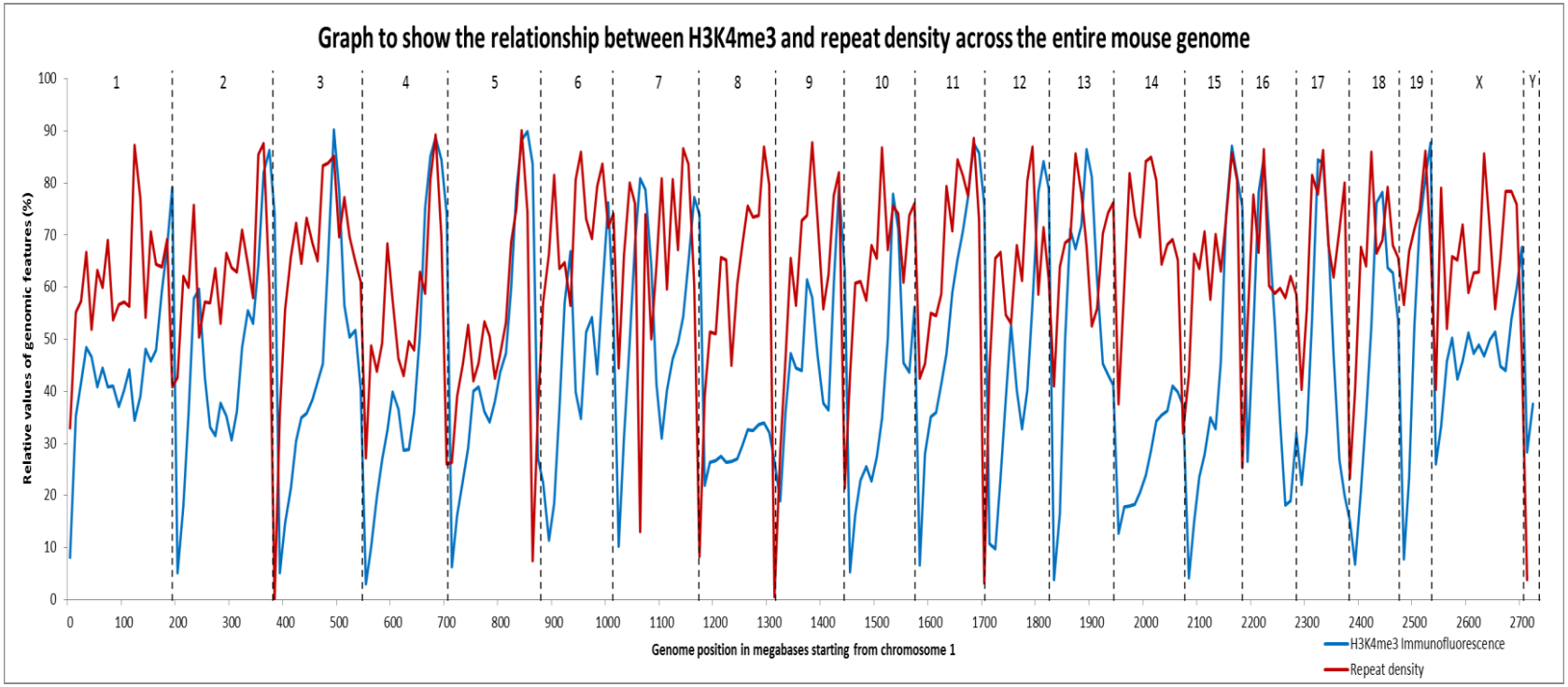


Figure 3.17 Line chart representing the density of H3K4me3 immunofluorescence across the entire mouse genome in undifferentiated OS25 ES cells (blue). Shown in red as a comparison is the NCBI data for repeat density. Chromosomes have been separated by dashed lines and the appropriate chromosome label has been included at the top of the graph.

percentage of the maximum emittance. The results were organised into 10 megabase windows in order to reflect the resolution of the microscope images. Chromosome emittance was compared to NCBI data, which was also delineated into 10 megabase windows. High percentage values indicate strong fluorescence, whereas low values indicate regions of sparse FITC staining. Analysis using this technique has produced a graph displaying the entire mouse genome in the form of a line chart. Included in the **appendix** are histograms which reflects the banding pattern of each individual chromosome in a graphical form. Error bars have been included in the **appendix** to denote the standard deviation between replicates and indicate the reproducibility of the data, but have been omitted from the line graphs so as to not obstruct the view.

The genome-wide picture of the correlation between H3K4me3 enrichment and the genomic features analysed here show a close relationship. For the majority of the genome, the lines representing H3K4me3 enrichment and gene density, CpG density and repeat density are very close to each other. Often peaks are found in the same position on the chromosome, and patterns of H3K4me3 enrichment are mirrored by the genomic features. Looking closely at the data, chromosome 5 has a fluorescent region at the telomere end of the chromosome, as shown in the photograph in **figure 3.11**, and a more subtle, smaller, region of fluorescence near the centromere. This is reflected in the FITC emittance graph below, which shows the highest fluorescence between 120 and 153 megabases from the centromeric end of the chromosome, and a small elevation in fluorescence between 40 and 60 megabases relative to surrounding chromatin. The alignment of these data with genomic features such as gene, CpG and repeat densities show a clear correlation (**figures 3.15-3.17**). Gene density on chromosome 5 is relatively higher

in a 20Mb window between 20 and 40Mb from the centromere. This may correspond to the high FITC emittance shown at 40-60Mb due to the not strictly linear nature of chromosomes on glass slides. There is also a clear elevation in gene density at the telomere end of the chromosome, from 110-150Mb, which also correlates with increased FITC emittance. CpG island and repeat density also show peaks on chromosome 5 at this 110-150Mb region of increased FITC emittance, confirming that gene density is closely related to H3K4me3 incidence on this chromosome.

Data in **figures 3.15-3.17** and in the **appendix** show that the H3K4me3 distribution throughout much of the genome corresponds closely with CpG island, gene and repeat-dense sequences. Chromosomes 2, 3, and 4 show similar profiles for FITC emittance and all three genomic features, with peaks in these coinciding in most cases. The peaks are slightly offset from each other, perhaps reflecting the fact that chromosomes on glass slides are non-linear and will be stretched or compressed in some areas over others. For all of the chromosomes in this study there is a strong correlation between peaks in FITC emittance and peaks in CpG islands, repeats and gene density, showing that the similarities in distribution patterns among these graphs reveal that H3K4me3 maintains its association with gene-rich areas of DNA even at the metaphase chromosome stage. The strongest FITC correlation among the literature sequence data sets is with CpG islands, evidenced by data for chromosome 2, which shows a K4me3 peak corresponding to a peak in CpG density, but no associated peak in gene density (**figures 3.15-3.16 and appendix**).

Some exceptions to the close relationship between H3K4me3 and these genomic features can be found on chromosomes 8 and 14 (**figures 3.15-3.17 and**

appendix). Data for these two chromosomes shows a fairly constant and moderate level of H3K4me3 enrichment across these chromosomes. However, the three genomic features under investigation do not correlate with this pattern. Both gene density and repeat density have pronounced peaks at chromosome 8 that are not reflected in the H3K4me3 data or in the NCBI data for CpG island density. This could indicate that H3K4me3 is preferentially deposited at CpG islands, which coincide in many cases with gene and repeat dense areas of the genome. However, the same is not the case for chromosome 14, which shows a distinct peak in density for all three genomic features that is not reflected in the epigenetic H3K4me3 distribution. One possible explanation for this observation is that the region in which CpG island density and gene density increases, but H3K4me3 remains low, contains several variable T-cell receptor gene loci. It would be important for these regions to contain CpG islands so that expression of these vital variable regions would not be silenced by DNA methylation, and this would also clearly be a region of high gene density. However, the embryonic cells have not been poised for T-cell expression yet and so are likely to not mark these sites for activation using H3K4me3. Taken as a whole, therefore, these data, in most cases, support the association of H3K4me3 with genes, repeats and CpG islands, which fits in with the tight association of trimethyl K4 with gene promoters and TSSs as a marker of active gene expression (Pekowska *et al.*, 2011).

The relatively pale stained chromosomes 1, X and Y all show a uniform distribution of gene density and do not have pronounced regions of higher and lower density of any of these three genomic features. This is reflected in the FITC emittance data, showing a relatively uniform distribution across the chromatid arms (**figure 3.11 and appendix**). The comparison between chromosomes 1 and

2 is particularly reflective of this, as despite having 15 megabase pairs more DNA than chromosome 2, chromosome 1 contains only 1944 genes, 729 fewer than its shorter counterpart (NCBI, 2013). These data show that trimethyl K4 levels are equivalent along the length of chromosome 1, and significantly lower than at regions on chromosome 2. This may indicate that genes are spread more evenly along chromosome 1, having no clusters of high gene density. Alternatively, chromosome 1 may simply lack the high-density H3K4me3-rich regions that account for a large proportion of genes on chromosome 2, and therefore has fewer genes as a result. The clustering nature of H3K4me3 banding could indicate that certain sections of the genome can be functionally agglomerated by recruitment of chromatin-binding proteins to areas of high methylation. These high-density areas may be activated at the same time, or be located close to each other for efficient transcription factor and Pol II recruitment. It is already known that there are gene clusters such as *Hox* clusters that contain functionally related genes, often marked with histone modifications in a similar pattern, in the case of *Hox*; this is with the bivalent markers of H3K4me3 and H3K27me3 (Soshnikova and Duboule, 2009, Kashyap *et al.*, 2011).

3.3.4 Analysis of H3K27me3 distribution on metaphase chromosomes

Tri-methyl H3K27, as already described, has a more widespread distribution across the genome than H3K4me3 (**figure 3.2**). This is also true at the gene level, whereas K4me3 is situated at the promoter and TSS of the gene, K27 methylation covers the entire gene and environs in a block of methylated chromatin (Pauler *et al.*, 2009). However, despite the repressive action of H3K27me3, it is still distributed in gene-rich euchromatin (**figure 3.18**). **Figures 3.11 and 3.18** show

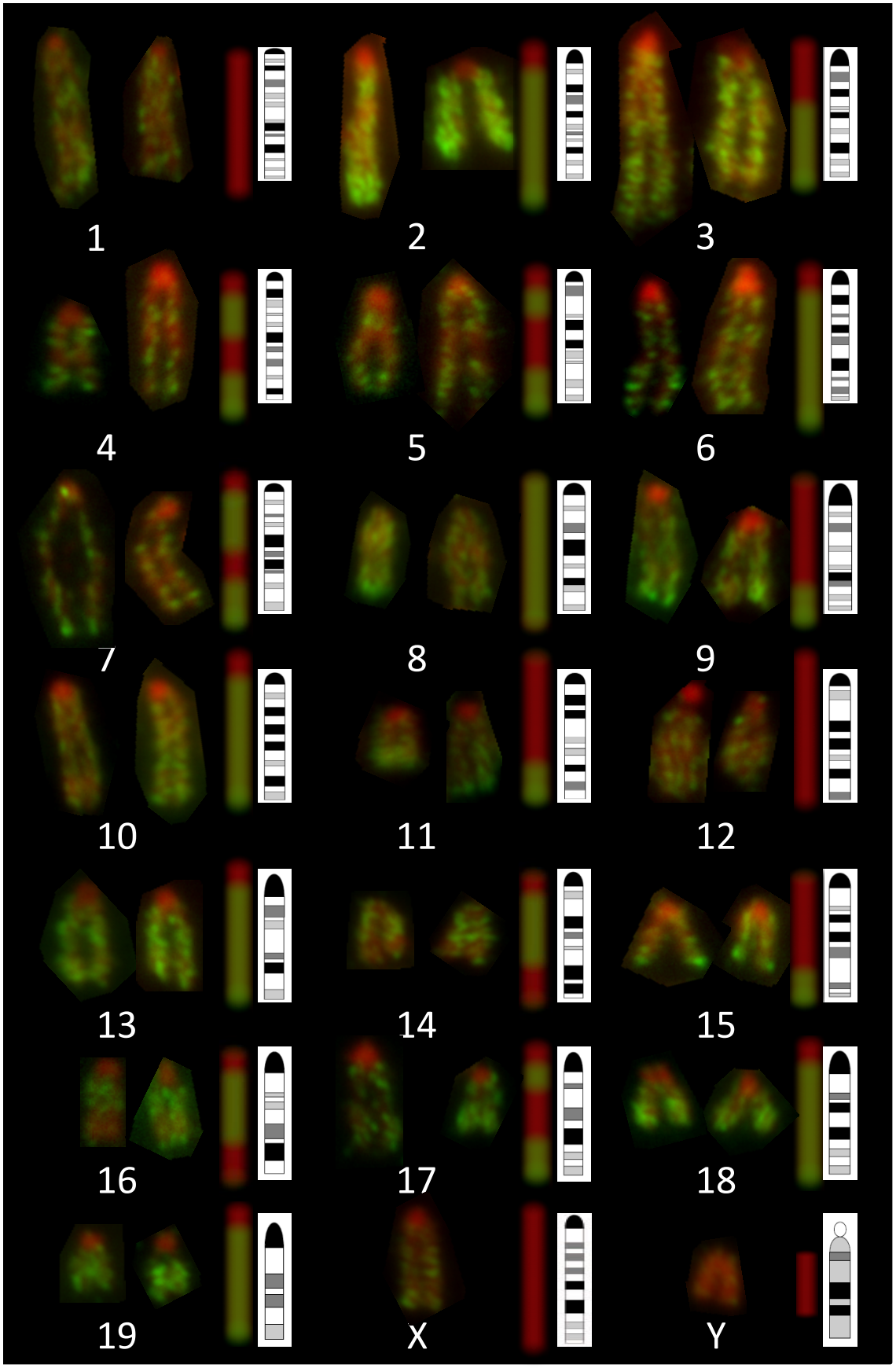
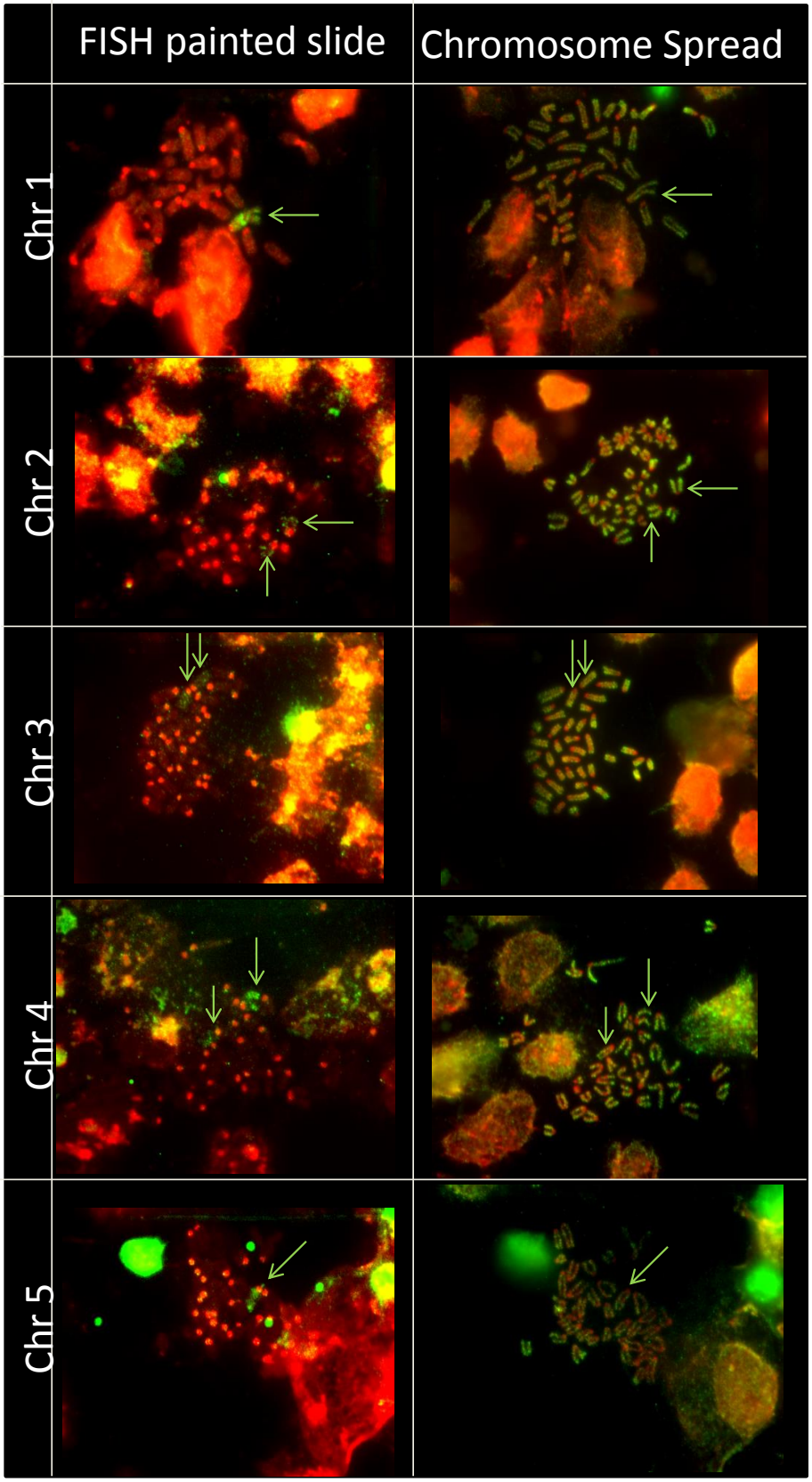
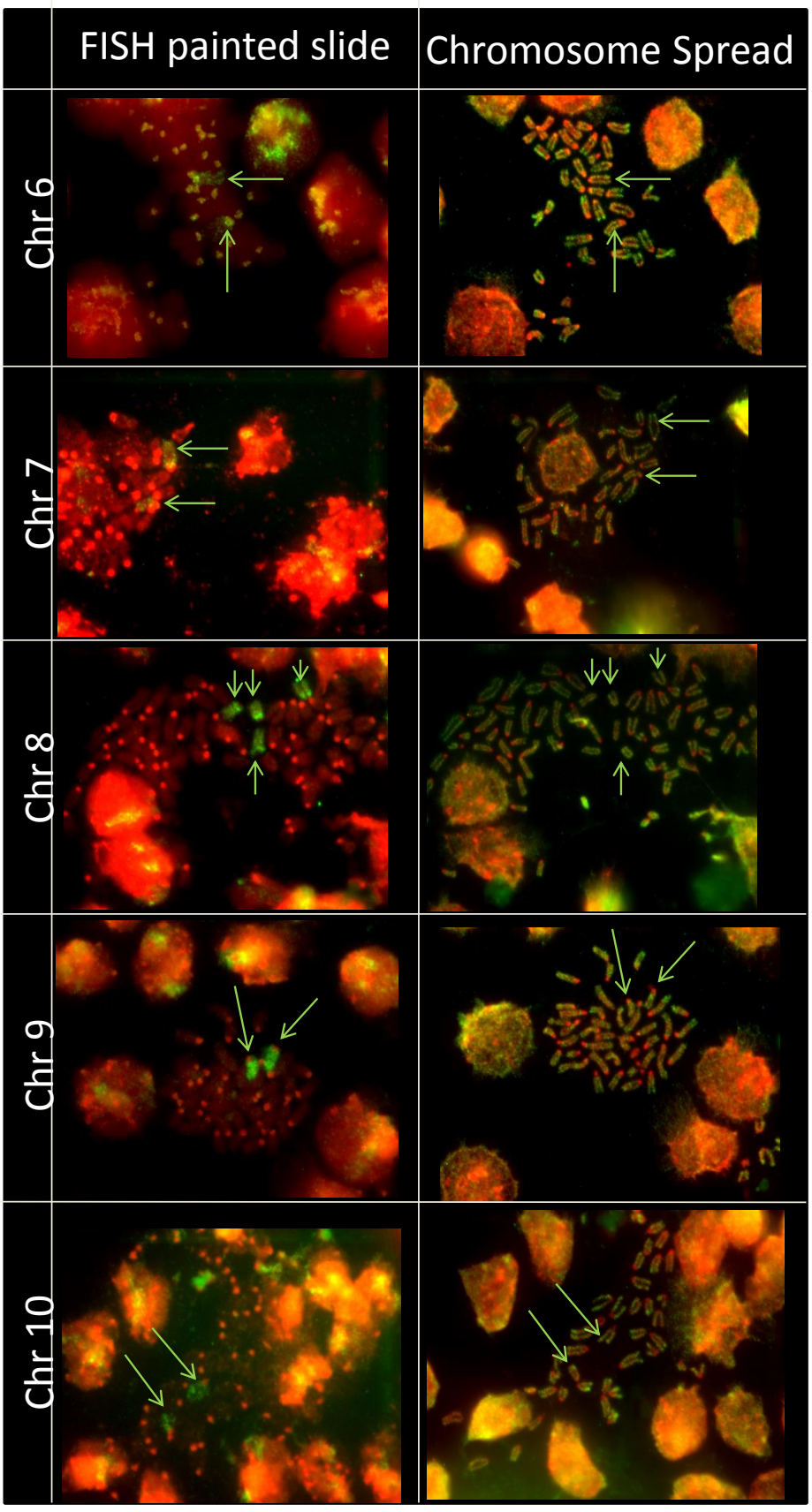

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Y

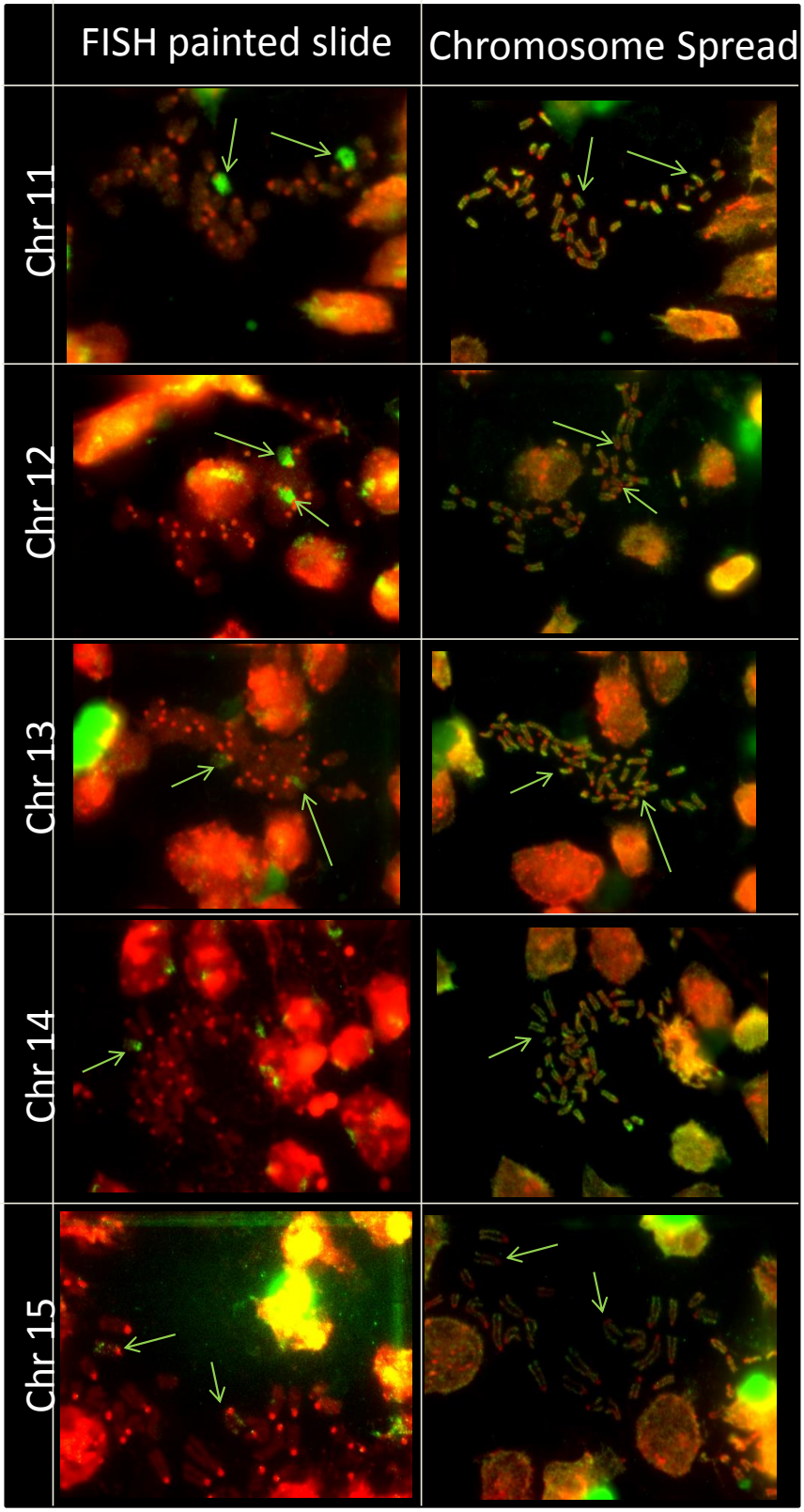
Figure 3.18 Karyotype of OS25 ES cells showing the H3K27me3 epigenome. Chromosomes were first immunostained with FITC (green) labelled antibodies to H3K27me3 and counterstained with DAPI (red), then washed and labelled with FITC-DNA probes specific for each chromosome. These were captured separately under a fluorescence microscope and aligned with their respective ideograms (university of Washington).

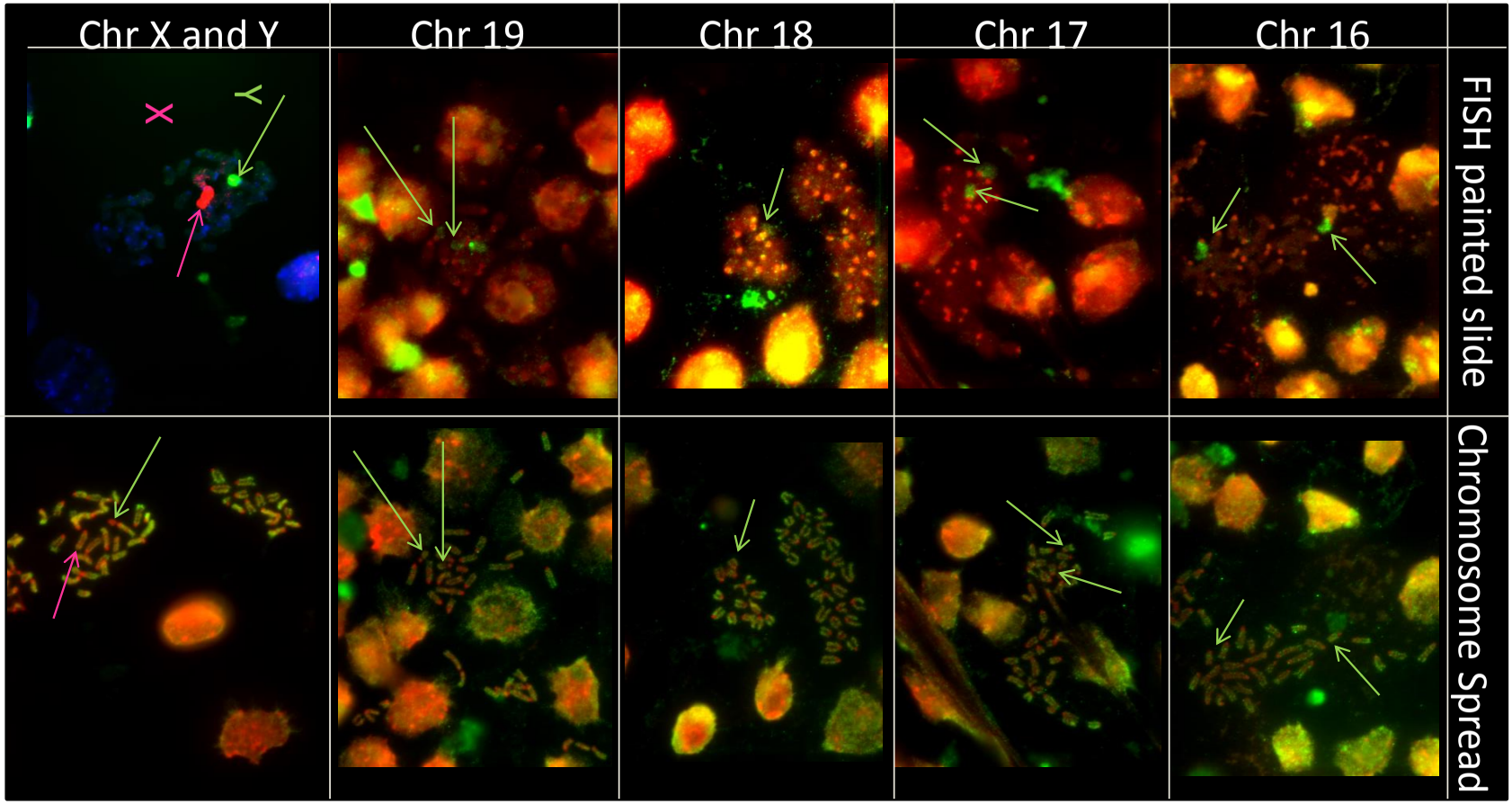
that the areas of low H3K27me3 density are less prominent than those for trimethyl K4, or alternatively that the less stained areas of the chromosome are only slightly poorer in trimethyl K27 than the more strongly stained loci; i.e. the bands are weaker. Whereas chromosomes stained with the H3K4me3 antibody displayed usually two or three bands of dense staining per chromosome, H3K27me3-stained samples show only two, one or even no clear bands on ES cell chromosomes (**figure 3.18**). There are also several chromosomes, such as chromosome 2, which have no clear banding pattern at all, and are blanketed by FITC fluorescent molecules, except at the centromere (**figure 3.18**). The FITC emitting sections of chromatin for H3K27me3 are also longer and cover more of the chromosome than the trimethyl K4 bands. However, H3K27me3 still shows preferred localisation to certain chromosomal regions over others, and the same process of FISH probe labelling can be used to identify each chromosome labelled with H3K27me3 and therefore construct an epigenetic karyotype describing the distribution of this mark across the metaphase genome.

The full karyotype of H3K27me3 labelled chromosomes is shown in **figure 3.18**. FISH slides used to construct these are shown in **figures 3.19-3.22**, with replicates of chromosomes shown in **figures 3.23-3.25**. These pictures show that the H3K27me3 epigenome contains pronounced biases towards some sections of the genome. The resulting banding pattern can be analysed as before against genomic data, and H3K4me3 epigenome data (**figure 3.7**). It is less clear where the boundaries of H3K27me3 bands lie, by the very nature of antibody fluorescence, but still a pattern can be assigned to the H3K27me3 distribution at metaphase, as boundaries between densely and sparsely fluorescent areas are still visible. By analysing the numerical fluorescent values extracted from these

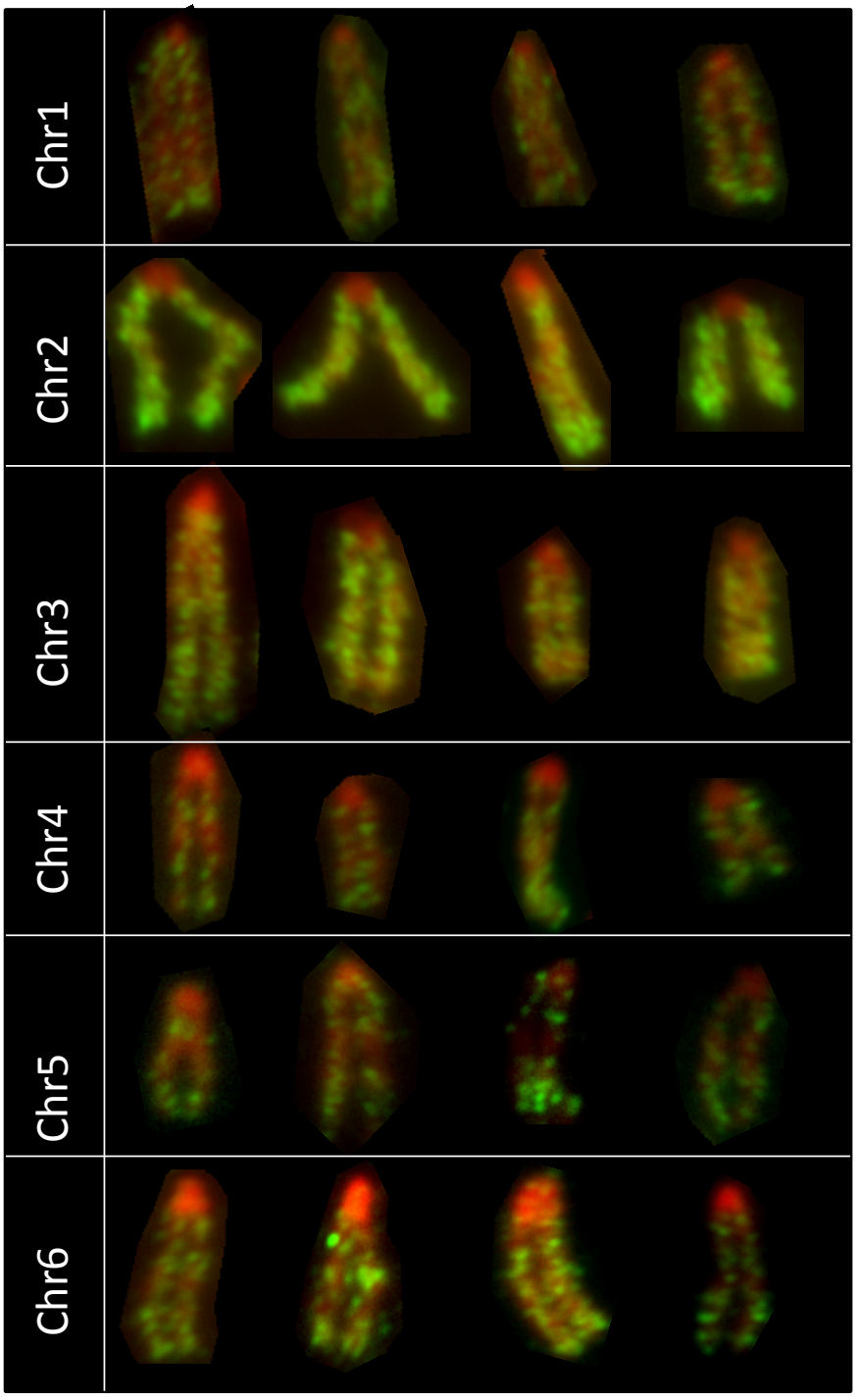

FISH painted slide
Chromosome Spread
Chr 1
Chr 2
Chr 3
Chr 4
Chr 5


FISH painted slide
Chromosome Spread
Chr 6
Chr 7
Chr 8
Chr 9
Chr 10


FISH painted slide
Chromosome Spread
Chr 11
Chr 12
Chr 13
Chr 14
Chr 15


Chr X and Y
Chr 19
Chr 18
Chr 17
Chr 16
X
Y
FISH painted slide
Chromosome Spread

Figures 3.19-3.22 – H3K27me3 FISH Chromosome paints. Left panel shows previously immunostained chromosome spread having been treated with FISH probes for chromosomes 1-19, X and Y(autosomes and Y chromosome in green, X chromosome shown in red) counterstained with DAPI (red). Right panel shows the original spread showing FITC staining for H3K27me3 antibodies (green) also counterstained with DAPI. Green arrows indicate the locations of the target chromosome within the spread. Scale bar indicates 10µm.


Chr1
Chr2
Chr3
Chr4
Chr5
Chr6

Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13

Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr X
Chr Y

Figures 3.23-3.25 Replicates of H3K27me3 chromosome staining by immunofluorescence. Chromosomes were identified by FISH, cut out from chromosome spread and aligned together on a single page. Four replicates for each separate chromosome indicate the reproducibility of the technique.

slides, it is possible to ascertain a clearer picture of the H3K27me3 distribution in ES cells than is possible by eye.

3.3.5 Banding patterns for H3K27me3 on identified chromosomes

The H3K27me3 banding pattern shows some similarities with the pattern for H3K4me3. One of the more clearly banded chromosomes in the H3K27me3 karyotype is chromosome 16, which contains a high-density region of FITC emittance near the centromere (corresponding to a gene-rich area), and a low-density region towards the telomere (**figures 3.25, 3.26 and 3.27**). This is analogous to the pattern seen for H3K4me3 on **figure 3.11**, suggesting that this is a region containing a high number of both trimethylated K4 and K27, perhaps a region with several bivalent genes or gene clusters. Clearly, however, for two marks which display very different banding patterns and distributions at metaphase, there are many instances where they do not line up together and one mark is found without the other. In keeping with its more widespread pattern of distribution, large clusters of K27me3 can be seen, for example on chromosome 2, which encompasses several of the more discrete, thin bands of trimethyl K4. This suggests that trimethyl H3K27 is deposited more broadly and persists more widely at metaphase than its tightly banded neighbour H3K4me3. This is most evident on chromosome 2, which shows K27 FITC emittance all across its arms, but K4 staining only in two bands at opposite ends of the chromosome arms (**figures 3.18 and 3.23**).

Other chromosomes appear to show similar banding patterns, such as chromosome 5 (**figure 3.23, 3.28 and 3.29**), which shows two distinct regions of high H3K27me3 banding intensity, separated by a region of weak staining.

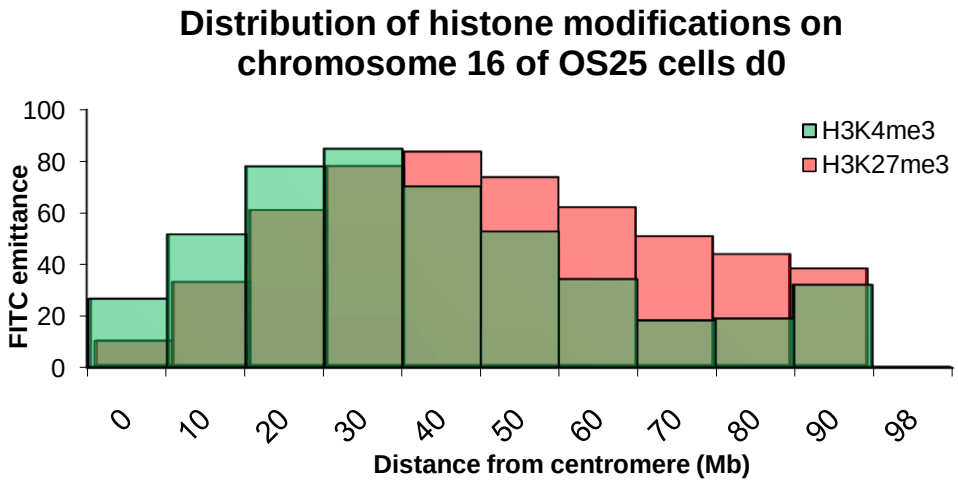


Figure 3.26 Histogram showing the relative fluorescence (%) of secondary antibodies against H3K4me3 (green) and H3K27me3 (red) histone modifications across mouse chromosome 16.

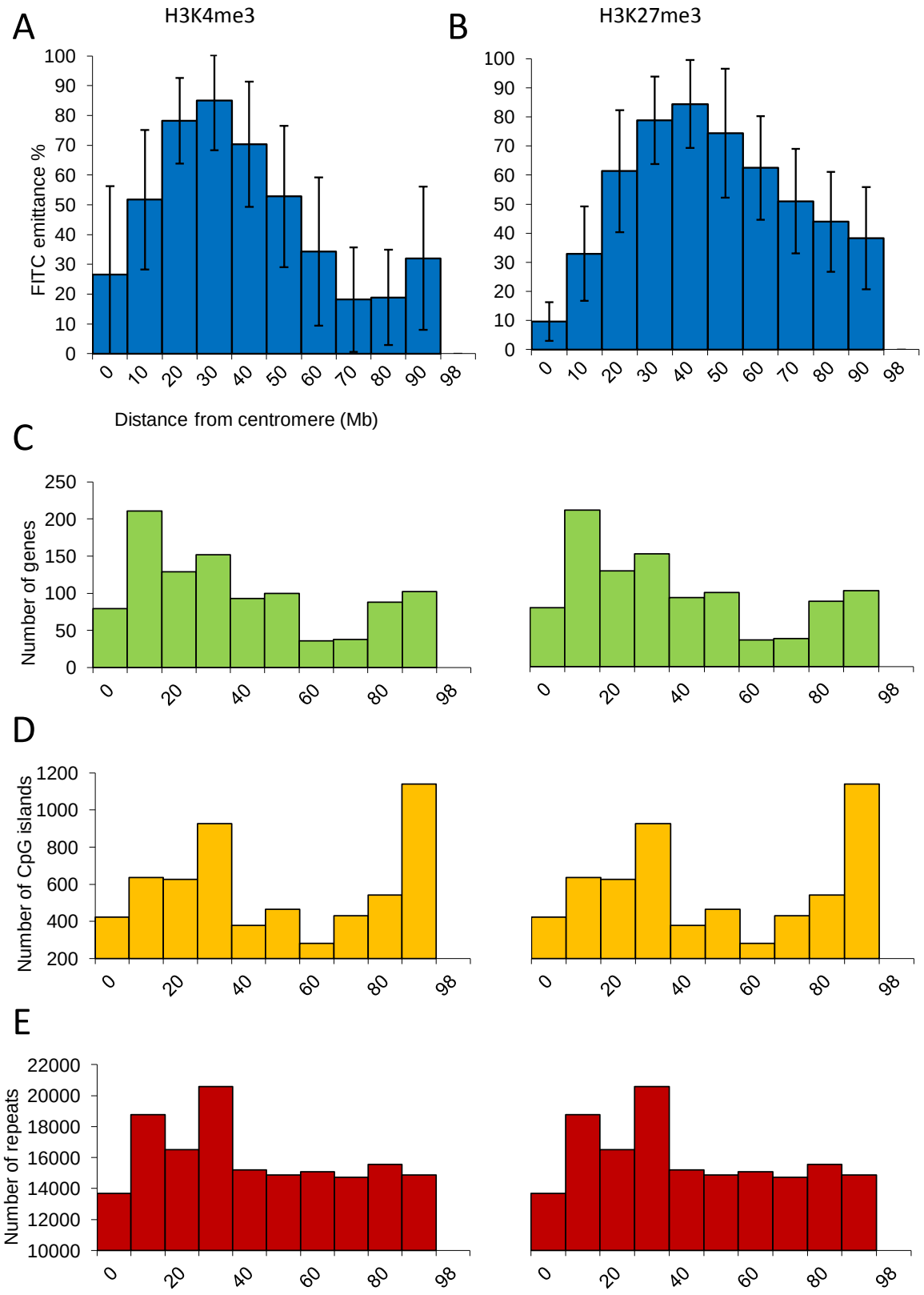


Figure 3.27 Alignment of **A.** H3K4me3 and **B.** H3K27me3 FITC emittance data from chromosome 16 with genomic features. Shown here are the NCBI data for **C.** genes, **D.** CpG islands and **E.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates

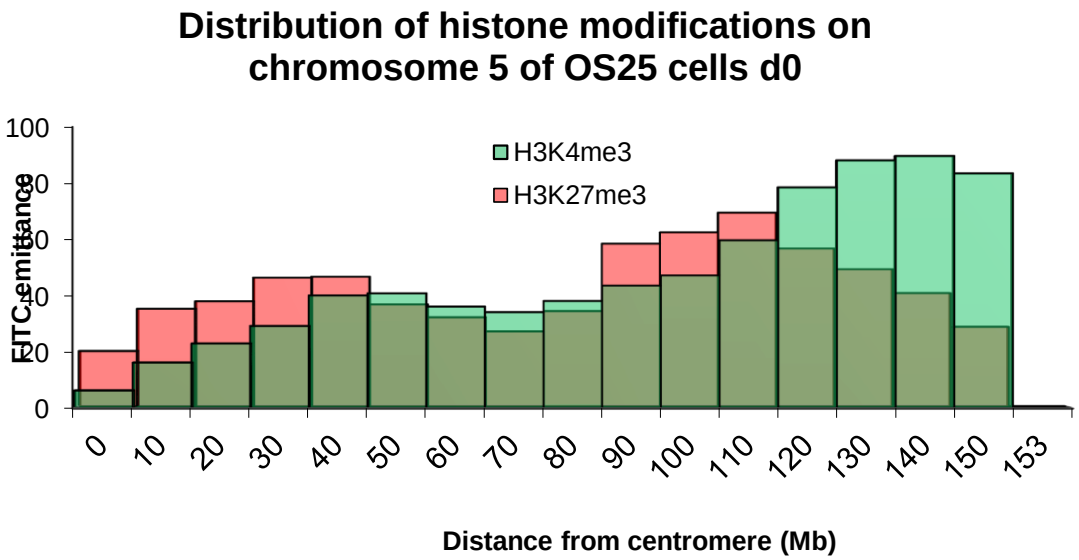


Figure 3.28 Histogram showing the relative fluorescence (%) of secondary antibodies against H3K4me3 (green) and H3K27me3 (red) histone modifications across mouse chromosome 5.

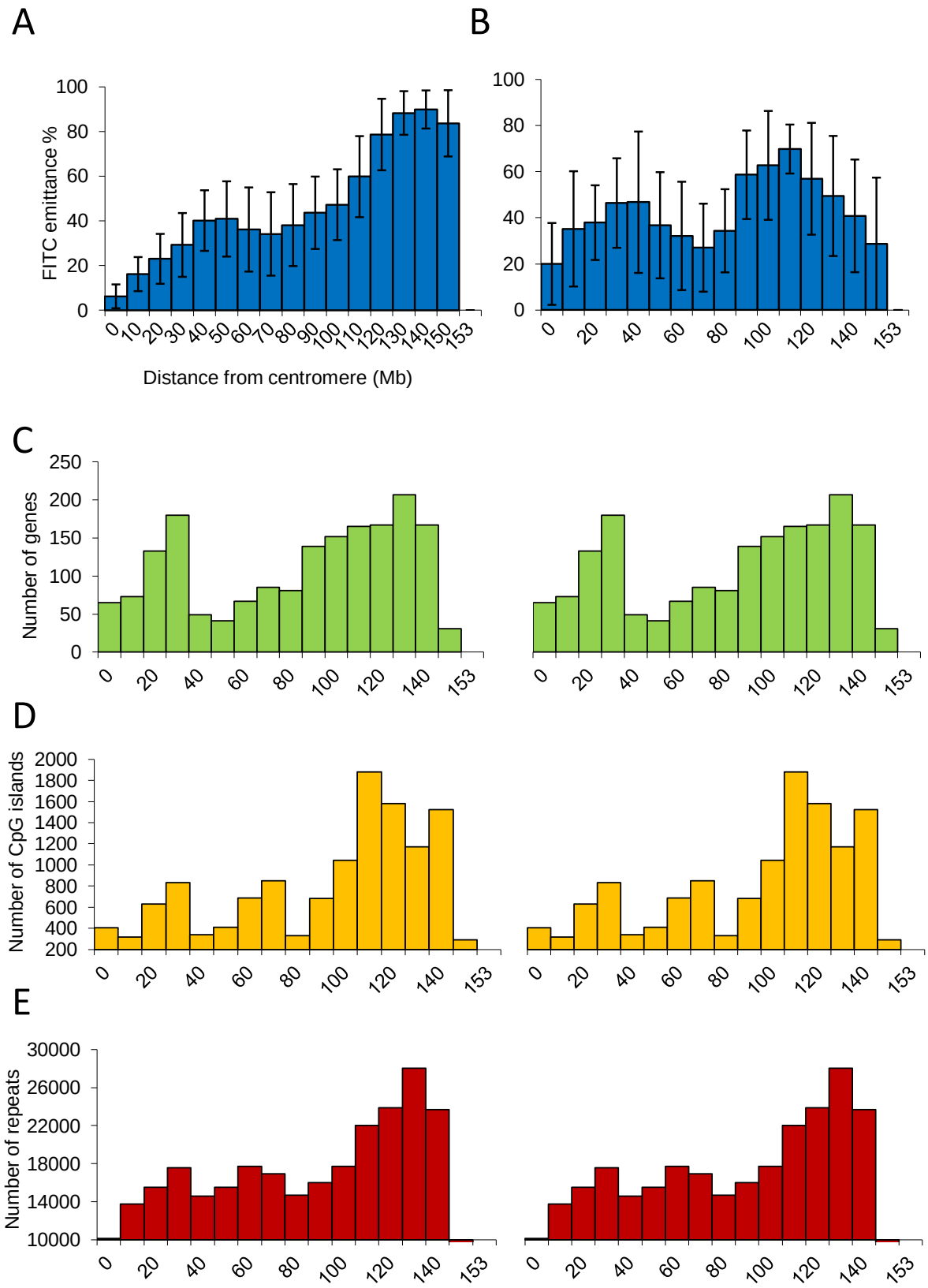


Figure 3.29 Alignment of **A.** H3K4me3 and **B.** H3K27me3 FITC emittance data from chromosome 5 with genomic features. Shown here are the NCBI data for **C.** genes, **D.** CpG islands and **E.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates

However, after analysis of several replicates of FITC-labelled chromosomes, there is a distinct difference between the distributions of the two bivalent marks. The peak at ~40Mb from the chromosome 5 centromere is prominent in H3K4me3, H3K27me3 and in genomic features; however at the telomere end of the chromosome, a peak in trimethyl K4 and in gene density is not reflected by a peak in K27me3. On the contrary, trimethyl K27 staining is reduced markedly at the telomere end of the chromosome. Unlike results seen for many of the other chromosomes, this reveals a portion of the genome where H3K4me3 and H3K27me3 do not appear to co-localise at the same concentrations.

Similarly, histone modification distribution patterns on chromosome 6 appear identical, but upon further analysis they have subtle differences (**figures 3.18, 3.30 and 3.31**). The karyotype for H3K27me3 shows that chromosome 6 appears to be stained across the entirety of the chromosome arms; however analysis of this data using the mathematical method described earlier results in a more detailed picture being obtained (**figures 3.30 and 3.31**). In this case, a peak in H3K4me3 at 50-70Mb correlates with gene density but not with CpG island density; whereas the FITC emittance histogram for H3K27me3 shows a gradual increase from the centromere to the telomere that more closely resembles the sequence data for CpG islands. This may highlight a divergence in the associations of these marks with genomic features. Whereas peaks in gene density and CpG island density often, quite logically, overlap; there are a few occasions where a lone peak in one of these sequence features can provide insight into what sequence feature is really correlated to which histone modification, and how closely epigenetic features follow underlying sequences.

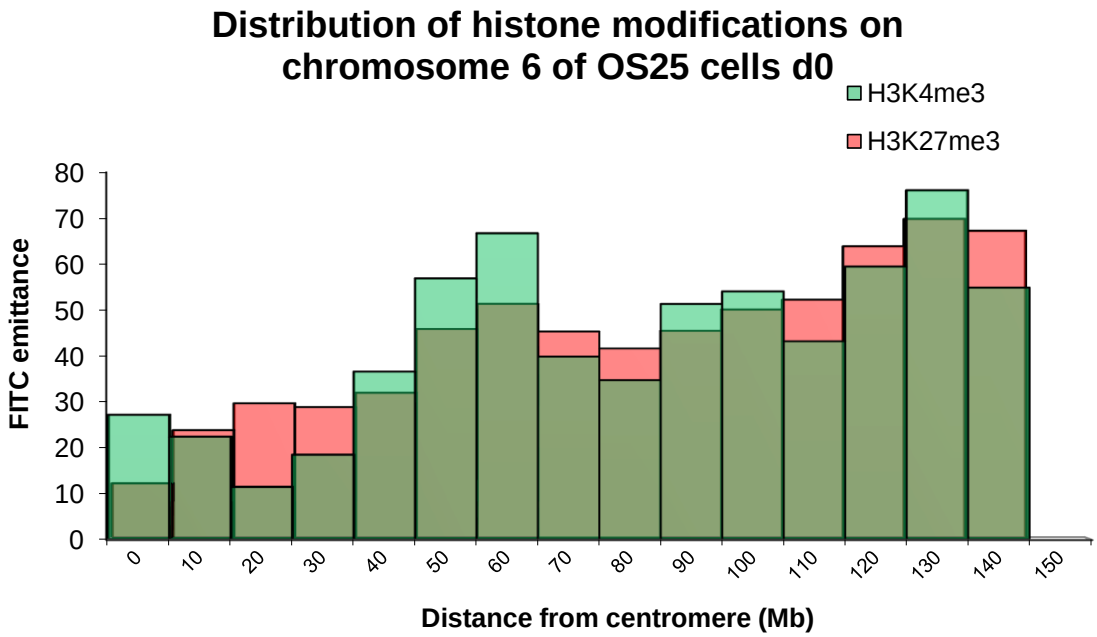


Figure 3.30 Histogram showing the relative fluorescence (%) of secondary antibodies against H3K4me3 (green) and H3K27me3 (red) histone modifications across mouse chromosome 6.

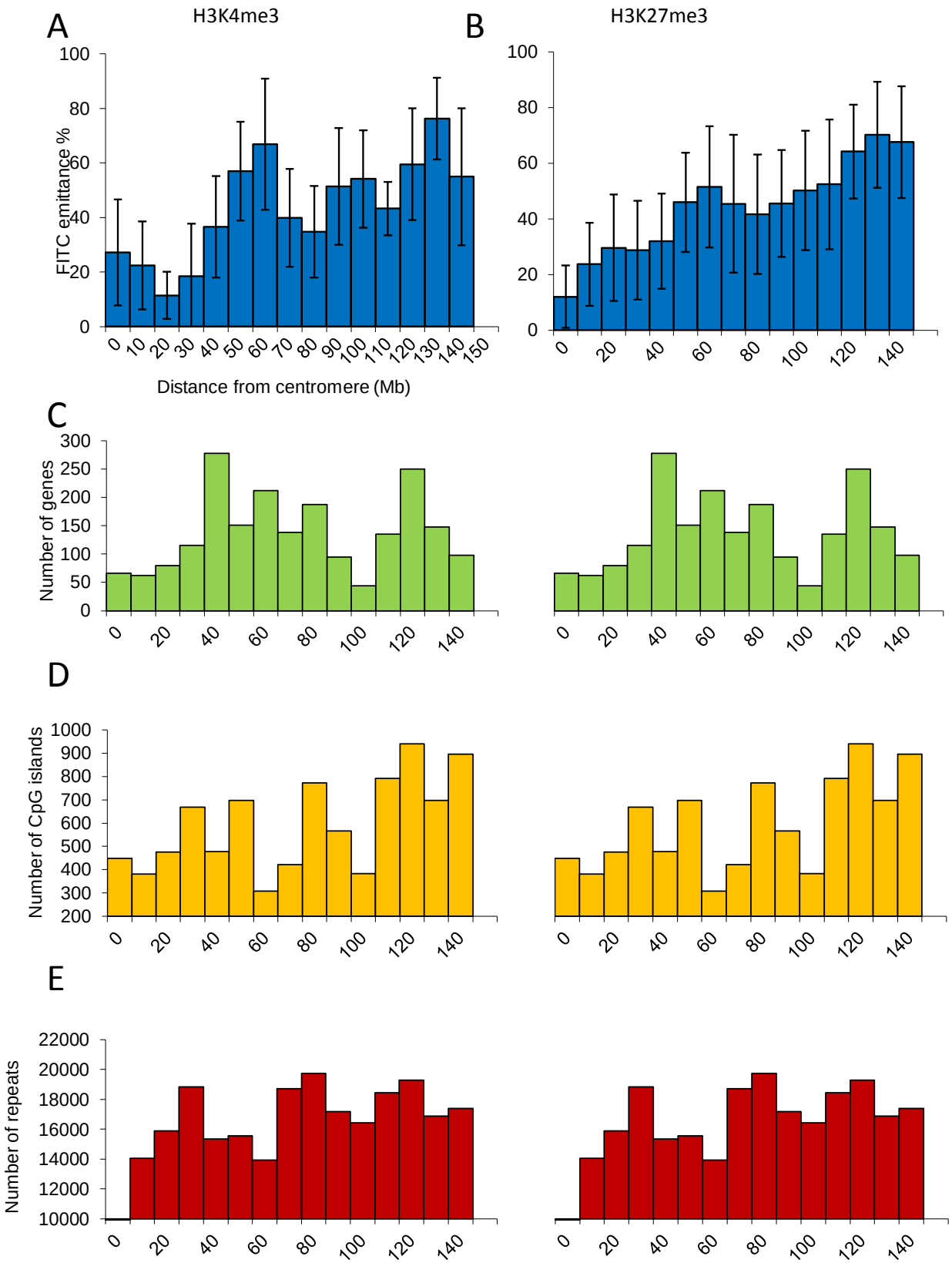


Figure 3.31 Alignment of **A.** H3K4me3 and **B.** H3K27me3 FITC emittance data from chromosome 6 with genomic features. Shown here are the NCBI data for **C.** genes, **D.** CpG islands and **E.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates

These observations highlight the diversity of the data and the amount of information that can be extracted from an epigenetic karyotype of the entire mouse genome. The relationship between histone modifications and DNA sequence is not fully understood, and further investigation into the links between genetics and epigenetics is warranted in order to discover the mechanisms involved in maintaining this delicate balance in eukaryotic systems.

4. RESULTS – NATIVE CHROMATIN IMMUNOPRECIPITATION

4.1 INTRODUCTION TO N-CHIP RESULTS

4.1.1 Epigenetic control of differentiation

Epigenetic processes are thought to be at the heart of the progress of differentiation. The production of a multitude of different cell types from a single progenitor is due to the epigenetic profiles of those resultant cells being different enough from each other that they have their own individual identity, or gene expression profile (Niwa, 2007). This is programmed not by their genetic sequence (which for all these cells are identical) but by their various histone modifications and DNA methylation patterns, for example. Although it has been implied that the bivalent markers H3K4me3 and H3K27me3 are part of the mechanism that alters gene expression on differentiation, it is not known exactly how these modifications change during differentiation, and what role other modifications have in this process (Bernstein *et al.*, 2006, Khromov *et al.*, 2011, Vastenhouw and Schier, 2012). This presents an opportunity to investigate the link between histone modifications and gene expression more closely, as any predictive properties that the epigenetic marks may have can be probed by assaying these marks over a period of time whilst cells are differentiating. To discover what association histone modifications have with gene expression over the time of the differentiation course, it was desirable to assay the gene expression of ES cells over a period of seven days. To determine how these marks change through the cell cycle, two cell populations were used; an asynchronous population and one treated with colcemid to stall cells at the G_2 and M phase transition (Kleinfeld and Sisken,

1966, Muntion *et al.*, 2012). Using these two populations of cells allows the determination of the effect of the shift in gene expression between S phase and M phase of the cell cycle.

Figures 4.1 and 4.2 show the cell cycle profiles for asynchronous and metaphase cells between 0 and 7 days of differentiation. The asynchronous population displays a classical stem cell distribution, with a large proportion of cells in S phase, and at least as many cells in G_2 as in G_1. Over the period of the time-course the profile begins to change, as a greater proportion of cells become stalled in G_1, and the percentage of cells in S and G_2 phases reduces. At d7 the cell cycle profile more closely resembles a differentiated cell than an ES cell, with a large peak in cell numbers at the G_1 phase, followed by a smaller G_2 peak, and a lower proportion of S phase cells. **Figures 4.1 and 4.2** show that colcemid-treated cell samples all contain a majority of cells in G_2 or M phase, far in excess of those found in the asynchronous samples at every timepoint.

4.1.2 Hoxa, pluripotency and histone modifications

The *Hoxa* cluster of genes is responsible for regulating the vertebral segmentation of the developing embryo and for controlling the early differentiation of the mammalian embryo (Lewis, 1978, Durston *et al.*, 2011). The cluster is organised so that each gene is arranged on the chromosome in order of how early in development it is expressed. As such, the cluster acts as a block of many genes, which is evident in its epigenetic signature, often displaying a histone modification, i.e. H3K27me3 being present throughout the cluster (Kashyap *et al.*, 2011). The *Hox* clusters represent an ideal model of study into the process of epigenetic changes through differentiation as it has a well-defined on-off switch at an early

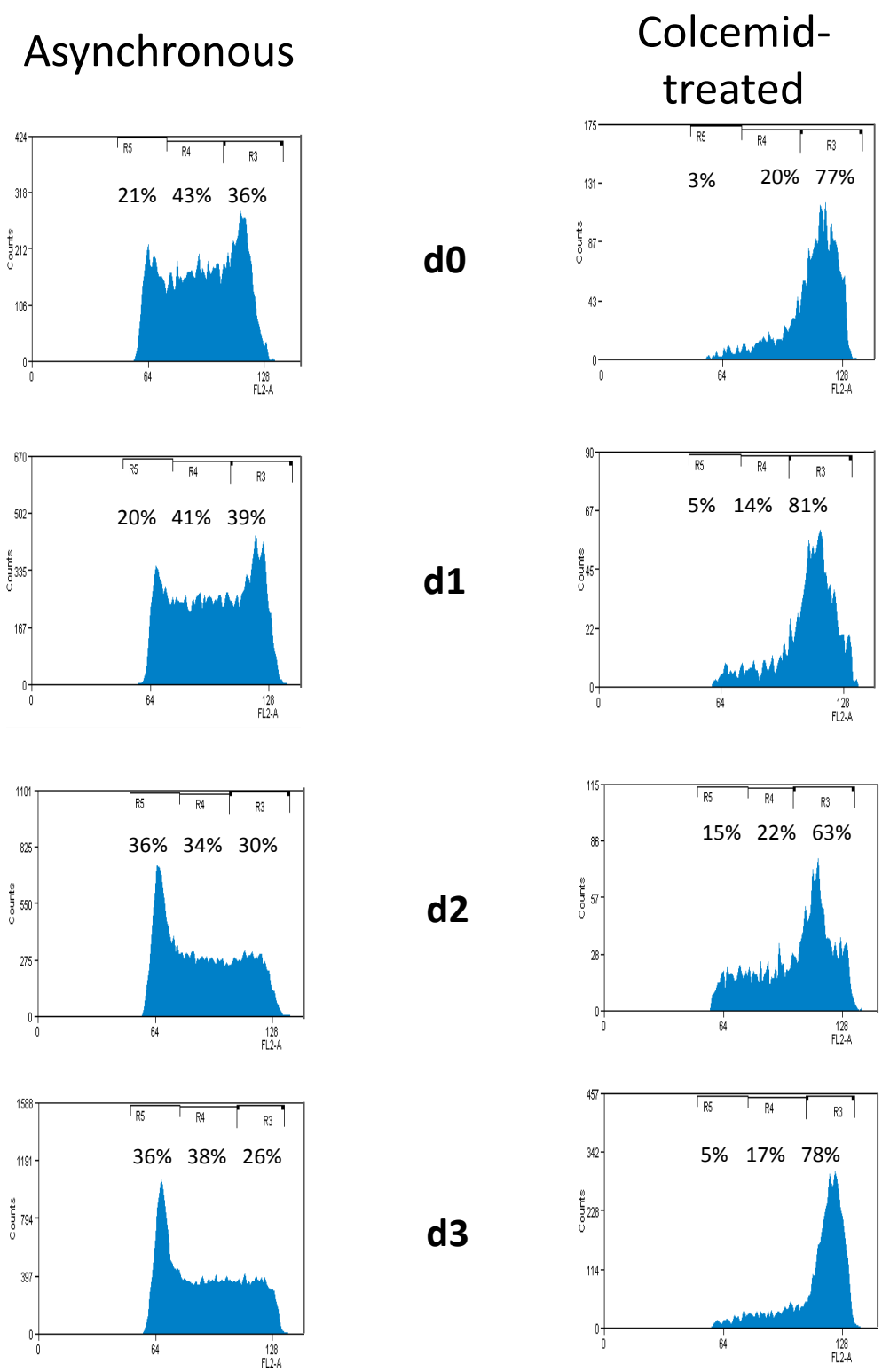


Figure 4.1 Graphs showing the DNA content of asynchronous and colcemid-treated CCE/R cells fixed with acetone and stained with propidium iodide. 2N cells (G_1) are distributed within R5, 4N G_2 and mitotic cells within R3, and S phase cells within R4. Percentage of cells in each population are also shown. D1-3 represents 1-3 days of differentiation, i.e 1-3 days after the removal of LIF. D0 cells are undifferentiated.

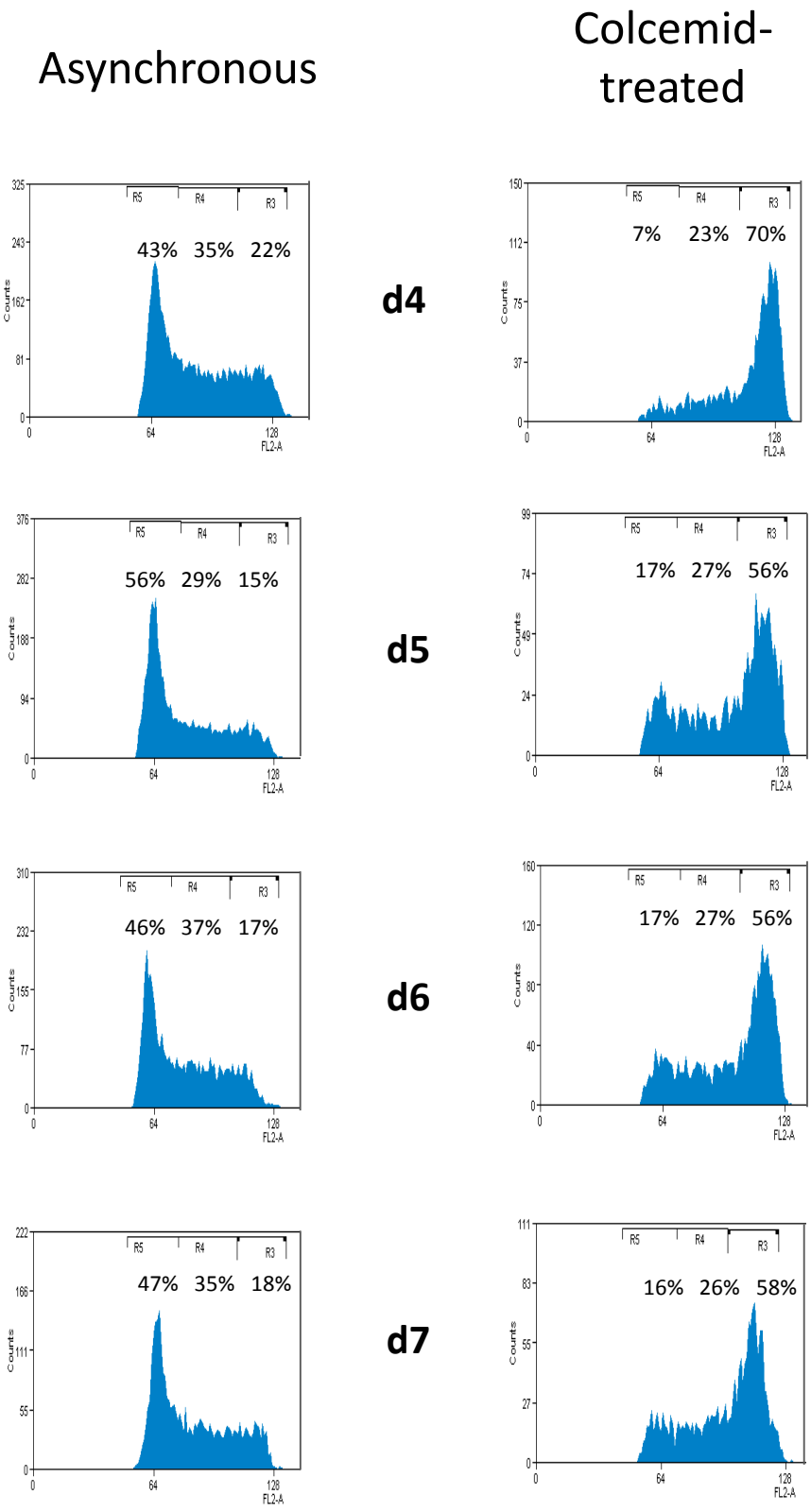


Figure 4.2 Graphs showing the DNA content of asynchronous and colcemid-treated CCE/R cells fixed with ethanol and stained with propidium iodide. 2N cells (G_1) are distributed within R5, 4N G_2 and mitotic cells within R3, and S phase cells within R4. Percentage of cells in each population are also shown. D4-7 represents 4-7 days after the removal of LIF.

stage of differentiation. The temporally expressed nature of each of the *Hoxa* genes also provides a further insight into the epigenetic signature that is associated with a future, past or present expression profile.

Hoxa genes generally present a bivalent motif at the transcription start site (TSS). This is the co-incident trimethylation of H3K4 and H3K27, and is often found at lineage-specific genes required after differentiation (Bernstein *et al.*, 2006, Vastenhouw and Schier, 2012). It is currently unknown to what extent these domains are critical in the control of differentiation, but it is theorised that the repressive H3K27me3 mark inhibits transcription of these lineage specific genes at the undifferentiated stage, and then that this mark is lost during differentiation to leave the underlying H3K4me3 histone modification to promote expression of the genes at the appropriate time (Bernstein *et al.*, 2006, Khromov *et al.*, 2011). As *Hox* genes fit into the category of developmentally regulated genes, it will be interesting to find out what epigenetic motifs occur at this locus before and after differentiation of embryonic stem cells.

Pluripotency can be thought of as a poised state whereby an embryonic stem cell remains uncommitted to any particular developmental path. This state is controlled by a few key genes and transcription factors like Oct4 (*Pou5f1)*, Nanog and Sox2 (Boyer *et al.*, 2005, Mikkelsen *et al.*, 2007). These proteins bind enhancers of most genes required for maintaining the pluripotent state, and also promote their own and each other's transcription to maintain a positive-feedback loop (Whyte *et al.*, 2013). A reduction in expression of these factors, usually due to environmental cues both extrinsic and intrinsic, leads to rapid differentiation and loss of pluripotency.

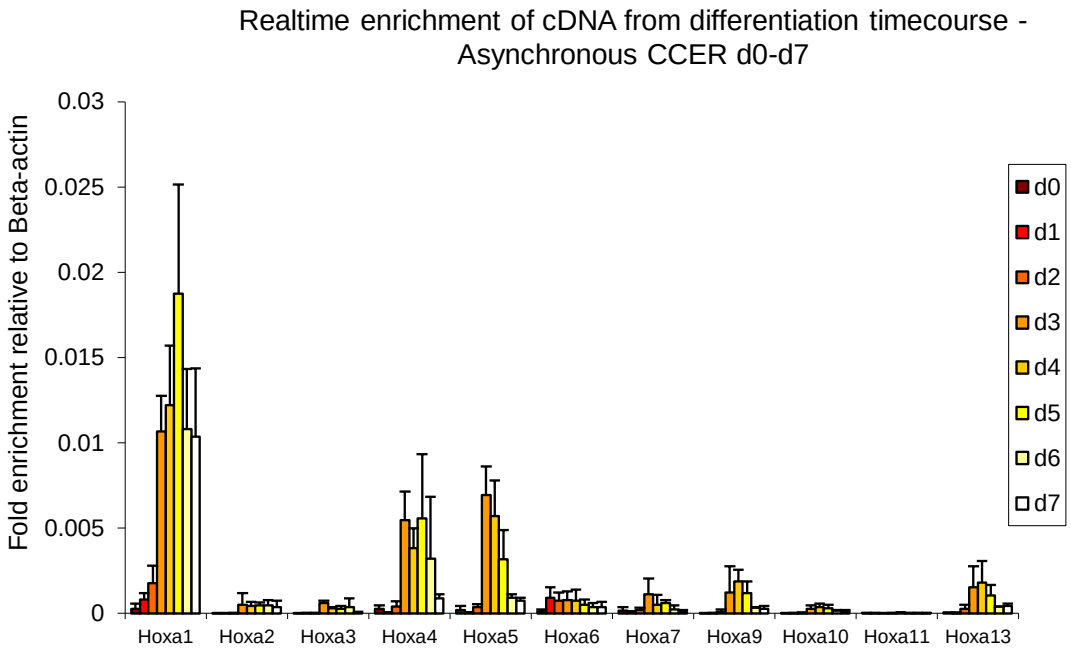
4.2 RNA EXPRESSION OF ES CELLS THROUGH DIFFERENTIATION

4.2.1 Analysis of ES cell expression changes through differentiation at the Hoxa cluster

Asynchronous cells were grown without LIF for seven days and sampled at the same time on each day during that period. Retinoic acid was added to the cells at d2, in order to stimulate differentiation of the stem cells and to encourage *Hox* expression (Chen and Gudas, 1996, Bami *et al.*, 2011, Kashyap *et al.*, 2011). Some cells were incubated with colcemid for 24hrs before harvesting, and some were left untreated as an asynchronous control group. RNA was extracted and reverse-transcriptase PCR was used to determine the levels of transcript for genes in the *Hoxa* cluster. This cluster was used as it is known to display higher levels of expression after differentiation, and it has also been found to be marked with the bivalent histone modifications H3K4me3 and H3K27me3, giving it a known identity as a differentiation-regulated gene cluster (Bernstein *et al.*, 2006, Kashyap *et al.*, 2011).

Data from **figure 4.3** shows that at day 0, expression of all genes in the *Hoxa* cluster is extremely low compared to the expression of the housekeeper gene β-actin. As shown by previous studies, the expression of *Hox* cluster genes is switched off in ES cells because the homeobox proteins are not needed until later in the differentiation process (Dolle and Duboule, 1989, Durston *et al.*, 2011). Therefore, it is likely that this level of expression is due to spurious transcription and that these genes are effectively silent. Some of the genes in this cluster show consistently elevated expression compared to the others on all eight days included

A



B

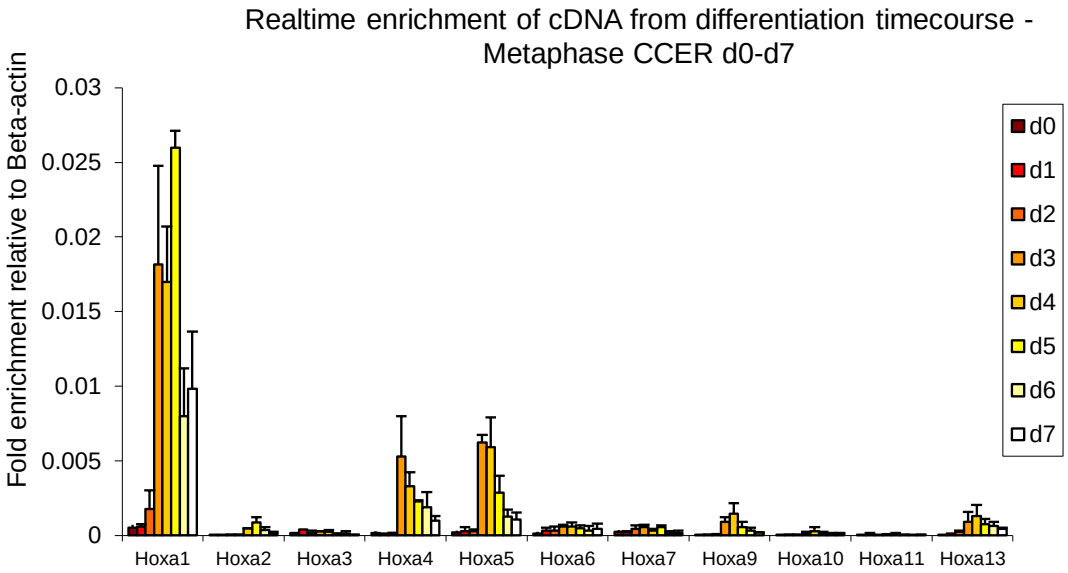


Figure 4.3 Real-time expression levels of Hoxa cluster genes in day 0 to day 7 differentiating CCE/R cells. **A** shows expression levels (y-axis) of Hoxa genes (x-axis) relative to ß-actin in asynchronous cells. **B**. Expression of Hoxa genes in colcemid-treated metaphase cells. Error bars show the standard error of the mean between three biological replicate experiments

in this study. For example, *Hoxa* 1, 4 and 5 are enriched over any of the other *Hoxa* genes, and is particularly evident in the case of *Hoxa1* versus *Hoxa11*, indicating that the former group may be beginning to be expressed over the 7 day period of the time-course, whereas the latter group remain inactive. *Hox* genes are normally classified in order of activation, i.e. *Hoxa1* would be first expressed before *Hoxa2*, and so on (Durston *et al.*, 2011). In accordance with this dogma, these results show that *Hoxa1* is rapidly expressed soon after the differentiation process had begun, as the enrichment increases from d0 to d1, then increases to peak at day 5 after LIF withdrawal, a value nearly 100 times higher than the baseline.

The fact that *Hox* expression increases quickly after addition of retinoic acid (RA) at d2 is a well-known feature of the RA response in ES cells (Kashyap *et al.*, 2011), however the increase in expression of *Hoxa* genes such as *Hoxa1* before RA has been introduced is unexpected and occurs well before the physiological requirement of the gene products. The relatively low levels of many *Hoxa* transcripts, even after d2, may indicate that the increase is not high enough to be effective *in vivo*, however it may point to a mechanism whereby a gradual build-up of transcript throughout differentiation may result in a positive feedback switch being turned on.

In contrast to other genes in the cluster; *Hoxa2* expression is only slightly enriched after 3 days of differentiation and maintains this low level of expression for the rest of the time-course. Activation of *Hoxa* genes occurs most frequently at day 3, with examples including *Hoxa2*, 3, 4, 5, 7, 9, 10 and 13, corresponding to 24h after RA was introduced into the cell culture medium. These genes display fluctuating levels of expression relative to each other; however they all share the

characteristic of an obvious leap in transcript levels between days 2 and 3. *Hoxa6* shows a drop in gene transcription after 5 days, which is also reflected in several other *Hoxa* genes, indicating that there is only a small window in which the peak in *Hoxa* gene expression is maintained, and that the effects of the RA have begun to wear off. This may indicate a small artificial response to RA that increases expression slightly in otherwise completely silent *Hoxa* genes.

4.2.2 Comparison between asynchronous and G_2/M stalled cell expression

The replication fork of the synthesis phase disrupts the histone modification landscape, and so in order to investigate fully the potentially altering effect of the cell cycle, it was desirable to analyse expression data before and after this point. The asynchronous population represents the “before synthesis” sample, as the majority of its cells are in G_1 or S phase. The colcemid-treated population contains mostly G_2 and M phase cells (**figures 4.1 and 4.2**), and is ideal for studying the expression of genes between DNA replication and cell division. G_2/M expression is shown in **figure 4.3,** which shows that *Hoxa* genes are expressed at roughly equivalent levels both before and after S phase. Some of the expression patterns from the asynchronous population are retained in G_2/M, such as genes *Hoxa1*, *4* and *5*, which display the highest levels of expression, and *Hoxa2, 3, 10* and *11,* which maintain particularly low levels of RNA transcript. The timings of transcription activation differ somewhat, however, as is evident in results for *Hoxa1*. The activation of this gene is evident in the G_2/M stalled populations, showing a large increase in enrichment between d2 and d3. Expression goes on to peak at d5, the same day as the asynchronous population, and at a similar enrichment value.

The data shown here confirm that for many of the genes in the *Hoxa* cluster, there is no great difference between expression in asynchronous populations and for G_2/M stalled populations. This means that any difference in epigenetic marks between pre-synthesis and post-synthesis chromatin would be due to the mechanisms of chromatin maintenance and not fluctuations in gene expression. It also suggests that any disruption in the epigenetic landscape that occurs at S phase does not contribute to a transient up- or down-regulation of gene transcription activity.

Another striking feature of these data is the low levels of enrichment of all of these *Hox* genes relative to the housekeeping gene *β-actin*. Even the highest level of expression of any *Hox* gene during the time-course was less than a tenth of the level of actin transcription. This may be because only small numbers of *Hox* transcript are necessary to complete its function of binding to DNA and to affect transcription of differentiation-related genes. In other words, more copies of β-actin RNA are needed as its product is required at all times throughout the cell in order to maintain its cytoskeleton. A gene such as *Hox* is needed only transiently, and perhaps not in such large quantities as actin, in order to fulfil its function. This is reflected in the expression results for the two pluripotency genes *Nanog* and *Pou5f1*. Both genes are involved with the maintenance of the undifferentiated state in ES cells and both are down-regulated once the cell begins to differentiate. This is reflected in these data that show a pronounced drop in *Nanog* expression in asynchronous cells at d3, and in *Pou5f1* expression at d2, and then further at d5. The enrichment of *Pou5f1* is lower than *Nanog* throughout the time-course despite its vital role as a nuclear transcription factor; however a clear pattern of transcription down-regulation is still visible. This suggests that either fewer copies

of Oct4 are required to maintain its functionality, or that the *Pou5f1* mRNA is more stable than that of *Nanog* and produces more copies of the protein from each mRNA molecule.

4.2.3 Expression of housekeeping and Hoxa neighbouring genes from d0 to d7

In order to examine whether the expression patterns of *Hoxa* are replicated in the surrounding chromatin environment, **figure 4.4** also shows expression of *Skap2* and *Evx1*, genes which are adjacent to the *Hoxa* cluster. Expression of *Skap2* is constant throughout differentiation, as expected for a housekeeping gene (Shimamura *et al.*, 2013). *Evx1* expression is at a consistently low level throughout the time-course as shown by the values of enrichment relative to *Beta-actin*. Neither of these boundary genes shows any response to retinoic acid and so remains representative of loci close to *Hoxa* whose expression is not affected by differentiation during this time-course.

After having analysed the expression of genes within a developmentally-regulated gene cluster, I then went on to assay the association of various histone modifications with these sequences to determine whether or not the changes in gene expression are influenced or predicted by the epigenetic background of the chromatin.

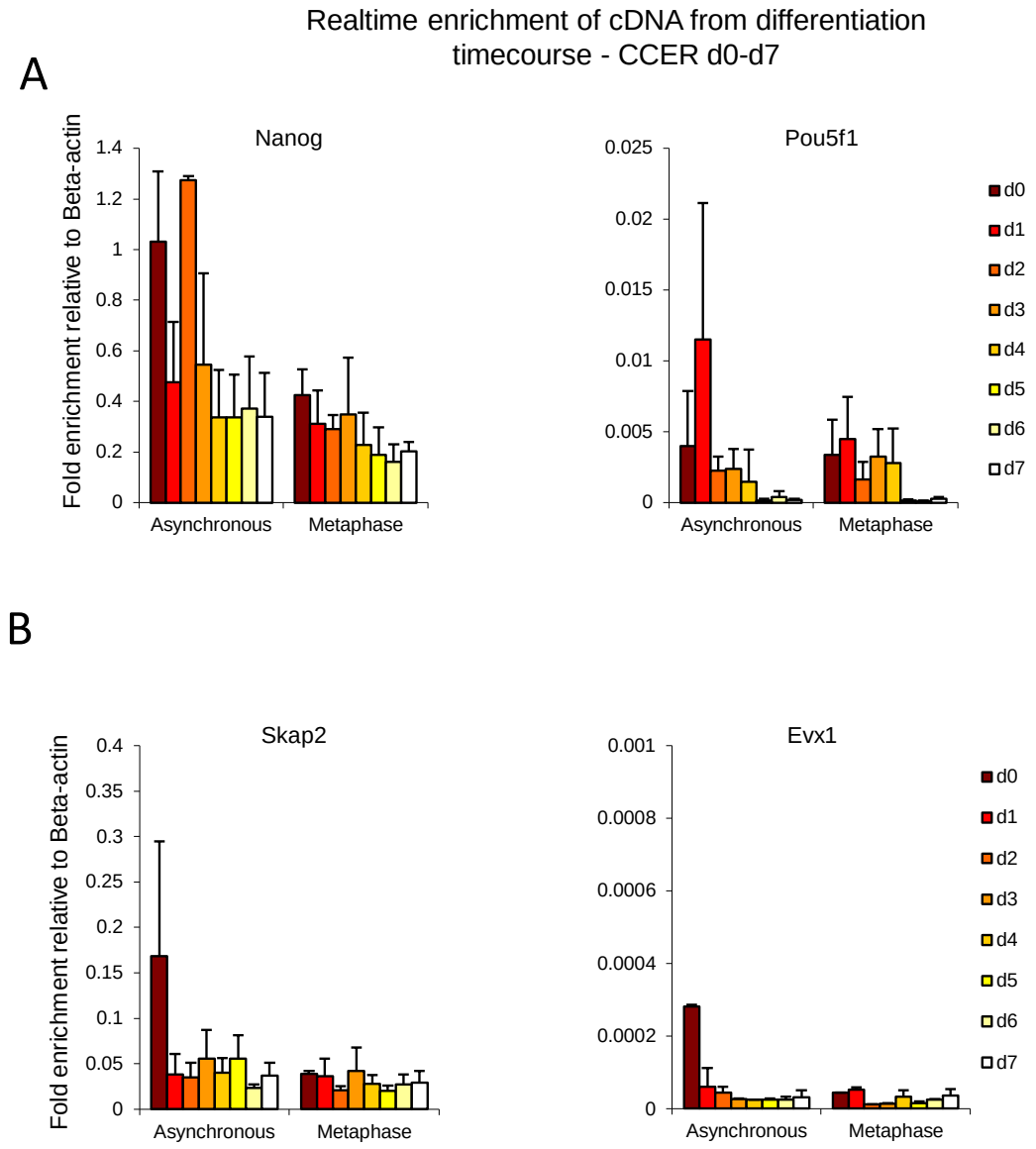


Figure 4.4 Real-time expression levels of day 0 to day 7 differentiating CCE/R cells, both untreated and cells treated for 24h with 5µl/ml colcemid. **A.** Expression of pluripotency-related genes. **B.** Expression of Hoxa cluster-neighbouring genes. D0-d7 represents days that cells were cultured in the absence of LIF (1µM retinoic acid added at d2). Real-time enrichments are relative to ß-actin transcript levels as determined by real-time PCR. Error bars show the standard error of the mean between three biological replicate experiments

4.3 HISTONE MODIFICATIONS IN ES CELLS

4.3.1 Analysis of histone modification enrichment by chromatin immunoprecipitation

Epigenetic information must be reconstituted after every cell cycle in order to be maintained as a heritable mark (Turner, 2007). The two critical phases that histone modifications must pass through are mitosis phase and synthesis phase. Mitosis phase, as described previously, presents a problem for the enzymatic maintenance of histone marks as transcription is silenced during mitosis and chromatin has a reduced accessibility to enzymes, so any modifications present on chromatin at this time would have to survive the compression and subsequent relaxation of chromosomes associated with M phase.

Synthesis, or S phase, represents another obstacle for the maintenance of epigenetic marks, because information must be copied from the old chromatin strand to the new. The mechanism by which this occurs is not fully understood, but it is thought that new histones go through a specific sequence of modifications that culminate in the correct marks being maintained on both sets of chromatin. However, it is not known whether this process occurs exactly faithfully, or at what time in the cell cycle the epigenetic profile is reconstituted. It is also not known whether the distributions or positions of these histone modifications are shifted at some point during the cell cycle, or if marks are quickly and accurately deposited onto new chromatin. To establish this, I analysed the genetic locations of histone modifications in cells which had been stalled in G_2/M phase, and compared them with asynchronous cells. This was done by using the N-ChIP technique with

antibodies to histone modifications H3K4me3, H3K27me3, H3K9me2, H3K9ac, H3K27ac and H4K20me3.

4.3.2 Enrichment of histone modifications at Gapdh, Nanog and Pou5f1

Figure 4.5 shows the level of enrichment of 6 different histone modifications on the housekeeper gene *Gapdh* and the pluripotency genes *Nanog* and *Pou5f1* in both asynchronous and G_2/M CCE/R cell populations. H3K4me3 enrichment does not change between asynchronous and G_2/M samples, either at *Gapdh* or *Pou5f1*. H3K4me3 increases at G_2/M at *Nanog*, however, and methylation of this mark at *Nanog* is higher than both *Pou5f1* and *Gapdh*. This reflects the high level of *Nanog* expression seen in **figure 4.4**, especially with respect to *Pou5f1*. All three of these genes are expressed in ES cells, and so the repressive mark H3K27me3 is depleted at all three marks, a state which does not change through the cell cycle. The repressive mark H4K20me3 is also depleted at these promoters, although H3K9me2 does show a moderate level of enrichment at all three loci. These histone modifications also do not change in enrichment as the cell cycle progresses. Although H3K9ac is enriched above the unbound fraction at each of these loci, this mark does appear to be reduced in the G_2/M population, as shown in **figure 4.4**. This shows a reduction in H3K9ac at *Gapdh*, *Nanog* and *Pou5f1* in the colcemid-treated cells as opposed to asynchronous cells. *Nanog* again displays a higher enrichment of this activating mark than *Pou5f1*, reflecting its increased transcription. This relationship is reversed, however, for another acetyl mark H3K27ac, which displays a greater level of enrichment at *Pou5f1* rather than

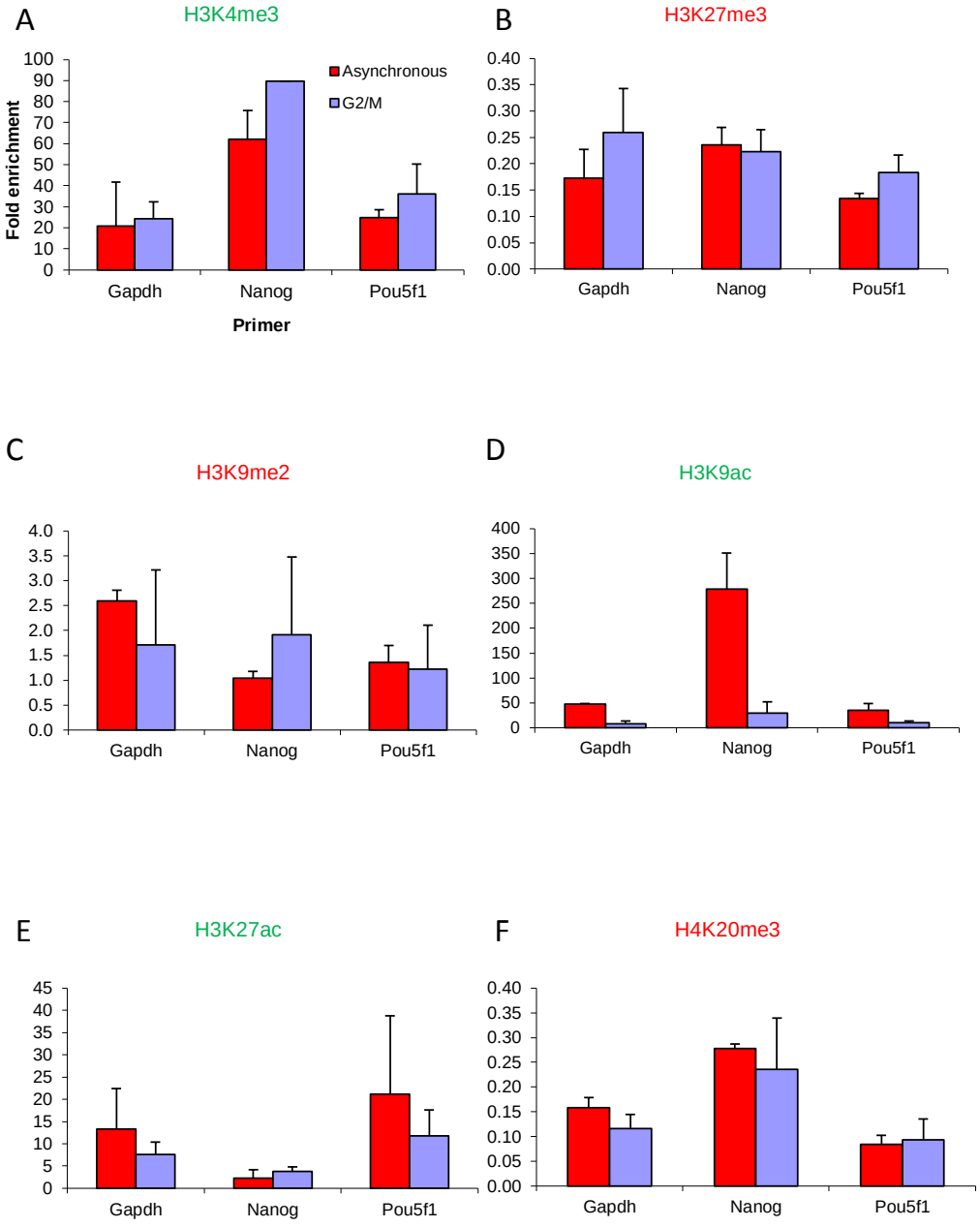


Figure 4.5 Enrichment of DNA extracted from ChIP pulldowns with **A.** H3K4me3 **B.** H3K27me3 **C.** H3K9me2 **D.** H3K9ac **E.** H3K27ac and **F.** H4K20me3 antibodies. Asynchronous denotes undifferentiated asynchronous CCE/R cells, G2/M indicates the same cells treated for 6 hours with 5µl/ml colcemid. Primers used are specific for the promoters of the metabolic gene Gapdh and the pluripotency-related genes Nanog and Pou5f1. Activating marks are highlighted in green, repressive marks in red. Error bars show the standard error of the mean between three biological replicate experiments

Nanog. This might indicate that H3K27ac is less directly reflective of expression levels. *Gapdh*, *Nanog* and *Pou5f1* all show enrichment of H3K27ac and no clear differences between asynchronous and G_2/M cells can be determined for this histone modification.

4.3.3 Enrichment of H3K4me3 and H3K27me3 across the Hoxa cluster

Results from **Figure 4.6** show that H3K4me3 is not distributed evenly across the entire *Hoxa* cluster in either asynchronous or G_2/M cells. Firstly, it is clear that there are no significant differences between the asynchronous and G_2/M samples across the cluster. Upstream of *Hoxa1*, trimethyl K4 shows a very low level of enrichment, revealing not only that it is depleted in intergenic sequences, but also that some gene promoters – in this case the housekeeper *Skap2* – are not marked by elevated H3K4me3 levels despite the relatively high expression of *Skap2* in these cells (**figure 4.4**).

Trimethyl H3K4 is enriched at *Hoxa1* compared to the unbound fraction control. However, at *Hoxa2*, the level of K4me3 enrichment is much lower. This reflects the future expression patterns of these two genes, as *Hoxa1* is expressed after d2 of differentiation whereas transcript levels of *Hoxa2* are not raised significantly throughout the time-course. This pattern is carried through into *Hoxa3* and *Hoxa4*, which show low and high levels of K4me3 respectively, corresponding to their low and high future expression in differentiating ES cells. The association between the level of expression and the level of methylated H3K4 continues through the rest of the *Hox* genes, as H3K4me3 enrichment drops from *Hoxa4* to *Hoxa5*, and drops

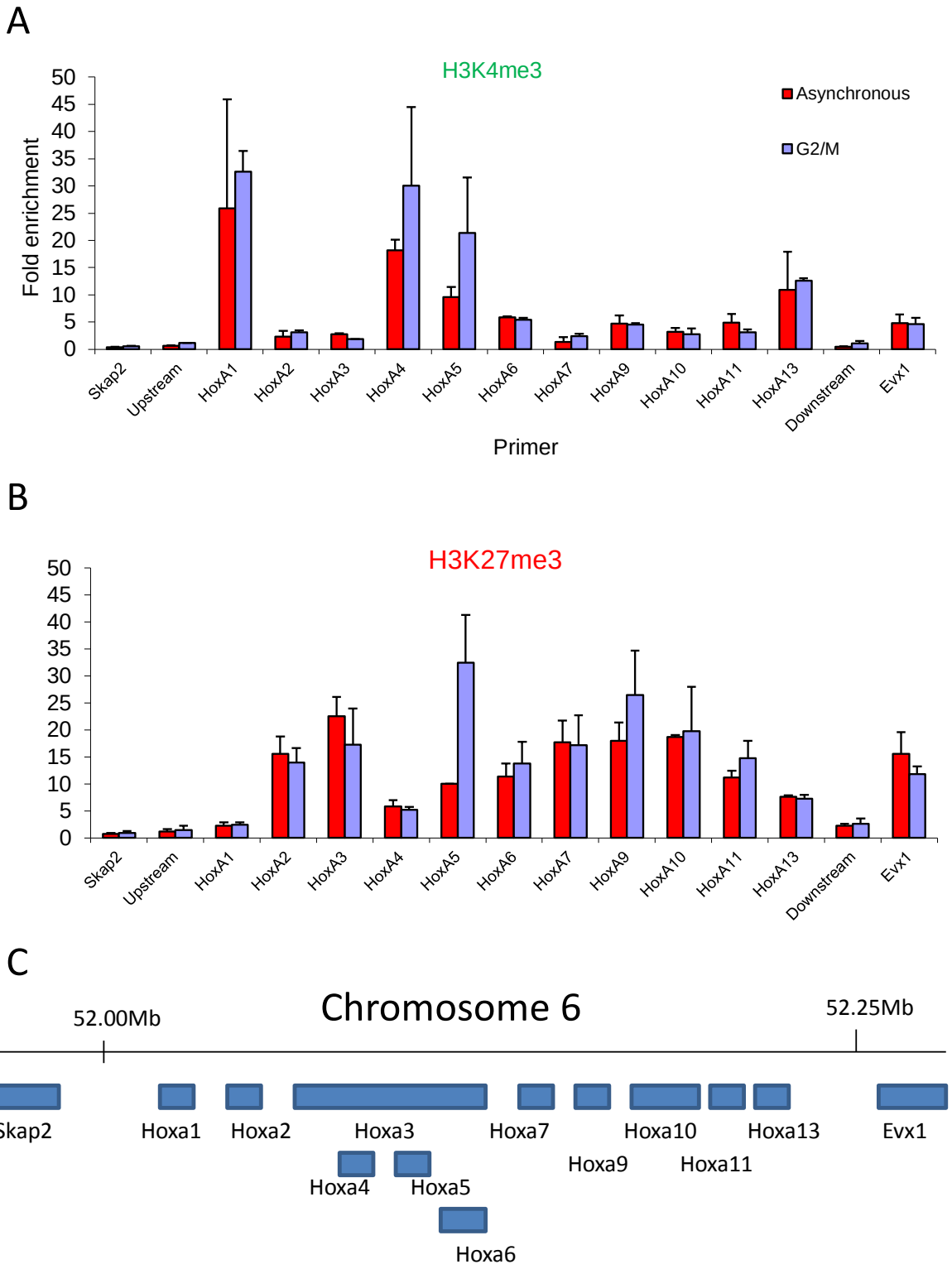


Figure 4.6 Enrichment of DNA extracted from ChIP pulldowns with **A.** H3K4me3 and **B.** H3K27me3 antibodies. Asynchronous denotes undifferentiated asynchronous CCE/R cells, G2/M indicates the same cells treated for 6 hours with 5µl/ml colcemid. Primers used are in and adjacent to the HoxA cluster on mouse chromosome 6. Primers are specific for the promoter of the labelled gene except for upstream and downstream primers which are complementary to intergenic sequences. **C.** Diagram of genome browser data showing the relative positions of Hoxa genes and neighbouring genes within the Hoxa cluster on mouse chromosome 6. Gene positions not to scale. Error bars show the standard error of the mean between three biological replicate experiments

further across *Hoxa6*, 7, 9, 10 and 11, before rising at *Hoxa13*. There is not a perfect correlation between *Hox* gene activity and H3K4me3 enrichment, as the transcription of *Hoxa7* is much greater than that of *Hoxa9*, which is not reflected in the ChIP data that shows *Hoxa7* as being the most trimethyl K4-depleted gene among the cluster. Another observation is that *Hoxa11* shows enrichment in this histone modification that is equivalent to several other genes in the cluster, whereas shown in **figure 4.3** is the clear repression of *Hoxa11* transcription at each day sampled compared to every other *Hox* gene. There is not a simple relationship between transcription and histone methylation, although these data do reveal a strong link implying H3K4me3 increases as a pre-emptive mark, as future transcription of underlying DNA increases.

Intergenic sequences both upstream and downstream of the *Hoxa* cluster have very low enrichments (**figure 4.6**); however, the closest genes on each side of *Hoxa* differ in their enrichment of trimethyl K4. *Evx1* has an enrichment of several times the unbound in both asynchronous and G_2/M populations and is equivalent to some *Hox* genes like *Hoxa9*, whereas enrichment for *Skap2* is less than 1, indicating a depletion of H3K4me3 at that site. This indicates that *Evx1*, a developmentally regulated gene, shows a similar enrichment of H3K4me3 to *Hoxa* genes, whereas *Skap2*, a housekeeping gene, does not (Dush and Martin, 1992, Fujioka *et al.*, 2003).

Data from N-ChIP experiments using H3K27me3 antibodies reveal more about the epigenetic landscape of the *Hoxa* cluster in ES cells. Again, the enrichment of this mark does not change significantly between asynchronous and G_2/M phase cells, with the possible exception of *Hoxa5*. This gene has a small enrichment in H3K27me3 in G_2/M compared to the asynchronous population. This is also

reflected in the results for H3K4me3, which show another slight increase. This difference is not reflected in the expression data, which shows a steady level of detected transcript between asynchronous and G_2/M cells. In fact, the expression time-course for *Hoxa5* shows a particularly close congruence between enrichments at each sampled day between both sets of cell populations. Therefore, any difference in the levels of histone modifications is not governed by changing expression, and vice versa.

Another striking feature of the H3K27me3 data is that it appears to complement the results obtained for H3K4me3. That is to say that genes at which H3K4me3 is enriched, are simultaneously depleted in H3K27me3. This is particularly evident for *Hoxa1*, which is greatly enriched in H3K4me3, but only slightly enriched in H3K27me3. Many other studies have noted the co-existence of H3K4me3 and H3K27me3 at *Hox* promoters, although this data shows a more nuanced picture of the epigenetic state of the promoters in ES cells (Bernstein *et al.*, 2006, Hattori *et al.*, 2013). Enrichment of both these marks relative to unbound fractions suggest that the bivalent motif is present, yet the predominance of trimethyl K4 as opposed to K27 reflects a predictive permissive chromatin environment that is confirmed by the elevated *Hoxa1* expression after differentiation relative to other genes in the cluster. This relationship can also be found inversely at genes *Hoxa2* and *Hoxa3*. These genes have low transcription relative to *Hoxa1*, and also show low levels of H3K4me3 and high levels of H3K27me3 relative to *Hoxa1*. Data from these experiments imply that the bivalent domain does not involve an undifferentiated state characterised by equal levels of trimethyl K4 and K27, but rather that at bivalent loci both modifications are often present, but with one or the other mark in clear prevalence over the other, determining its transcriptional output. The outlying

DNA sequences also reveal information about the bivalent mark. The intergenic sequences are devoid of both modifications, confirming the association of these marks with gene-promoter regions and not intergenic DNA (Bernstein *et al.*, 2006, Roudier *et al.*, 2011). Both marks are also depleted at *Skap2*, a gene which is expressed in these cells but which is not thought to be related to the process of differentiation. *Evx1*, on the other hand, contains both H3K4me3 and H3K27me3 and is more heavily enriched in H3K27me3. This pattern is similar to the level of modifications present at *Hoxa2*, and like this gene, *Evx1* shows a low level of expression at day 0. *Evx1* is also closer in terms of locus to the *Hoxa* cluster than *Skap2*, however both the upstream and downstream regions between the cluster and these outlying genes show depletion of histone modifications, confirming that these observations are not due to a motif from the *Hoxa* cluster indiscriminately spreading to outlying genes, but that another discrete bivalent mark has been deposited on *Evx1* to influence its epigenetic signature. Bivalency is thought to play a major role in the control of differentiation, and indeed *Evx1* is a gene which has been implicated in the repression of transcription related to the control of the differentiation process (Dush and Martin, 1992, Fujioka *et al.*, 2003). So in a sense, the bivalent mark can distinguish the *Evx1* gene that is differentiation related from *Skap2* that isn't, therefore implicating epigenetic factors in predictively marking, or even determining, the behaviour of underlying genes by the combination of histone modifications on chromatin.

As ever, the number and variety of epigenetic modifications means that a simple pattern of one modification influencing gene transcription alteration is too simplistic. The pattern of expression for undifferentiated cells found here, however, does intimate a strong relationship between bivalent modification enrichment and

gene expression in the *Hoxa* cluster. The pattern of high trimethyl K4 and low trimethyl K27 is shared by *Hoxa1* and *Hoxa4*, whereas the reverse pattern is evident for genes *Hoxa2, 3, 6, 7, 9, 10 and 11*. *Hoxa13* shows only a slight enrichment in H3K4me3 compared to H3K27me3, and *Hoxa5* shows a unique distribution of trimethyl K27. The incongruous results obtained for *Hoxa5* may reflect its unusual position at the confluence of several transcription units in the *Hox* cluster (Coulombe *et al.*, 2010). This may result in the epigenetic signature of this locus being more complex or perhaps differently organised than the other *Hox* loci. This may represent a gap in the bivalently controlled Hox cluster whereby future *Hoxa*5 expression is regulated by some other means.

Despite the expression profiles of *Hoxa4* and *Hoxa5* being very similar, these two genes show differences in their enrichments of histone modifications. Methylation at H3K4 is lower at *Hoxa5* than *Hoxa4*, and H3K27me3 is higher. However, a more obvious feature of these data is that both histone marks show increased enrichment at G_2/M at the *Hoxa5* locus. In asynchronous cells, the levels of H3K4me3 and H3K27me3 are equivalent, whereas at G_2/M the increase in H3K4me3 enrichment is overtaken by a larger increase in H3K27me3. At G_2/M then, the chromatin environment at *Hoxa5* would appear to be repressive, despite the high enrichment of H3K4me3. However, this shift in chromatin modifications does not affect expression at all, as *Hoxa5* expression is equal between asynchronous and G_2/M in undifferentiated cells and in cells at every stage of differentiation during the sampled time-course. It is possible that the information encoded by these histone modifications extends to more than simply gene expression and that the increase in the bivalent mark seen at G_2/M could signal an as yet unidentified pathway to interact with this section of chromatin, perhaps to

assist with some cell cycle-related aspect of the differentiation process. Alternatively, histone modifications at *Hoxa5* may be selectively deposited later than those of other genes, perhaps owing to its transcriptional complexity, therefore appearing to be enriched in G_2 or M phase compared to the rest of the *Hoxa* cluster (Coulombe *et al.*, 2010).

4.3.4 Histone modifications at H3K9 at Hoxa gene promoters

The bivalent mark is well-known to be found at certain developmental loci, but these loci, as well as others, have a variety of histone modifications that are deposited and removed to maintain a fine balance of epigenetic markers that have some control over gene expression as well as a plethora of other cellular processes. Other common marks include modifications on lysine 9 of histone H3, methylation of which is known to be associated with transcription repression and heterochromatin (Puschendorf *et al.*, 2008, Alder *et al.*, 2010). H3K9me2 in particular has been linked to deactivating expression of underlying genes (Rea *et al.*, 2000). K9 acetylation, on the other hand, is commonly associated with activating neighbouring expression and is often found at the promoters and bodies of active genes (Agalioti *et al.*, 2002).

Figure 4.7 shows that the distributions of two histone modifications on H3K9 are very different. H3K9me2 enrichment is constant across the entire *Hoxa* cluster, remaining at a value of around 1 to indicate an average or low prevalence of this histone mark across this region. H3K9me2 enrichment also shows no significant differences between G_2/M and asynchronous populations. Acetylated H3K9, however, shows great variation between genes in the *Hoxa* cluster, and between G_2/M and asynchronous cell populations. High levels of acetyl-K9 throughout the

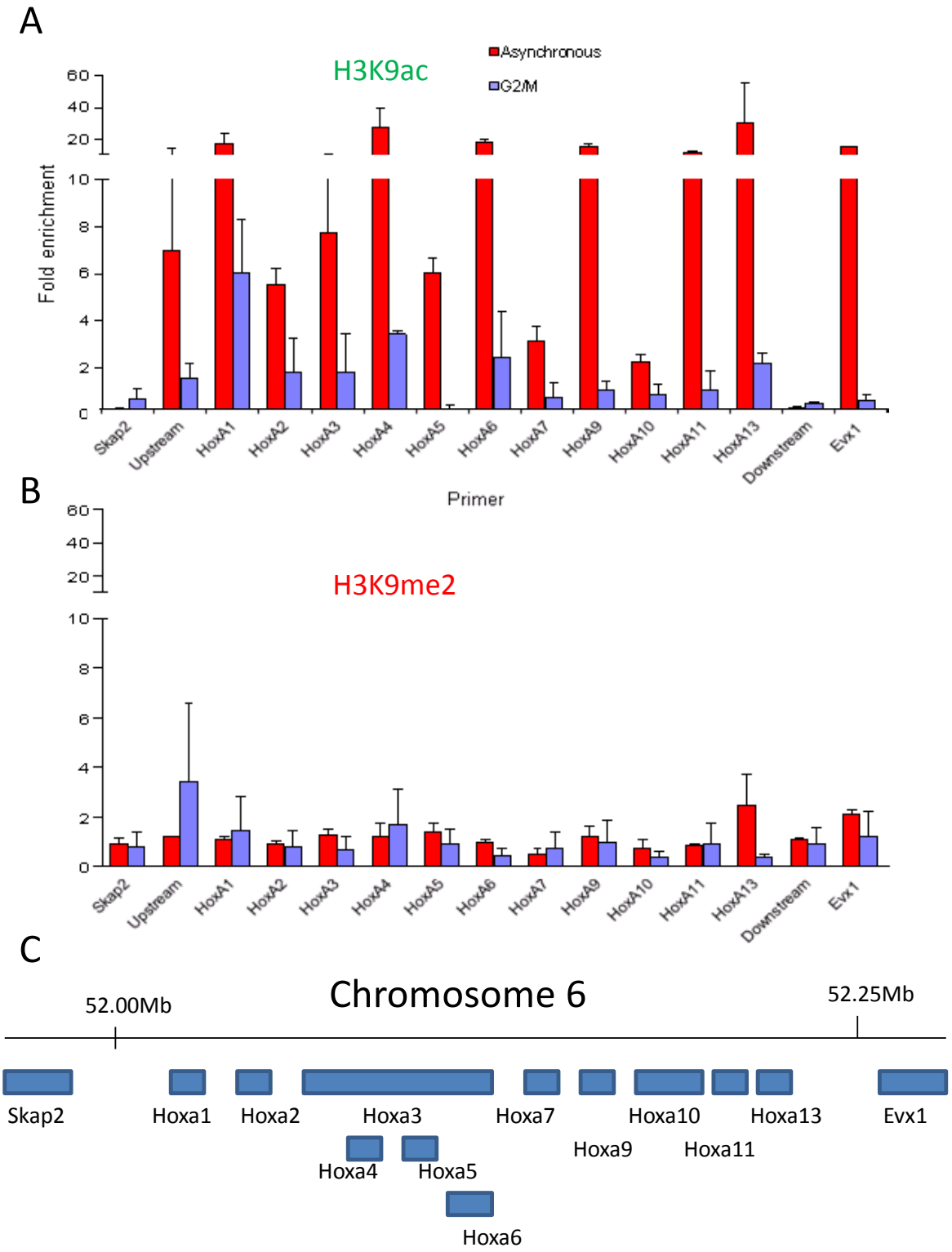


Figure 4.7 Enrichment of DNA extracted from ChIP pulldowns with **A.** H3K9ac and **B.** H3K9me2 antibodies. Asynchronous denotes undifferentiated asynchronous CCE/R cells, G2/M indicates the same cells treated for 6 hours with 5µl/ml colcemid. Primers used are in and adjacent to the HoxA cluster on mouse chromosome 6. Primers are specific for the promoter of the labelled gene except for upstream and downstream primers which are complementary to intergenic sequences. **C.** Diagram of genome browser data showing the relative positions of Hoxa genes and neighbouring genes within the Hoxa cluster on mouse chromosome 6. Gene positions not to scale. Error bars show the standard error of the mean between three biological replicate experiments

Hoxa cluster could explain the exclusion of the methyl modification from this residue. At every locus studied apart from *Skap2* and the downstream intergenic sequence, acetyl K9 enrichment was greater in the asynchronous population than in the G_2/M sample. This included the upstream intergenic region as well as all *Hoxa* genes and the related outlying gene *Evx1*. Unlike the H3K4me3 mark, H3K9ac enrichment does not correlate consistently with gene expression, as *Hoxa5* acetyl-K9 levels are low despite high expression, and the *Hoxa11* locus is highly acetylated (in asynchronous cells) and yet is not expressed in ES cells. The frequency of K9 acetylation changes between *Hoxa* cluster genes in both G_2/M and asynchronous cells, from high enrichment at *Hoxa4* and *Hoxa13* (asynchronous), to low enrichment at *Hoxa10* (asynchronous). Results for G_2/M cells fluctuate even more significantly, as some genes show enrichment in H3K9ac, such as *Hoxa1, 2, 3, 4, 6 and 13*, and some show no enrichment or even depletion in acetylation; i.e. *Hoxa5, 7, 9, 10, 11* and *Evx1*. Genes such as *Hoxa9* show a switch from strong enrichment in the asynchronous population to depletion in G_2/M. These incidences of the loss of acetylation in G_2/M may be due to the rapid turnover of acetyl marks in chromatin. As mitosis phase begins and chromatin becomes less accessible to HAT enzymes, it becomes more difficult for acetyl marks to be deposited (Patzlaff *et al.*, 2010). Therefore, acetyl marks may be lost in G_2/M cell populations more quickly than methyl marks. This process may be the reason behind the reversion at many loci to background levels (i.e. enrichment between 0.5 and 2) of H3K9ac. This also has implications for epigenetic inheritance, as not only would histone modifications need to be re-established after DNA replication, but also be reinstated after mitosis.

4.3.5 H3K27ac and H4K20me3 distributions at Hoxa

In order to investigate modifications on other residues and other chromosomes, I then studied the histone modifications H3K27ac and H4K20me3. I studied K27 acetylation as a counterpoint to methylation on the same residue, as in the case of H3K9, to investigate further whether two modifications on the same histone tail residue can be enriched. As is shown in **figure 4.8**, H3K27ac enrichment is at a constant level of between 0.5 and 2 times the unbound control fraction across the entire *Hoxa* cluster and surrounding DNA. As such, there is no evidence of a co-localisation of acetyl and methyl marks on H3K27. The levels of acetyl-K27, unlike those of acetyl K9, do not change between G_2/M and asynchronous populations, as they show no difference to background (unbound) levels of this modification in both samples, so no specific acetylation has taken place.

I analysed the distribution of H4K20me3 to find out if modifications on another histone would show a different pattern to those of histone H3. I also chose this modification as earlier chromosome immunofluorescence data showed that there was some association of this normally heterochromatic mark with euchromatin across all chromosome arms in ES cells. However, in keeping with its link with heterochromatin, H4K20me3 was found to be heavily depleted across most of the *Hoxa* cluster. A few loci show enrichments of around 1.0, such as *Hoxa6* and *Evx1*, but for many loci the values were well below 0.5-fold of the unbound. On the whole, asynchronous populations and G_2/M cells did not differ, but exceptions such as *Hoxa2, 6 and 7* show a definite preference for H4K20me3 in the G_2/M population. One explanation for this occurrence is that in some cases H4K20me3

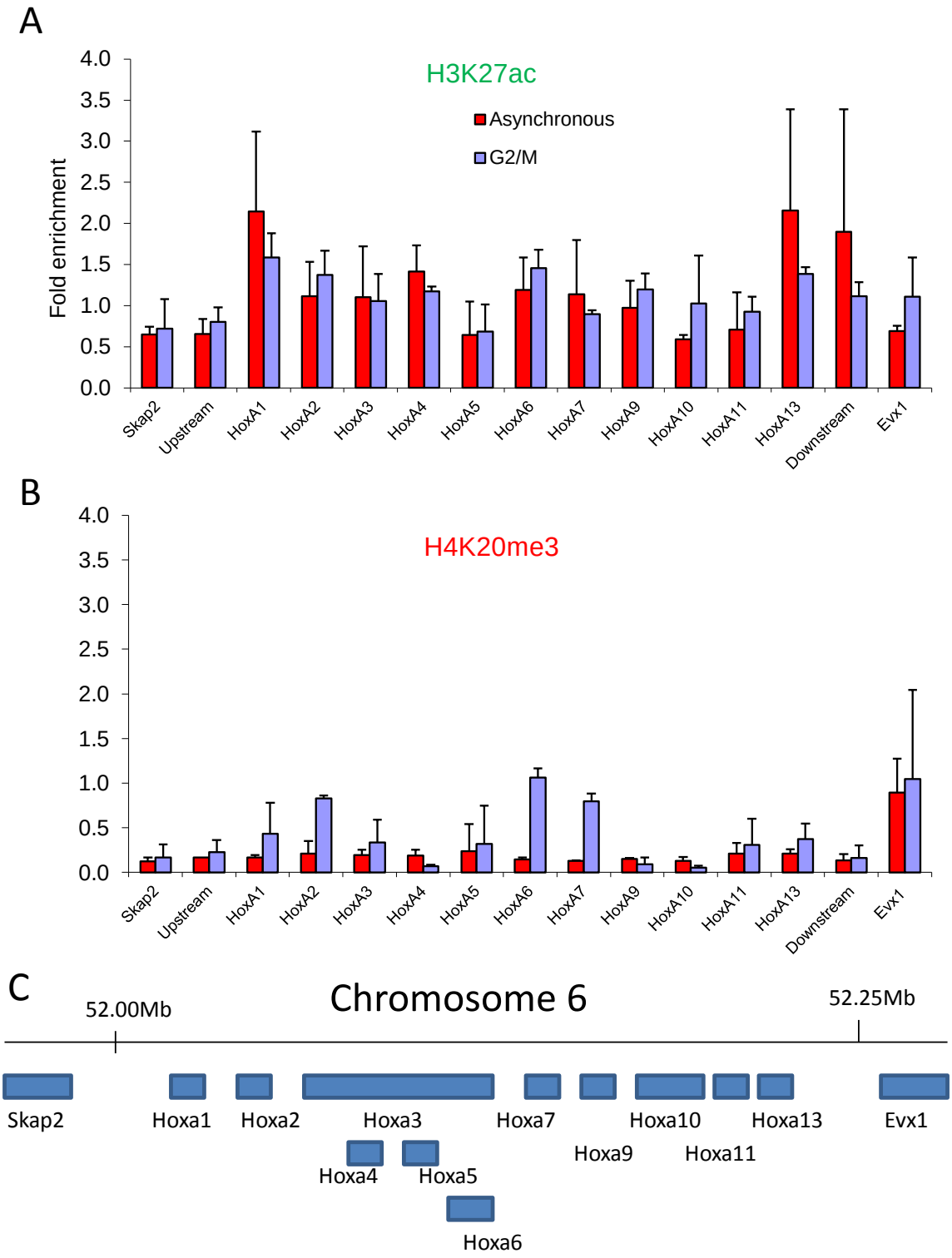


Figure 4.8 Enrichment of DNA extracted from ChIP pulldowns with **A.** H3K27ac and **B.** H4K20me3 antibodies. Asynchronous denotes undifferentiated asynchronous CCE/R cells, G2/M indicates the same cells treated for 6 hours with 5µl/ml colcemid. Primers used are in and adjacent to the HoxA cluster on mouse chromosome 6. Primers are specific for the promoter of the labelled gene except for upstream and downstream primers which are complementary to intergenic sequences. **C.** Diagram of genome browser data showing the relative positions of Hoxa genes and neighbouring genes within the Hoxa cluster on mouse chromosome 6. Gene positions not to scale. Error bars show the standard error of the mean between three biological replicate experiments

is deposited at G_2 phase as part of the process involved in cell division. Once its function is completed, the mark is removed and a subsequent depletion of the mark is found during G_1/S. The fact that even the elevated levels of H4K20me3 at G_2/M is not significantly above the background, however, suggests that these marks are not specifically deposited on these loci after DNA synthesis, and removed after G_2/M phase as part of the normal nuclear checking procedure characteristic of the eukaryotic cell cycle.

These results show clear differences between methyl and acetyl marks on histone H3, possibly highlighting their respective persistence through the cell cycle. The link between histone modifications and gene expression has been shown to be much stronger for the bivalent epigenetic markers H3K4me3 and H3K27me3 than for the marks H3K9me2 and H3K9ac. This could be due to the fact that the *Hoxa* cluster is transiently expressed during differentiation and the bivalent marker is closely linked to this process, perhaps so much as to be the defining epigenetic motif in order to drive expression.

4.4 HISTONE MODIFICATIONS IN DIFFERENTIATING CELLS

4.4.1 Dynamics of histone modifications through early embryonic differentiation

To fully investigate the distributions of histone modifications at *Hoxa*, I went on to repeat these N-ChIP experiments in differentiated cells for those modifications which showed enrichment at the *Hoxa* cluster, i.e. H3K4me3, H3K27me3 and H3K9ac. I used the time point of 5 days after LIF removal, because this was the point at which expression of *Hoxa* genes was highest, making it an ideal time to

analyse the behaviour of histone modifications during differentiation and their relation to gene expression.

4.4.2 Histone modifications at Gapdh, Nanog and Pou5f1 after differentiation

Figure 4.9 shows the changes in H3K4me3, H3K27me3 and H3K9ac between an undifferentiated and a d5 differentiated cell population. Methylation of H3K4me3 has increased slightly at *Gapdh* after differentiation, but has decreased markedly at *Nanog* and *Pou5f1*, reflecting the drop in transcription of these two pluripotency genes. There is also a simultaneous increase in H3K27me3 at *Nanog* and *Pou5f1* after differentiation, although the level of the repressive mark does not rise above the unbound fraction. H3K27me3 remains at a similar level at *Gapdh* at both d0 and d5. H3K9ac shows large changes between undifferentiated and differentiated cell populations, undergoing deacetylation from d0 to d5 as shown in **figure 4.9**. Although H3K9ac is enriched at all three loci, there has also been a change in how H3K9ac is transmitted through the cell cycle. At d0, acetylation at H3K9 decreases at G_2/M. At d5, however, the mark is stably enriched between both cell populations. *Nanog* has also become the least acetylated of the three genes, from being the most acetylated at d0. This is most likely to be a reflection of the drop in expression of *Nanog* and *Pou5f1* during the first five days of differentiation.

4.4.3 Bivalent histone modifications after differentiation at the Hoxa cluster

Figure 4.10 shows the results of N-ChIP experiments from ES cells and differentiated cells for both bivalent histone modifications H3K4me3 and

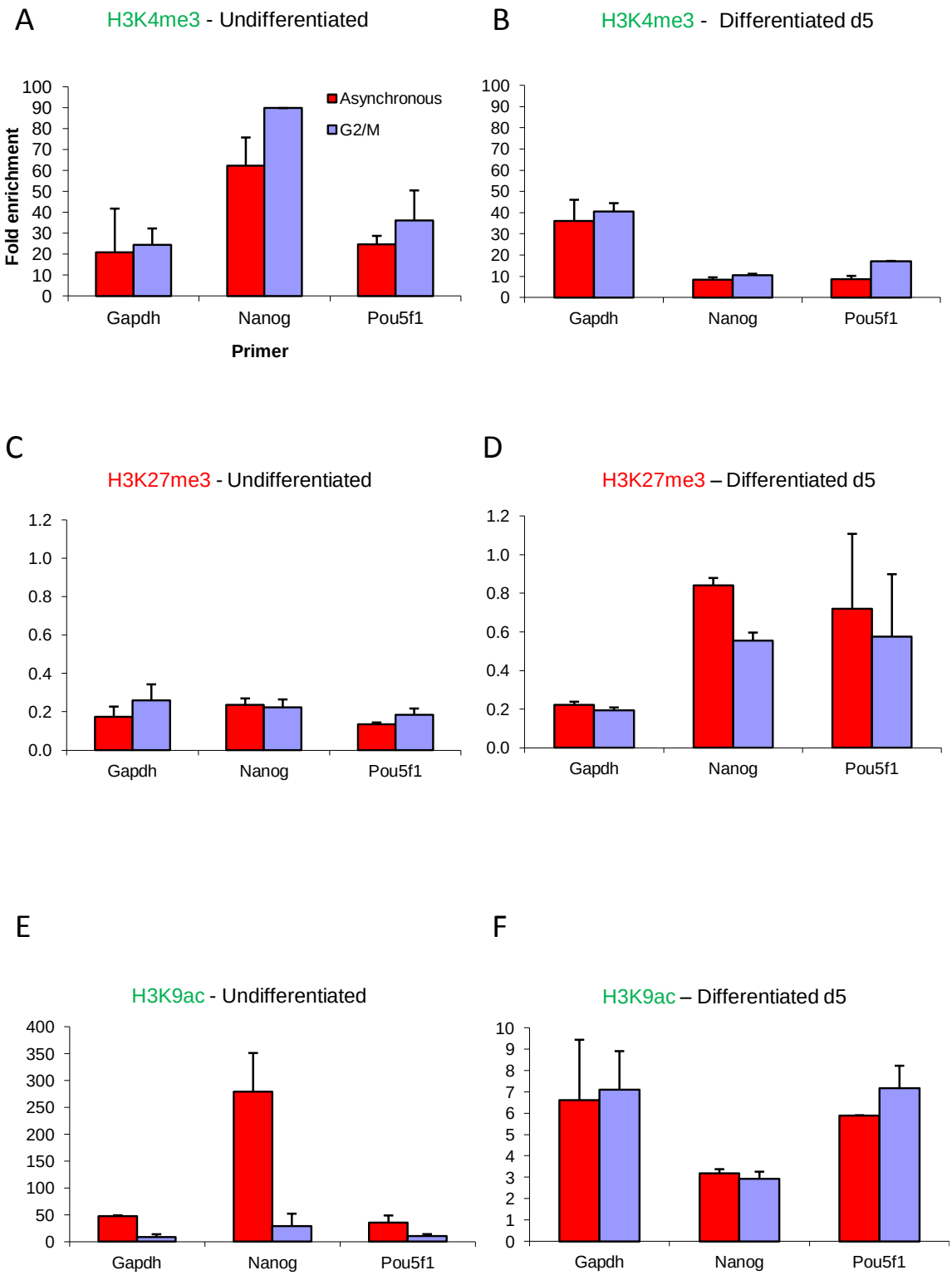
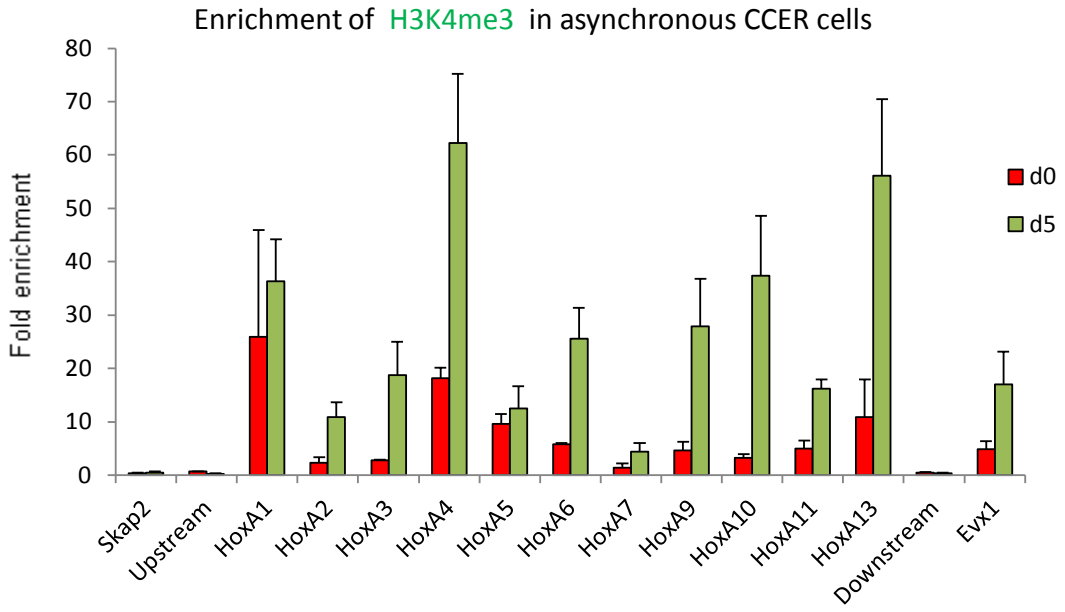


Figure 4.9 Enrichment of DNA extracted from ChIP pulldowns with **A+B.** H3K4me3 **C+D.** H3K27me3 and **E+F.** H3K9ac antibodies. Asynchronous denotes undifferentiated asynchronous CCE/R cells, G2/M indicates the same cells treated for 6 hours with 5µl/ml colcemid. Primers used are specific for the promoters of the metabolic gene Gapdh and the pluripotency-related genes Nanog and Pou5f1. Panels on the left display results obtained from an undifferentiated cell population, panels on the right represent data from cells sampled after 5 days of differentiation. Error bars show the standard error of the mean between three replicate experiments

A



B

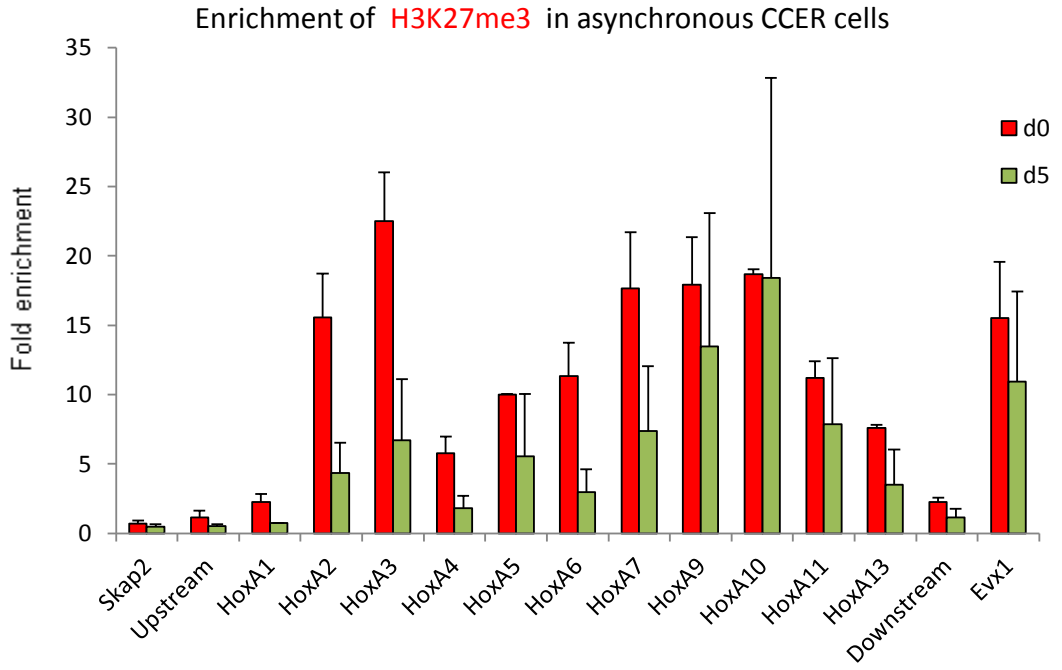


Figure 4.10 Real-time enrichment (Bound/Unbound) of DNA extracted from ChIP pulldowns. Data shown is from asynchronous CCE/R cells. **A** shows results obtained with H3K4me3 antibody from undifferentiated cells (red), and from cells differentiated for 5 days (green). **B** shows results obtained with H3K27me3 antibody from undifferentiated and d5 differentiated cells . Primers used are targeted to promoters for genes in and adjacent to the HoxA cluster on mouse chromosome 6. Primers are specific for the promoter of the labelled gene except for upstream and downstream primers which are complementary to intergenic sequences. Error bars show the standard error of the mean between three biological replicate experiments

H3K27me3 across the *Hoxa* cluster. Data for trimethyl K4 shows that methylation here has increased from day 0 to d5 at each gene promoter across the *Hoxa* cluster. **Figure 4.9,** however, shows that K4 methylation decreased at pluripotency genes in the same cell populations. Although H3K4me3 levels remain low at sequences upstream of the cluster, the *Hoxa* genes themselves show clear enrichment of K4me3 compared to the d0 samples. *Hoxa1* shows an increase in enrichment from d0 to d5, and an even larger increase in the G_2/M population, representing a doubling in the concentration of trimethyl-K4 at this locus at G_2/M. *Hoxa2*, one of the *Hox* genes with the lowest level of H3K4me3 at d0, shows a large increase in this mark after five days of differentiation. Methylation at K4 shows an increase between d0 and d5 populations at *Hoxa3* and similar increases can be found at every *Hoxa* gene in the study, as well as at *Evx1*.

The highest fold increases in H3K4me3 occur at genes that have low levels of enrichment in the undifferentiated cell population, the most drastic of these being the increase at *Hoxa10* from d0 to d5. This represents a shift in the epigenetic signature to a state that is strongly associated with promoting underlying gene transcription. The mechanism of bivalency suggests that one of the two marks becomes dominant after differentiation, negating the effect of the other. These data suggest that H3K4me3 is increasing in prevalence after 5 days of differentiation and thereby creating a more active chromatin environment, therefore fuelling the increase in transcription seen during the course of the expression study.

The increase in methylation seen within the *Hoxa* cluster is not indiscriminate, although *Hoxa1* and *Hoxa4* still show the highest enrichment of H3K4me3 in the cluster, methylation at other genes has risen by varying extents. *Hoxa5*, for

example, is highly enriched at d0, more so than for *Hoxa6*. However at d5, the pattern was reversed and *Hoxa6* displays higher levels of methyl K4. *Hoxa5* may represent the exception once again to the general distribution of H3K4me3 at the *Hoxa* cluster, as enrichment at this locus does not greatly increase upon differentiation, and in fact in the metaphase sample the H3K4me3 enrichment is decreased. Expression of *Hoxa5* and *Hoxa4* follows the same pattern throughout the 7 days of differentiation; however the histone modification profiles at these two genes are very different, perhaps due to the unique nature of *Hoxa5*. This tells us that complex epigenetic regulation processes are occurring that are very discretely and specifically targeted to sections of chromatin without affecting neighbouring sequences. The rise in methylation of H3K4 at *Hoxa4* is not spread to *Hoxa3* or *Hoxa5*, making the difference in epigenetic signatures between these genes even more pronounced. Therefore, it is likely that K4 chromatin modifying enzymes exert careful control over the epigenetic fate of these genes and regulate each individually rather than depositing methyl marks over a large area of chromatin in order to facilitate decondensation of chromatin. This would also explain the rise in K4 methylation at most genes in the *Hoxa* cluster, and the simultaneous decrease in H3K4me3 at *Skap2*, *Nanog* and *Pou5f1*.

Since the bivalent mark is linked to differentiation, and *Hoxa* is known to be activated during the course of cell development, one might expect the level of the repressive mark H3K27me3 to be reduced during this process (Kashyap *et al.*, 2011). Indeed, as shown in **Figure 4.10**, most *Hoxa* cluster genes undergo demethylation at lysine 27 after 5 days of differentiation. Again, there are exceptions, this time *Hoxa9* and 10, which show no marked change in H3K27me3 between undifferentiated and d5 cells. Unlike the pattern shown in H3K4me3, the

levels of trimethyl K27 relative to other genes in the cluster remain unchanged. The marks relatively high in K27me3 in d0 cells remain high after differentiation, but with a reduced enrichment over the unbound fraction. This trend is only bucked by the particularly high enrichment of K27me3 found at *Hoxa5* in G_2/M stalled cells, which is lost nearly entirely after differentiation to render the asynchronous and G_2/M populations equal within this sample. Again, this could indicate that *Hoxa*5 alone is not developmentally regulated by the bivalent chromatin signature and is controlled instead by other histone modifications, or other epigenetic regulators entirely. This could challenge the notion that the Hox cluster represents one regulatory region of chromatin and that all underlying genes are affected by the same epigenetic mechanisms.

This evidence points towards a reorganisation of chromatin during the process of differentiation with respect to the composition of the bivalent mark. Not only is the more repressive H3K27me3 mark reduced as differentiation progresses, but the more permissive histone modification H3K4me3 is up-regulated, with the main effect of increasing expression of *Hoxa* genes after the removal of LIF. It should be noted that although enrichment of H3K27me3 has been reduced after differentiation, it is still enriched above background levels at most *Hox* loci and significantly above levels at other genes such as the housekeepers *Gapdh* and *Skap2*, and pluripotency genes *Pou5f1* and *Nanog* (**figures 4.9 and 4.10**). These pluripotency genes have been shown to lose their associations with H3K4me3 after differentiation, however they do not show a concomitant increase in H3K27me3 deposition, as one might expect given results for *Hoxa* genes. The difference here is that the pluripotency genes are not thought to be associated with a bivalent histone modification mark, and so are not regulated in the same

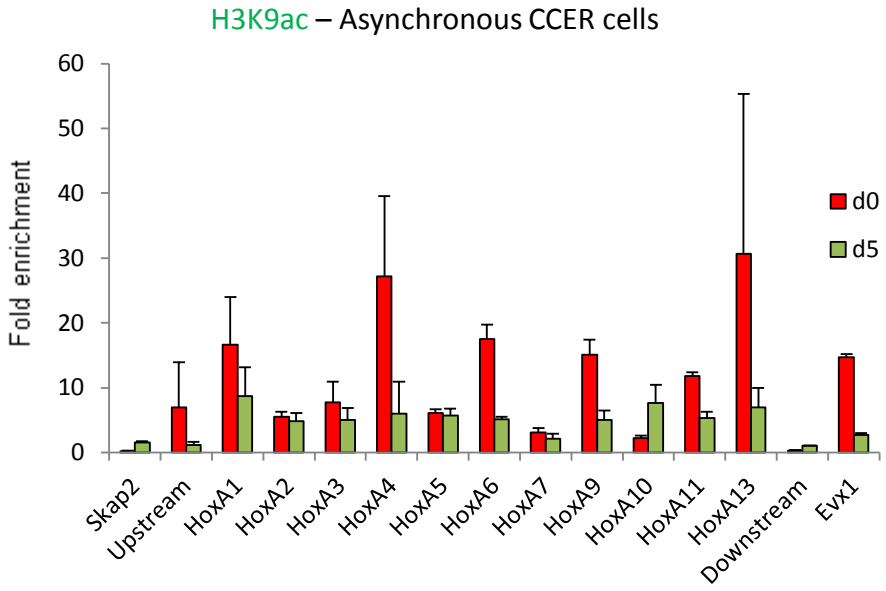
way as those genes in the *Hoxa* cluster. They do still appear to be epigenetically regulated, however, as the loss of H3K4me3 after differentiation points to control of transcription through epigenetic means. It may be that the expression of these genes are controlled simply through K4 methylation, or that some other repressive modification is playing the part of H3K27me3 in another bivalent motif, or that this process is controlled by a complex variety of factors which do not conform to any simple linear relationship.

4.4.4 H3K9ac enrichment through differentiation and the cell cycle

The epigenetic code could potentially include hundreds of separate chromatin modifications that combine to influence cell phenotype (Strahl and Allis, 2000, Turner, 2000). Keeping this in mind, I looked at another histone modification outside of the bivalent paradigm, and studying acetylation through differentiation rather than methylation. As previously discussed, acetylation is thought to have a more direct influence on chromatin structure by disrupting bonds within histone molecules to allow access to transcription factors and DNA binding enzymes (Allfrey *et al.*, 1964, Marushige, 1976, Grunstein, 1997). I therefore investigated histone 3 lysine 9 acetylation in day 5 differentiated cells as opposed to ES cells. **Figure 4.11** shows that the distribution of this mark changes drastically as differentiation progresses, and that the persistence of this mark through the cell cycle is also more stably maintained.

Results in the asynchronous population show that in the *Hoxa* cluster, H3K9ac is consistently depleted in d0 G_2/M cells compared to an asynchronous population. However, after differentiation, the difference between G_2/M and asynchronous cells is equalised, and H3K9ac is maintained at a constant level at each locus

A



B

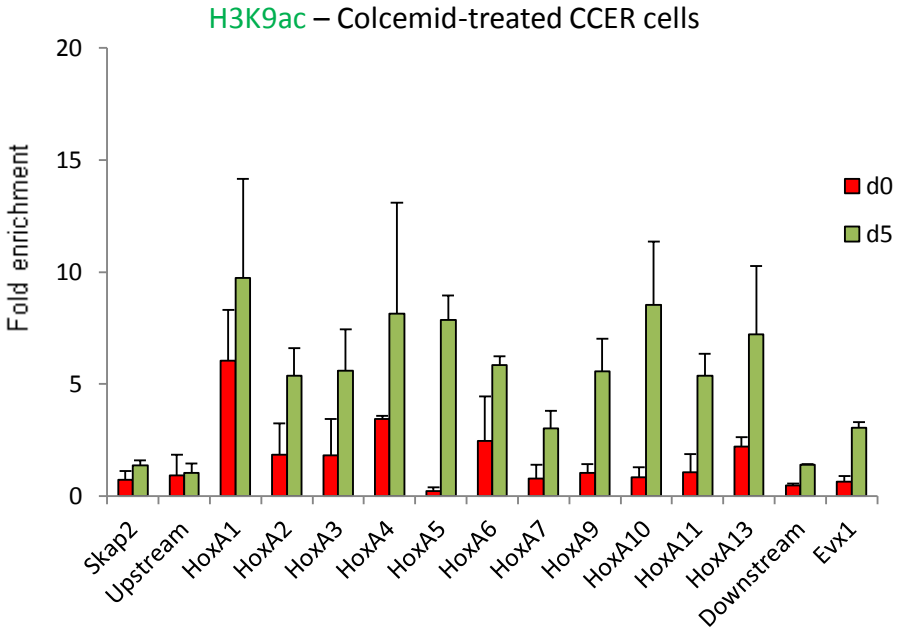


Figure 4.11 Real-time enrichment (Bound/Unbound) of DNA extracted from ChIP pulldowns. **A** shows results obtained with H3K9ac antibody from asynchronous CCER cells, **B** from CCER cells treated with colcemid (G_2/M). Red bars denote undifferentiated cells, green bars represent cells differentiated for 5 days in the absence of LIF. Primers used are in and adjacent to the HoxA cluster on mouse chromosome 6. Primers are specific for the promoter of the labelled gene except for upstream and downstream primers which are complementary to intergenic sequences. Error bars show the standard error of the mean between three biological replicate experiments

throughout the cell cycle. This is due not only to a reduction in the enrichment of K9ac in asynchronous cells, but also to an increase in acetylation in G_2/M-specific cells. *Hoxa1*, for example, showed an asynchronous enrichment at d0, dropping markedly in G_2/M in the undifferentiated sample. After differentiation, these values equalised at an intermediate level. *Hoxa2*, on the other hand, does not show a great depletion of H3K9ac from asynchronous d0 to d5 cells, but does show an increase in acetyl K9 in the G_2/M population. This pattern is also true for *Hoxa5* and *Hoxa7*. All other members of the *Hoxa* cluster display a trend of equalising, whereby acetylation is reduced in the asynchronous population, but increased for the G_2/M population. These results can be explained by the reconciliation of two distinct processes. Firstly, the process of differentiation appears to have negated the effect of the cell cycle on the maintenance of the H3K9ac mark. The reduction of K9ac in the G_2/M population in d0 cells can be attributed to the loss of acetylation during mitosis phase, although there is still widespread H3K9ac enrichment, as shown in **figure 4.7** and in chromosome pictures (**figure 3.5**). The process of differentiation profoundly affects the nature and timings of the cell cycle, and also has the effect of increasing the doubling time of affected cells. As a result, cells spend a lot more of their time being stalled in the gap phases rather than going through mitosis. Therefore, the loss of acetyl marks at metaphase would represent an even more transient event in the cell cycle of a differentiating cell, and would therefore not greatly affect the results obtained from a G_2/M population. This is accompanied by a simultaneous down-regulation of the H3K9ac modification, as evidenced by the reduction in acetylation on many of the *Hoxa* genes after 5 days of differentiation. The increase in H3K4me3 at the same

time may render the activating H3K9ac mark redundant, and so *Hoxa* chromatin becomes less acetylated as a result.

Considering these experiments as a whole, the data lends further weight to the theory that histone modifications can be predictive of gene expression, as seen by the results of the H3K4me3 ChIP experiment at the *Hoxa* cluster (**figure 4.6**). These data also show that the differentiation process alters the epigenetic landscape at key developmentally-regulated genes such as *Hoxa* to create a more permissive chromatin environment (**figure 4.10**), underlining the importance of dynamic epigenome control during key cellular processes.

5. ANALYSIS OF DATA IN THE CONTEXT OF PREVIOUS STUDIES

5.1 THE HOXA CLUSTER IN THIS AND PREVIOUS STUDIES

5.1.1 Hoxa in undifferentiated mouse ES cells

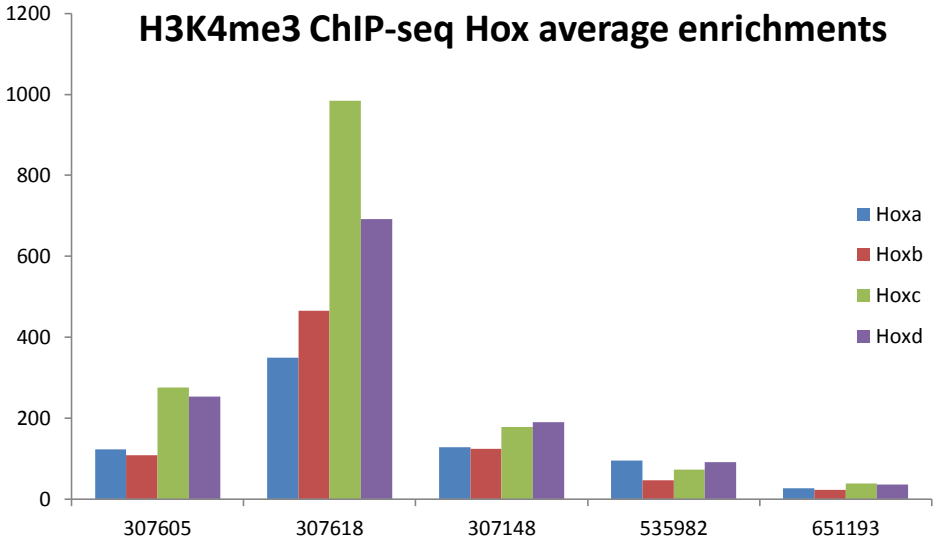
This chapter shows the data described in the previous chapters compared side by side with other similar studies involving embryonic cells. This comparison is between studies using undifferentiated murine embryonic stem cells, showing ChIP results against the H3K4me3 histone modification within the *Hoxa* cluster. The datasets shown in **table 5.1** were chosen as they study mouse embryonic cells, using ChIP-sequencing to obtain results for H3K4me3 and H3K27me3 enrichment across the mouse genome. This should allow a comparison to be made with my own data. All results have been normalised to the highest enrichment found in that data set, in order to easily compare different sets of data. **Figure 5.1** shows the variation between experimental datasets and why normalisation is necessary. The enrichment of H3K4me3 is highest in all literature studies at *Hoxa*1, as in the original data shown in black. There is good agreement among all studies at the *Hoxa*2 and *Hoxa*3 loci, which all show low enrichment of H3K4me3. The amount of variability between studies of this nature is evident at *Hoxa*4, as is shown in **figure 5.2**. My data shows a high enrichment of H3K4me3 at *Hoxa*4 of over half the *Hoxa*1 level, whereas the literature data shows an enrichment anywhere between 10% and 40% of the enrichment of H3K4me3 at *Hoxa*1. In one study (#307605), the level of H3K4me3 is fairly consistent between *Hoxa*2, a3 and a4, whereas in all the others, the enrichment of H3K4me3 increases between *Hoxa*3 and *Hoxa*4.

Results for *Hoxa*5 are also not consistent, with several studies showing low enrichment of H3K4me3, and others including mine showing a higher incidence.

Accession	Antibody	Cell line	Cell type	ChIP method	Read resolution
307605	H3K4me3 Abcam 8580	V6.5	ES cell undifferentiated	X-ChIP	25bp with 300bp window
307618	Abcam 8580	V6.5	ES cell undifferentiated	X-ChIP	25bp with 300bp window
307148	Abcam 8580	V6.5	ES cell undifferentiated	X-ChIP	25bp with 200bp window
535982	Abcam 8580	E14Tg2A	ES cell undifferentiated	X-ChIP	Unspecified: 250bp window
651193	H3K4me3 Millipore 07-473	E14	ES cell undifferentiated	X-ChIP	25bp with 300bp window
686992	H3K27me3 Upstate 07-449	129Sv-C57Bl/6	ES cell undifferentiated	X-ChIP	25bp with 200bp window
686993	Upstate 07-449	129Sv-C57Bl/6	ES cell undifferentiated	X-ChIP	25bp with 200bp window
686994	Upstate 07-449	129Sv-C57Bl/6	ES cell undifferentiated	X-ChIP	25bp with 200bp window
487549	Upstate 07-449	C6	ES cell undifferentiated	X-ChIP	25bp with 200bp window
307608	H3K4me3 Abcam 8580	129SvJae x C57BL/6	Mouse embryonic fibroblast MEF	X-ChIP	25bp with 300bp window
279934	Abcam 8580	MCV6	MEF	X-ChIP	25bp with 300bp window
279936	Abcam 8580	MCV6	MEF	X-ChIP	25bp with 300bp window
279938	Abcam 8580	MCV8	MEF	X-ChIP	25bp with 300bp window
279942	Abcam 8580	MCV8.1	MEF	X-ChIP	25bp with 300bp window
307609	H3K27me3 Upstate 07-449	129SvJae x C57BL/6	MEF	X-ChIP	25bp with 300bp window
397409	H3K27me3 in-house	LF2	ES cell d10 differentiated	X-ChIP	25bp with 300bp window
397411	H3K27me3 in-house	E14	ES cell d10 differentiated	X-ChIP	25bp with 300bp window
279937	Upstate 07-449	MCV6	MEF	X-ChIP	25bp with 300bp window
279939	Upstate 07-449	MCV8	MEF	X-ChIP	25bp with 300bp window

Table 5.1 Table showing the characteristics of datasets used in Chapter 5. GSM accession numbers are shown in the first column, and antibodies used, cell line and cell type, ChIP method and sequencing resolution are also shown.

A



B

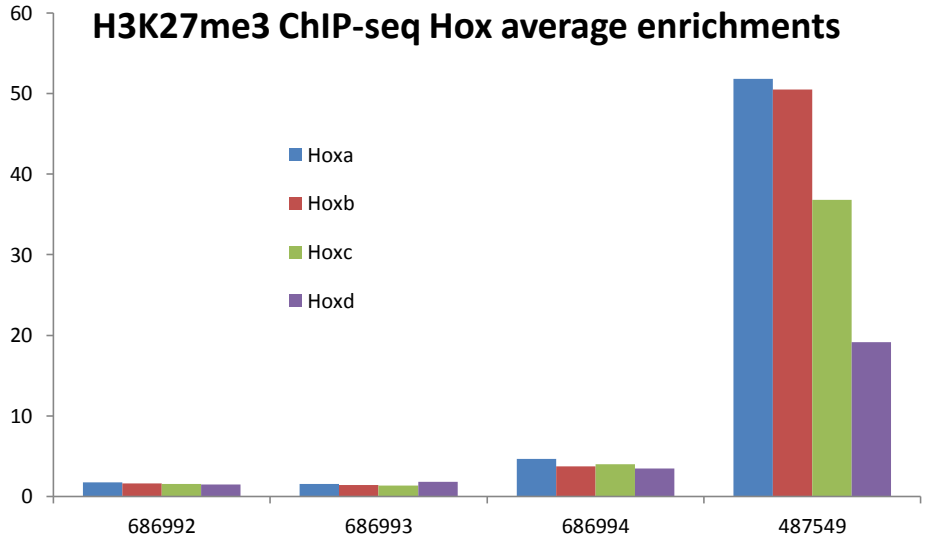


Figure 5.1 Graph to show the mean enrichments of **A** H3K4me3 and **B** H3K27me3 across the 4 Hox clusters in mice. Shown on the x-axis are the GSM accession numbers of the ChIP-seq datasets. The y-axis shows the level of enrichment of each modification.

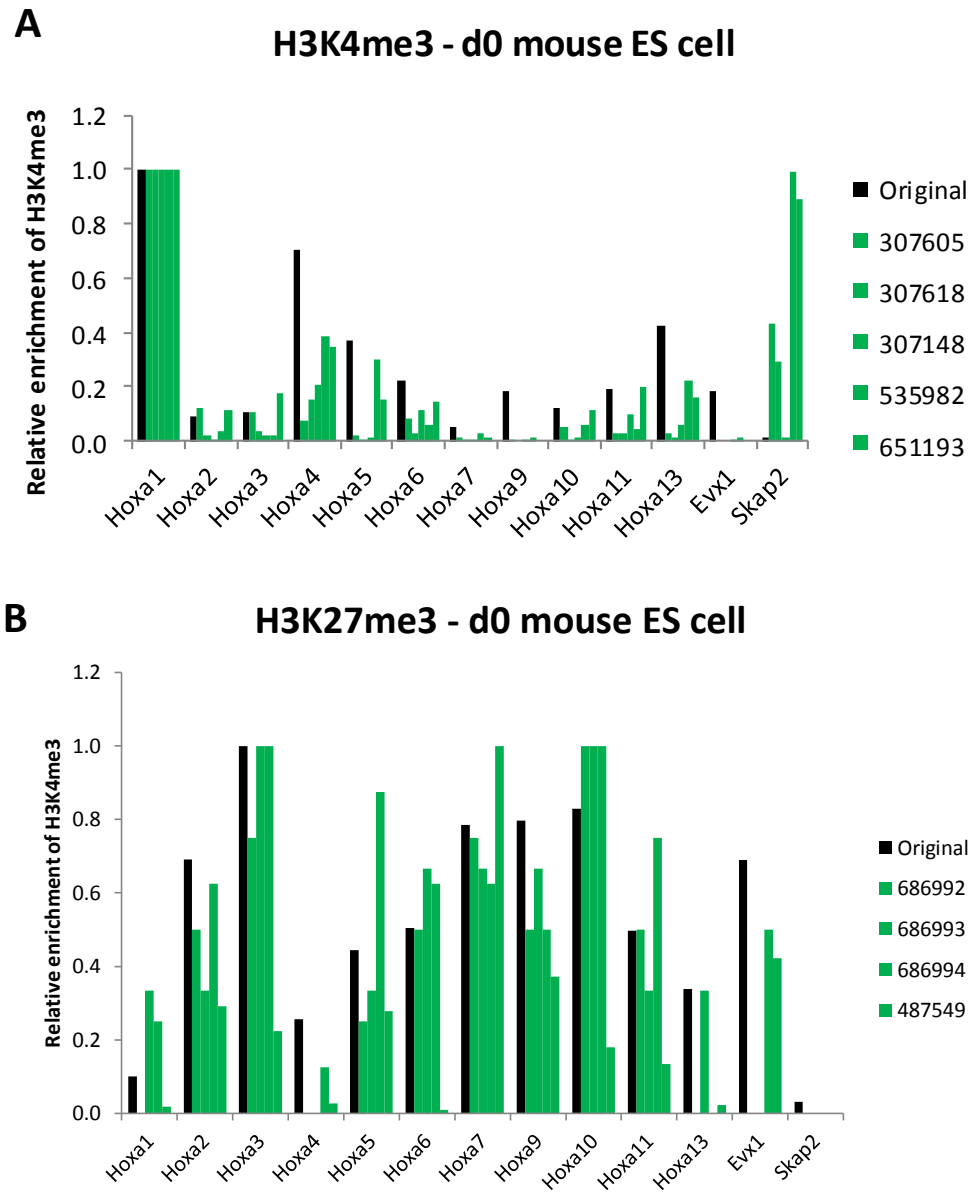


Figure 5.2 Graphs showing my original data compared with several datasets obtained from previous studies in the literature. Original data is shown in black, literature data is shown in green. NCBI GSM numbers are shown in the legend to denote the datasets. **A** shows the H3K4me3 distribution in undifferentiated mouse cells, **B** shows the H3K27me3 distribution; both histone modifications were analysed at Hoxa cluster gene promoters. All data were normalised to the maximum of that dataset, shown by a value of 1.0, so that comparisons may be made between data

Enrichment of H3K4me3 at *Hoxa*6 is less variable, with all results coming between 10-20% of the maximum. *Hoxa*7 results are also quite consistent, with all studies reporting a low enrichment of H3K4me3. *Hoxa*9 is where the greatest difference between my results and the literature results lies. My results show an enrichment of H3K4me3 of between 10 and 20% of the maximum enrichment, whereas the literature studies all show low enrichments of less than 5% of the maximum. *Hoxa*10 shows low enrichment for all studies, as does *Hoxa*11, with moderately good agreement between the data. *Hoxa*13 shows fairly high enrichment in my data, which is reflected in some studies, although enrichment is very low in others. Results for the two genes that surround the *Hoxa* cluster show large differences between my data and those in previous studies. There is an extremely low enrichment of H3K4me3 at Evx1 in all previous studies. However, in my study, the enrichment is higher, at around 15-20% of the maximum. Conversely, enrichment of H3K4me3 at Skap2 is particularly high in several literature datasets, some even approaching the same level of H3K4me3 as at *Hoxa*1, but is extremely low in my data.

As has been noted already in this study, the distribution of H3K27me3 across the *Hoxa* cluster generally opposes the distribution of H3K4me3. This is also reflected in **figure 5.2**, which shows a much lower relative enrichment of H3K27me3 at *Hoxa*1 than H3K4me3. My data is analogous to literature studies, which also show an enrichment of H3K27me3 at *Hoxa*1 at below 1/3 of the maximum enrichment across the cluster. The amount of H3K27me3 enrichment increases in *Hoxa*2 and then again to a maximum at *Hoxa*3 in my data. This agrees with the previous studies, two of which also have maxima at *Hoxa*3. This affirms the association of *Hoxa*2 and *Hoxa*3 with low H3K4me3 and high H3K27me3 levels. *Hoxa*4 shows

high H3K4me3 enrichment in my study, and relatively low enrichment in literature data. H3K27me3 is lower in all studies, as shown by the falls in enrichment levels in **figure 5.2**. H3K27me3 is maintained at a high level throughout the rest of the *Hoxa* cluster, in contrast to H3K4me3, which is generally low from *Hoxa*5 to *Hoxa*13 in literature datasets. The bordering genes Evx1 and Skap2 have opposing H3K4me3/H3K27me3 levels in the literature studies. For example, H3K27me3 is enriched at Evx1 but not at Skap2, whereas H3K4me3 is enriched at Skap2 but not at Evx1. This conflicts with my data, which shows a low level of H3K4me3 at Skap2, and not such a great difference between H3K4me3 and H3K27me3 at Evx1. Such a clear difference between datasets is somewhat surprising, given that these two genes are not part of the *Hoxa* cluster proper, and also that Skap2 in particular is a housekeeping gene and should not be subject to a great deal of expression regulation since it is required to be 'always-on'. Interestingly, the literature data agrees with mine is so far as the H3K27me3 is enriched at negligible levels in d0 cells.

5.1.2 Hoxa in differentiated mouse ES cells

Figure 5.3 shows the distribution of H3K4me3 across the *Hoxa* cluster in differentiated cells. It is immediately clear that there is more variation between literature datasets, and also my own data, than there is for ESCs. For example, GSM307608 shows a similar enrichment of H3K4me3 at *Hoxa*1 to my study, and a very low level at *Hoxa*2, well below the value shown for my data. GSM279934, however, has low enrichment of H3K4me3 at *Hoxa*1, and a high enrichment at *Hoxa*2; the highest anywhere in the *Hoxa* cluster for that dataset. My data shows the highest enrichment of H3K4me3 differentiated cells occurring at the *Hoxa*4

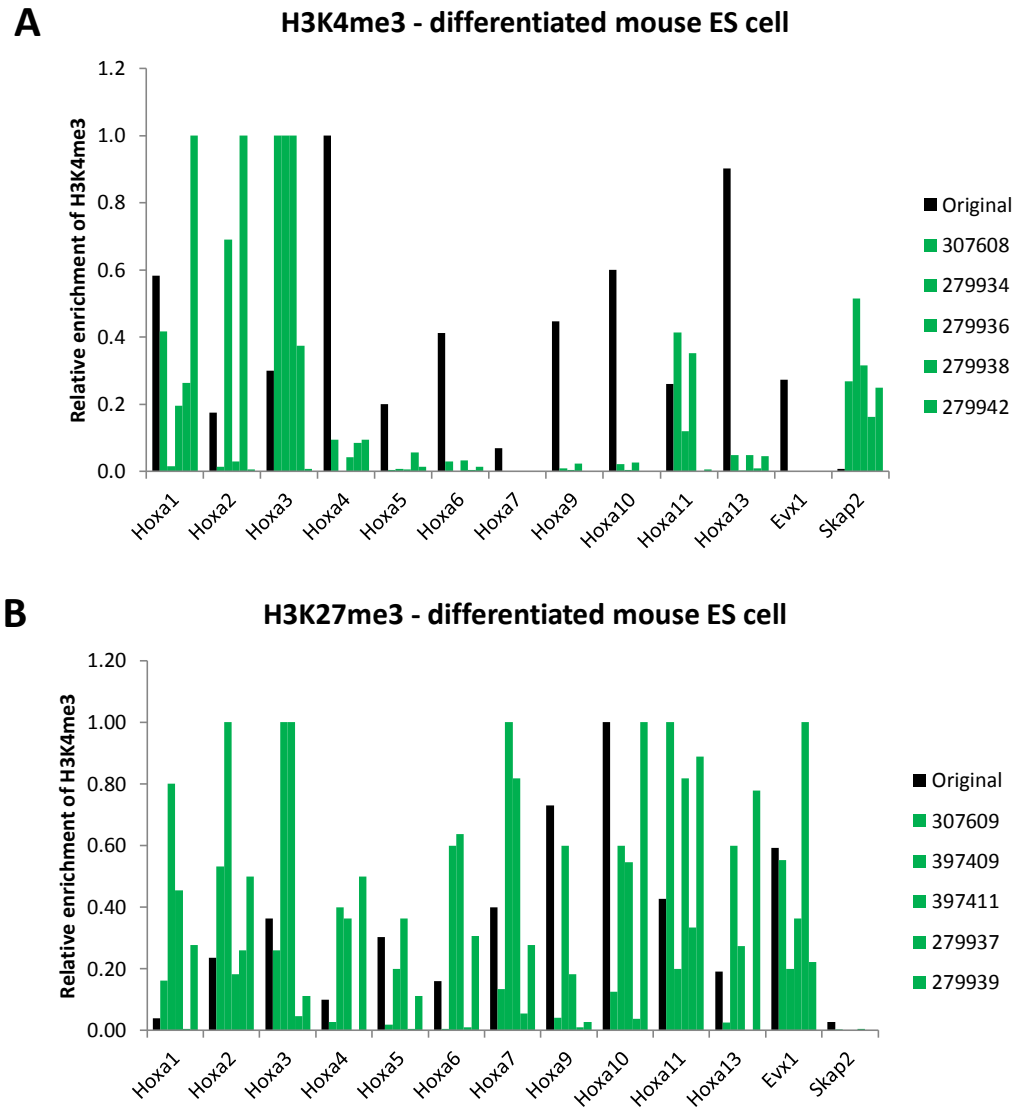


Figure 5.3 Graphs showing my original data compared with several datasets obtained from previous studies in the literature. **A** shows the H3K4me3 distribution in differentiated mouse cells, **B** shows the H3K27me3 distribution; both histone modifications were analysed at Hoxa cluster gene promoters. All data were normalised to the maximum of that dataset, shown by a value of 1.0, so that comparisons may be made between data

promoter. However, the literature data consistently show a low level of enrichment at this locus. This is again the case for *Hoxa*5-13, with the exception of *Hoxa*11, which shows a high level of enrichment in several studies. However, my data show varying levels of H3K4me3 enrichment, but always at a higher relative level than those shown in the literature studies. Results for Skap2 and Evx1 are similar to the undifferentiated data, and once again the literature data disagrees with the results of my study.

Literature data for the distribution of H3K27me3 across the *Hoxa* cluster in differentiated cells is widely variable. The five different samples shown in **figure 5.3** show differences at many *Hoxa* loci. Firstly at *Hoxa*1, each dataset is different to the next; some showing high H3K27me3 enrichment, others showing moderate enrichment, and one displaying very low enrichment. My data shows low levels of H3K27me3 in d5 differentiated cells, agreeing with one study but not others. At *Hoxa*2 a similar picture can be seen, whereby my study agrees with two literature samples in denoting modest enrichment of H3K27me3 at this locus, but not other studies. This pattern continues throughout the *Hoxa* cluster, whereby literature datasets do not agree, with the exception of Skap2 H3K27me3 enrichment, which is consistently very low in all literature datasets and my own data. The amount of variation between datasets has been represented in **figures 5.4 and 5.5** by error bars denoting standard deviation between studies. The error bars show that the variation between studies is greater when looking at differentiated cells rather than undifferentiated embryonic stem cells. This is most likely due to the fact that the embryonic stem cell is a very well-defined stage in the development of the cell, and has been standardised to some degree between laboratories. Studies employing ChIP-sequencing to look at histone modifications in more differentiated

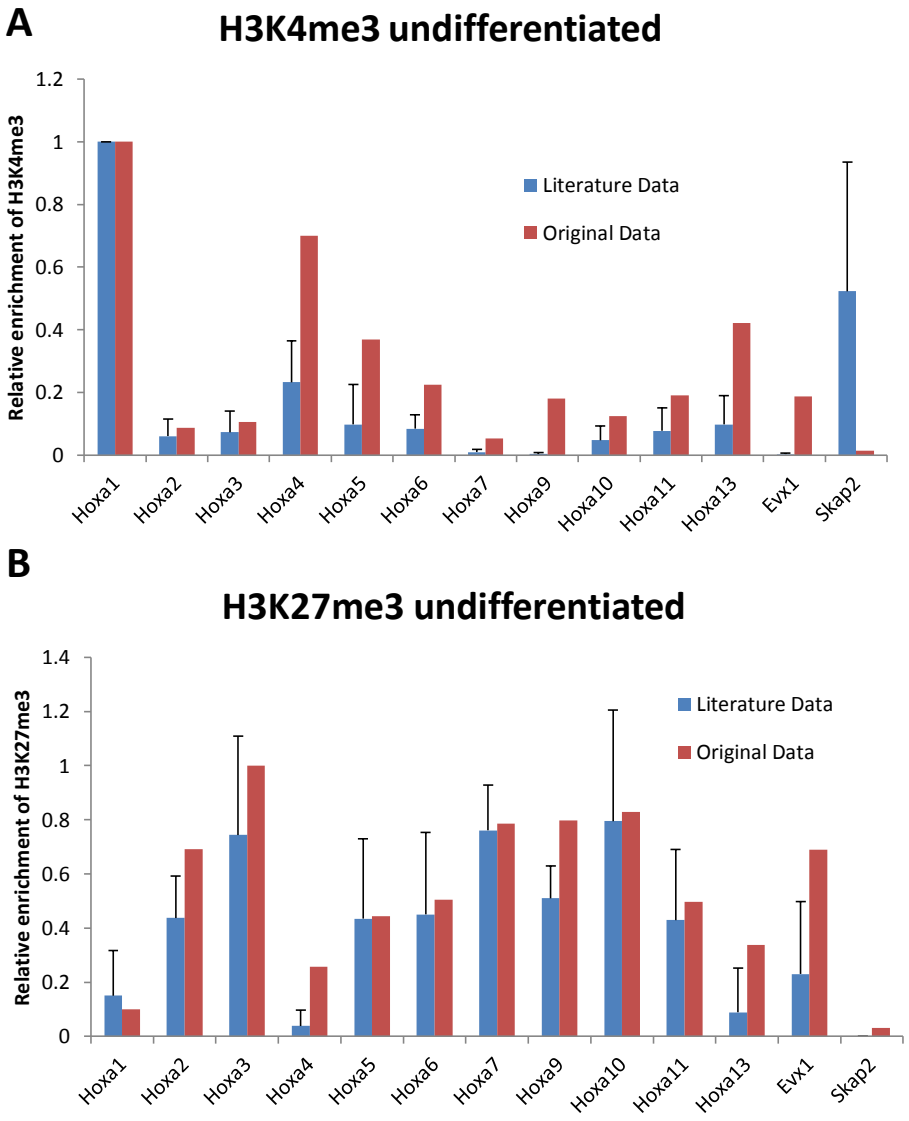


Figure 5.4 Re-interpretation of figure 5.2, showing the literature data mean as a single bar, and error bars to denote standard deviation of the results in order to represent the variation between studies. **A** shows enrichment of H3K4me3 across the Hoxa cluster in undifferentiated mouse cells. **B** shows enrichment of H3K27me3 at the Hoxa cluster, also in undifferentiated cells.

embryonic cells may differ widely in their culture methods and differentiation paths of the studied cultures. There is also greater variation among H3K27me3 d0 results when compared to H3K4me3 d0; however my results follow the literature pattern of H3K27me3 methylation closely, indicating that the methylation pattern found in this study is reflective of previous datasets obtained in other laboratories.

The correlation between my data and literature data is not as strong for d0 H3K4me3 as for d0 H3K27me3, and there are several clear differences in **figure 5.4** between the mean literature values and my results. However, the patterns of methylation of H3K4me3 appear similar. For example, although there is a difference at *Hoxa*4 between datasets in **figure 5.4**, it is still the second-most enriched site in the *Hoxa* cluster after *Hoxa*1 in literature studies, whereas my data shows *Hoxa*1 and *Hoxa*4 being more equal in H3K4me3 enrichment, both above any other *Hoxa* gene promoter. In fact, the differences between these datasets would be reduced considerably by the omission of the *Hoxa*1 bar. *Hoxa*1 has a very high relative enrichment to the other 10 *Hoxa* genes in the literature data. The normalisation of the data to the highest signal has caused the enrichment at other loci to appear low by comparison. The reason why *Hoxa*1 is relatively higher in the literature data is not clear, although it may be due to the use of differing cell lines, or possibly the use of ES cells that have been split through many more passages, causing more cells to begin differentiating and upregulating expression of the Hox genes.

5.1.3 Comparison between literature studies and original study

These analyses have shown that there is a strong basis to support the observation that my original data is supported by previous datasets for undifferentiated ES

cells. Data for H3K27me3 are particularly similar between literature and original studies, with only minor differences. Data for H3K4me3 is also correlated between original and literature sets, although there are more clear differences than for H3K27me3, the pattern of H3K4me3 methylation across the *Hoxa* cluster is broadly the same (**figure 5.4**). This, as has previously been mentioned, indicates a pre-establishment of histone modification in anticipation of the *Hoxa* cluster being transcribed after differentiation.

The data shown here for differentiated cells shows a large amount of variation between both the literature datasets themselves, and also my original dataset (**figure 5.5**). There is little clear correlation between previous studies and my own, especially for H3K4me3 data. The difficulty in analysing these datasets together comes from the differences between populations of differentiated cells. My study used d5 differentiated embryonic stem cells from the CCE mouse lineage. As no ChIP-seq studies have been done using this exact cell line looking at H3K4me3 and H3K27me3, studies using other types of differentiated embryonic cells have been analysed. Clearly, this can include a variety of cell lines, and even methods of differentiation and so therefore it is not surprising that a wide amount of variation has been observed. However, it remains the case that an inverse relationship can be found at most *Hoxa* loci between H3K4me3 and H3K27me3 enrichment. The highest H3K27me3 enrichment in **figure 5.5** can be found at *Hoxa*2, a3, a7 and a10 in the literature data, and at *Hoxa*3, a7, a9, a10 and a11 in my original data. All of these genes show low H3K4me3 enrichment in undifferentiated cells, although literature data shows that *Hoxa*2 and a3 are enriched with H3K4me3 in differentiated cells, and my data indicates a high level of H3K4me3 at *Hoxa*9 and a10. Referring back to **figure 4.10**, it is apparent that

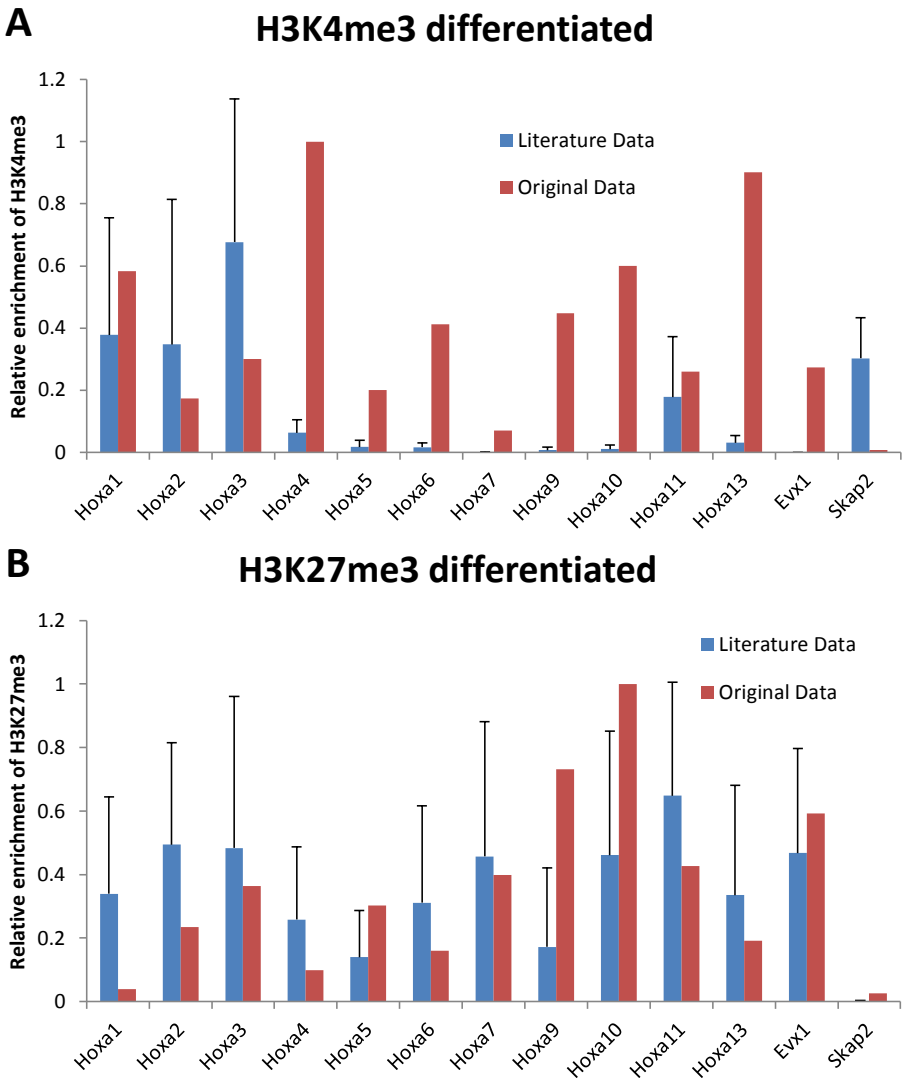


Figure 5.5 Re-interpretation of figure 5.3, showing the literature data mean as a single bar, and error bars to denote standard deviation of the results in order to represent the variation between studies. **A** shows enrichment of H3K4me3 across the Hoxa cluster in undifferentiated mouse cells. **B** shows enrichment of H3K27me3 at the Hoxa cluster, also in undifferentiated cells.

the absolute enrichment of H3K27me3 has decreased after differentiation at most *Hoxa* loci, but not at *Hoxa*9 and a10, whereas H3K4me3 has greatly increased at the same loci. This is not the case in the literature data shown in **figure 5.5**, which shows that H3K4me3 at *Hoxa*9 and a10 remains at negligible levels. The means of the literature datasets indicate that rather than an inverse correlation between H3K4me3 and H3K27me3 being present (as it is at d0) in differentiated cells, there are concomitant high levels of both marks at *Hoxa*1, a2, a3 and a11. However, the variation between studies as shown by the large error bars in **figure 5.5** suggests that inter-sample disagreement may be skewing the results.

Figure 5.6 shows a direct comparison between H3K4me3 and H3K27me3 levels at each *Hoxa* locus, both in my data and in literature studies. Presenting the data in this way, it is clear that the relationship between these two marks is similar in both datasets, and that my data (for d0 cells at least) corroborates earlier studies. The clear predominance of H3K4me3 over H3K27me3 at *Hoxa*1 and *Hoxa*4 is visible for original and literature data, as is the reverse relationship at *Hoxa*2, a3, a6, a7, a9, a10 and a11. At *Hoxa*13, levels of these two marks are broadly equivalent in my data, and also in literature data. There is a difference between the two sets of data at *Hoxa*5, where my study shows equivalence between the relative levels of H3K4me3 and H3K27me3, whereas the literature data shows a higher level of H3K27me3. However, the error between literature experiments shows that the overall levels of H3K4me3 and H3K27me3 could be more similar than the graphs suggest. As described earlier, the *Hoxa* border gene Skap2 has a clear enrichment of H3K4me3 and negligible amounts of H3K27me3 in the literature studies that is not reflected in my own data, which shows low enrichment of both marks. This difference could be caused by the use of two

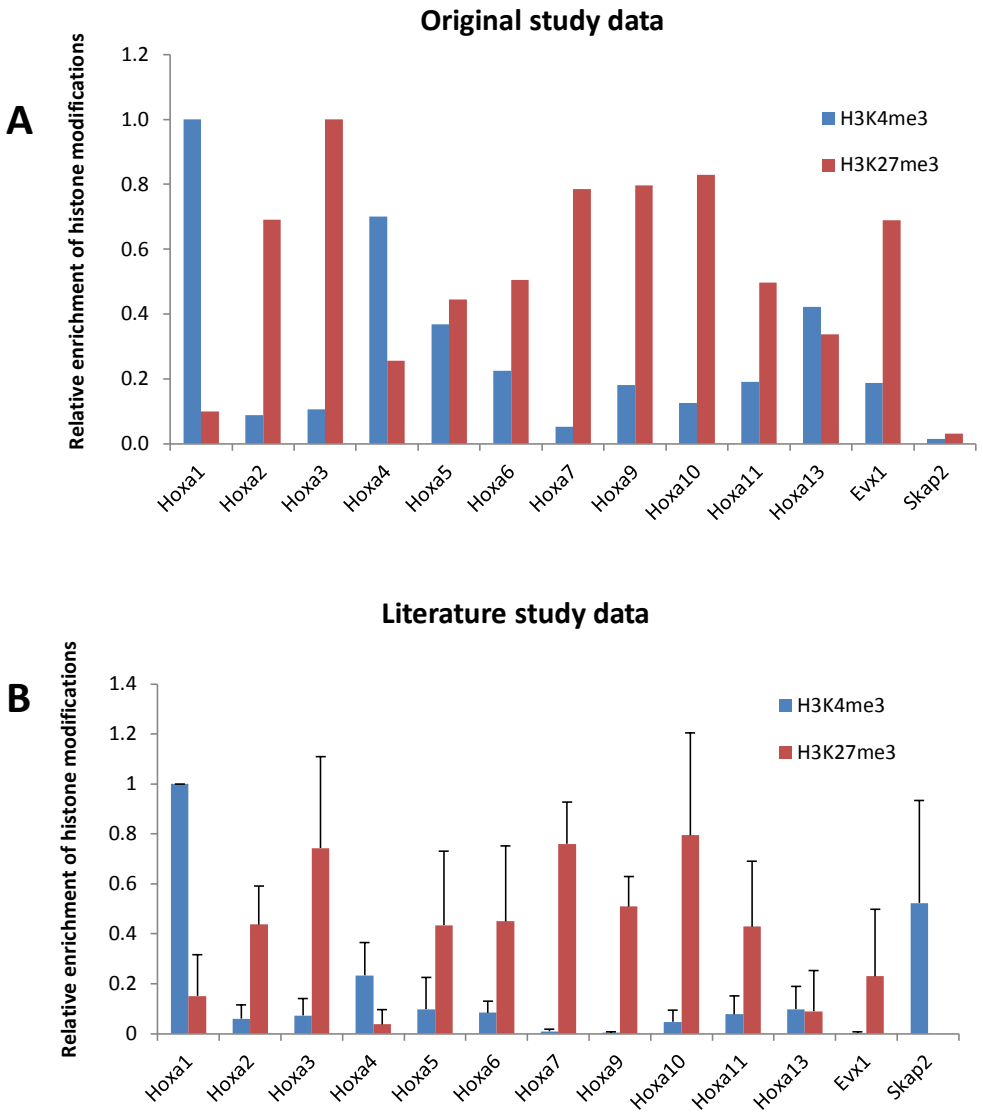


Figure 5.6 Graphs to show the relationship between H3K4me3 (blue) and H3K27me3 (red) in d0 undifferentiated mouse ES cells. **A** shows enrichment of H3K4me3 and H3K27me3 in studies described in this thesis. **B** shows enrichment of H3K4me3 and H3K27me3 in studies performed in other laboratories, where data is available publicly. Error bars show the standard deviation between studies.

different techniques, ChIP-sequencing for the literature data, and targeted real-time PCR for my data.

The close similarity between my data and previous studies for the enrichment of H3K4me3 and H3K27me3 in undifferentiated mouse embryonic stem cells confirms that the bivalent model for the control of expression of developmentally regulated loci is too simplistic. These data point towards a more sophisticated mechanism of epigenetic control that carefully regulates the expression of these genes, and also marks chromatin based on the future expression of the underlying gene sequences. As shown earlier, the expression of *Hoxa* genes once induced using retinoic acid is not a linear process. *Hoxa*1 and *Hoxa*4 are two genes that are quickly and strongly upregulated after differentiation cues have been introduced. This is reflected in the chromatin landscape for the *Hoxa* cluster at d0 even before cells have started to differentiate, showing the predominance of permissive chromatin marks (H3K4me3) over restrictive ones (H3K27me3). This implies that these histone modifications are marking particular genes, so that further down the differentiation line, the transcription machinery can recognise, with the assistance of chromatin-binding proteins, which genes are to be activated first.

5.2 APPLYING CHIP-SEQ ANALYSIS TO OTHER GENES AND CLUSTERS

5.2.1 Analysing other Hox clusters using literature data

As shown above, hypotheses drawn from differentiated datasets are likely to be unreliable; however, the data for undifferentiated cells are similar enough to support a hypothesis that an inverse relationship exists between H3K4me3 and H3K27me3 at the *Hoxa* cluster and perhaps at other developmentally-regulated

sites as well. This hypothesis can be tested by applying this observation to genomic areas other than the *Hoxa* cluster, such as at other Hox clusters like Hoxb, c and d.

Figure 5.7 shows the results obtained from the same literature datasets as above, this time looking at the three other Hox clusters. These clusters of Hox genes are similar to *Hoxa* in that they determine rostral-caudal axis patterning in developing embryos. From the ChIP-seq data displayed here, it appears that the Hoxd1 gene has been marked for early expression, due to the relative dominance of the H3K4me3 signal over H3K27me3. This appears similar to the *Hoxa*1 histone modification profile seen in **figure 5.6**. However, Hoxb1 does not display this motif, showing a relatively low H3K4me3 enrichment. The enrichment of H3K4me3 over H3K27me3 is also less pronounced for the first Hoxc gene in the cluster, Hoxc4. There are genes in each cluster that show a greatly enriched H3K4me3 residue accompanied by a low level of H3K27me3 trimethylation, although they are at different positions within the cluster. Hoxb7, and to a lesser extent Hoxb9, show such a pattern. Hoxc13 is also heavily methylated at H3K4me3 relative to the rest of the cluster, although its H3K27me3 remains low. These observations could indicate that the four Hox clusters are regulated by the bivalent signature differently or perhaps during separate stages of the organism's development.

The three Hox clusters shown in **figure 5.7** show very different profiles of H3K4me3 and H3K27me3 enrichment. Hoxb shows an accumulation of H3K27me3 mainly in the middle of the cluster, at Hoxb3, b5 and b6, whereas H3K4me3 is enriched at Hoxb7 and b9. The Hoxc cluster shows a more moderate level of enrichment of both H3K4me3 and H3K27me3, as in many places both

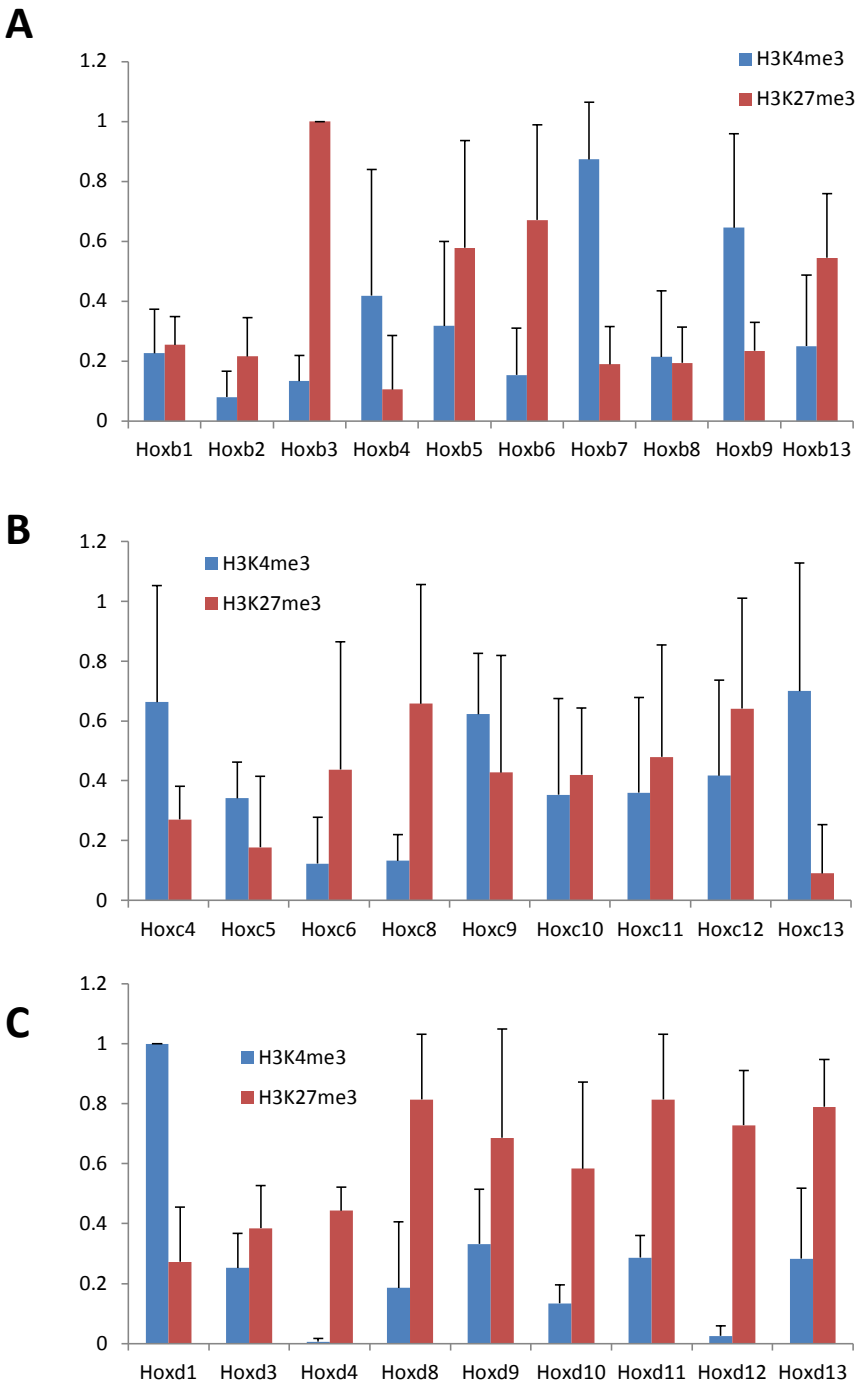


Figure 5.7 Graph to show the relationship between H3K4me3 (blue) and H3K27me3 (red) in d0 undifferentiated mouse ES cells across **A** the Hoxb cluster, **B** the Hoxc cluster and **C** the Hoxd cluster. Above shows enrichment of H3K4me3 and H3K27me3 in literature studies performed in other laboratories, where data is available publicly. Error bars show the standard deviation between studies. Fluorescence relative to the maximum observed fluorescence is shown on the Y-axis.

modifications are enriched to roughly the same degree. There is an excess of the H3K4me3 mark on Hoxc13, and an excess of H3K27me3 on Hoxc6 and c8 however, which could denote targeted epigenetic marking.

Hoxd genes appear to be more generally marked with H3K27me3, with most genes showing a relatively high level of this mark, and a low level of H3K4me3. This is not the case at Hoxd1, which shows preferential H3K4me3 enrichment, or at Hoxd3, which shows relative equivalence between the two marks. This may indicate that, unlike Hoxb and Hoxc clusters, Hoxd may be regulated in a similar way to *Hoxa*. Expression data from other studies confirm that Hoxd1 and Hoxd3 are expressed earlier and more strongly than other genes in the cluster after retinoic acid activation (Zha *et al.*, 2012). Histone modifications may have a part to play in this process by foreshadowing the increase in expression by increasing the ratio of H3K4me3 to H3K27me3 at Hoxd loci. As differentiation, or retinoic acid activation of Hox, continues, the more distal Hoxd genes will become higher in H3K4me3 methylation and become epigenetically poised for expression. This same process could occur for Hoxb and Hoxc, but the characteristic epigenetic signature may become apparent later in the differentiation process than for *Hoxa* and Hoxd. The presence of the bivalent marks at all Hox clusters imply that this mechanism could be used to regulate the expression of these genes, although this has not yet been proven.

5.2.2 Analysis of other developmental genes using ChIP-seq

To explore the role of the two bivalent histone modifications H3K4me3 and H3K27me3, I have analysed the public ChIP-seq data for other genes not related to Hox clusters, namely *Pou5f1* and *Nanog*, two key regulators of pluripotency, as

well as *Shh* and *Foxd4*, which are genes that have important functions during embryonic development (later than the ES cell stage). **Figure 5.8** shows results from my study for *Pou5f1* and *Nanog*, as well as literature study data for these two genes as well as *Shh* and *Foxd4*. It is clearly shown that both the pluripotency genes are rich in H3K4me3 methylation at their promoters, whereas H3K27me3 is almost entirely absent from these loci, in both literature and my original datasets. As d0 ES cells require these genes to be expressed to maintain pluripotence, a strongly permissive chromatin environment is required.

Shh and *Foxd4* are developmental genes that are required during embryo development, but have functions different to those of Hox genes (Huangfu and Anderson, 2006; Moody, *et al.,* 2013). Results from **figure 5.8** show that both of these genes display a high level of H3K27me3 enrichment, coupled with a low frequency of H3K4me3. This confirms a repressive chromatin environment at d0, consistent with its lack of expression. However, these genes do not display the bivalent characteristic of some of the *Hoxa* and Hoxd genes, which are markers for future expression of those genes at later developmental stages. Bivalency is clearly a specialised characteristic of some developmental loci, which is used to carefully control gene expression in developing vertebrates. A simple on/off switch, such as is required to down regulate pluripotency genes after differentiation, does not appear to involve both bivalent histone modifications as a precursor to cessation of transcription in these genes.

The data in the literature shown above have shown a good level of agreement with the results I have obtained myself (**figure 5.4**), and therefore some confidence can be applied to the literature data shown for other genes and gene clusters. Considering all the data analysed within this chapter, it seems that the

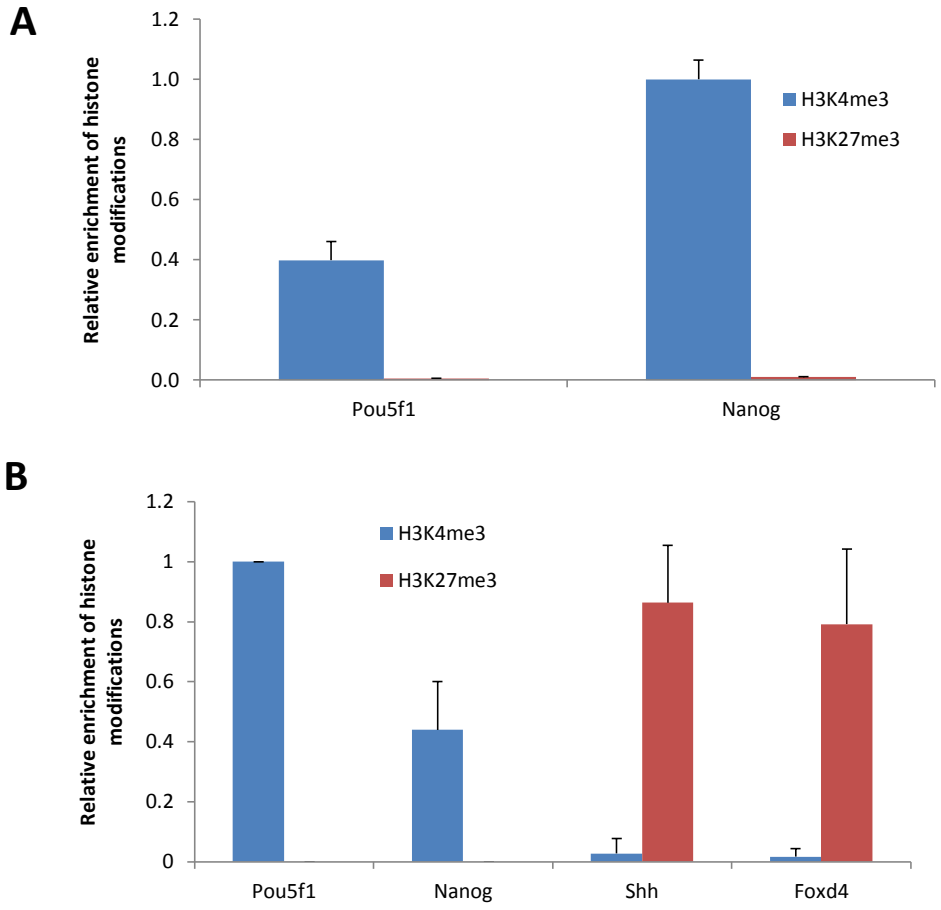


Figure 5.8 Graph to show the relationship between H3K4me3 (blue) and H3K27me3 (red) in d0 undifferentiated mouse ES cells. **A** shows enrichment of H3K4me3 and H3K27me3 in data from chapter 4 of this study. **B** shows results from literature studies performed in other laboratories, where data is available publicly. Error bars show the standard deviation between studies.

bivalent mark is present specifically at developmentally regulated sites that are yet to be expressed in embryonic cells. The bivalent mark does not occur at the heavily transcribed Pou5f1 and Nanog loci, key pluripotency markers, but does occur at all four Hox gene clusters. However, there is little overlap between H3K4me3 and H3K27me3 at Shh and Foxd1, other loci that are up regulated after differentiation. This could indicate that bivalency is specific to individual sites and are not the exclusive mechanism by which differentiation-triggered transcriptional control is exerted. Certainly, the bivalent profiles across the four Hox clusters are very different, and it is evident that each cluster has its own unique bivalent signature (**figure 5.7**). This could reflect the differences between the clusters with regard to the timing of expression during the developmental process, as well as the physiological necessity for the levels of the protein products of those genes to be translated in the developing embryo. The epigenetic profile of the Hox clusters, however, are clearly distinguishable from the state of bivalent marks H3K4me3 and H3K27me3 at other loci such as Shh, Foxd1, Pou5f1 and Nanog. These loci show either one or the other mark in abundance, but not both together. This observation supports the association between a bivalent motif and the regulation of differentiation.

5.3 COMPARISON OF CHIP-SEQ AND IMMUNOFLUORESCENCE

Having demonstrated that ChIP data in this study are comparable to ChIP-sequencing data in other literature studies, I have also compared the chapter 3 immunofluorescence data to these datasets. However, unlike the data in chapter 4, which are mostly derived from asynchronous cell populations, immunofluorescence data are all taken from metaphase-specific cell populations.

As ChIP-seq experiments are performed on asynchronous cells in these datasets, by comparing these to my own data, it is analogous to comparing metaphase cells with asynchronous cells with a tiny proportion of cells in metaphase. Other notable differences between these experiments are that immunofluorescence is performed on unfixed chromatin, whereas all ChIP-seq data here use formaldehyde-fixed and sonically sheared DNA-histone fragments, eventually sequencing 200-300 base pair fragments fragments using sliding-window analysis to produce 25bp resolution maps. This technique has the potential to introduce bias into the experiment, as it has been previously shown that using X-chip can cause artificially strong signals for interactions in regions of low nucleosome density when using ChIP-seq (Schwartz, *et al.,* 2005). However, **figure 5.9** shows the enrichment of H3K4me3 across four different mouse chromosomes, both before and after ES cell differentiation. The patterns seen in these graphs closely resemble the karyotypes seen in **figures 3.11-3.14** and the histone modification genome maps in **figures 3.15-3.17**. For example, the region at c. 50Mb on chromosome 5 shows low H3K4me3 enrichment in the ChIP-seq data; this corresponds with the fluorescence data, which also shows a lack of H3K4me3 enrichment as evidenced by the lack of fluorescence in this region. Similarly, chromosomes 7 and 16 also have regions of low H3K4me3 density that correspond to low fluorescence in the immunofluorescent karyotype, at c. 70 megabases from the centromere on both chromosomes. However, the graph for chromosome 1 illustrates the difference between data obtained by ChIP-seq and that from immunofluorescent techniques. Data from **figure 3.11-3.12** suggests that chromosome 1 does not have any large chromosomal areas of intense H3K4me3 density. The analysed graphs shown in **figures 3.15-3.17** show a low level of

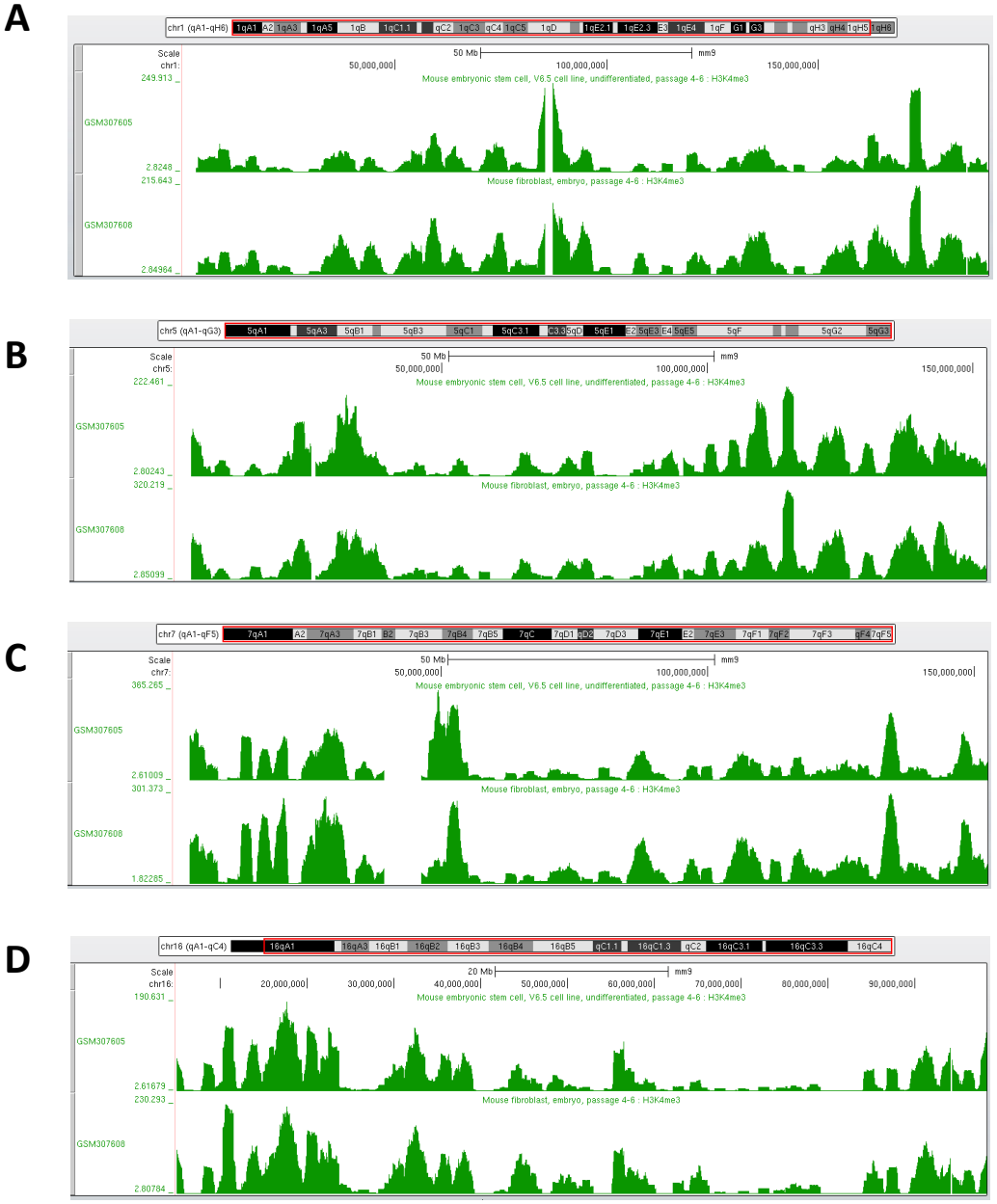


Figure 5.9 UCSC browser data showing ChIP-seq results for ES cells (upper – accession GSM307605) and embryonic fibroblasts (lower – accession GSM307608). The data bars indicate the level of enrichment of H3K4me3 throughout **A -** mouse chromosome 1, **B** – chromosome 5, **C** – chromosome 7, **D** – chromosome 16. Chromosome ideograms and scale bar is shown above

fluctuation of fluorescence, with an increase in fluorescence at the chromosome 1 telomere end. **Figure 5.9** shows the same general picture, with H3K4me3 enrichment fluctuating to a small extent from the mean, with an increase in density at the telomere end. However, these fluctuations are more well-defined than those shown in **figure 3.15**, and may correspond to functional areas of the genome. These comparisons are again obfuscated by the differences in techniques used, as ChIP-seq measures genetic distances between epigenetic features in base pairs, whereas the immunofluorescence technique measures physical chromosomal distances, which often does not reflect the distance in base pairs. This may cause some heavily fluorescent areas on chromosomes to appear in a different genetic location on the epigenomic map, either because the physical chromosome molecule has been compressed or relaxed at various regions within it, causing a distortion in the results.

Also shown in **figure 5.9** are graphs showing the H3K4me3 distributions across chromosomes in differentiated embryonic fibroblasts. On the whole, it appears that there are no major differences in the H3K4me3 distributions on this scale. However, a closer inspection of the data reveals that there are subtle differences in the enrichment of H3K4me3 at certain loci. For example, at 50 megabases from the chromosome 7 centromere, there is a double peak of strong H3K4me3 enrichment in the ES cell chart. However, after differentiation into fibroblasts, this has changed to a single peak. The leftmost peak has reduced H3K4me3 methylation after differentiation at this locus. This shows a specific response and a targeted change in epigenetic state caused by the progress of differentiation.

6. DISCUSSION

6.1 THE MOUSE EPIGENOME

6.1.1 Histone modifications and their maintenance in metaphase

One of the major current focuses in the field of molecular biology is the question of what the precise role of epigenetic marks is, and how they are maintained. Many theories have been suggested as to why the varied array of epigenetic modifications have arisen, but a definitive model for the establishment, maintenance and effect of these has not so far been achieved. These questions have potentially critical importance in our understanding of how eukaryotes organise not only their gene expression patterns, but also their diversely differentiated cell types. Much work has been done on the relationship between histone modifications and gene expression, and the effects and functions of many of these modifications are known (Rea *et al.*, 2000, Schotta *et al.*, 2004, McManus *et al.*, 2006, Guenther *et al.*, 2007, Pauler *et al.*, 2009, Wen *et al.*, 2009, Pasini *et al.*, 2010, Pekowska *et al.*, 2011, Zhang *et al.*, 2012). Some important studies have also inquired into the fundamental nature of the marks themselves. Where are they located (Bernstein *et al.*, 2006, Guenther *et al.*, 2007, Puschendorf *et al.*, 2008, Marks *et al.*, 2009)? How are they maintained (Breiling *et al.*, 2004, Francis, 2009, Francis *et al.*, 2009, Vermilyea *et al.*, 2009)? How are they re-established after replication (Groth *et al.*, 2007, Scharf *et al.*, 2009)? And how do they interact with each other (Forneris *et al.*, 2005, Zippo *et al.*, 2009, Kowenz-Leutz *et al.*, 2010)?

The first part of this study shows that during metaphase, histone modifications maintain individual and discrete distributions across the epigenome, and that some of these distributions change upon early embryonic differentiation (**figures 3.1 - 3.5**). These distributions show that many histone modifications are persistent enough to be maintained even during the dense packaging of chromatin found at metaphase. It has also been shown that the chromosome painting method can lead to a robust and illuminating epigenomic karyotype, showing a single-cell at-once epigenome (**figures 3.7 and 3.11**).

6.1.2 Investigating the epigenome of Mus musculus

With the completion of the Human Genome Project in 2003, the field of molecular biology can be thought of as being in the post-genomic era (Venter *et al.*, 2001). Whole-genome sequencing is now a relatively routine affair, and the focus of research is now more to do with what those DNA sequences mean, what they code for, and how they are expressed *in vivo*. The latter question has a lot to do with the positioning and interactions of epigenetic modifications in eukaryotes. Histone modifications have been linked with influencing transcription since the 1960s (Allfrey *et al.*, 1964). However, the evidence of how they function and interact with the genome is not as advanced as the knowledge we have around DNA and the features of the genetic code. This is understandable, as there are known to be many different histone modifications, unlike the four bases found in DNA, and more and more are being discovered (Bannister and Kouzarides, 2011, Arnaudo and Garcia, 2013). There are several initiatives currently proceeding with the aim of closing that gap in knowledge between genome and epigenome, and to produce a map of epigenetic signatures such as histone modifications similar to

that of the genome map (Birney *et al.*, 2007). This will have an enormous and profound impact on the field of biology, as not only the DNA sequence of a region will be known, but also the chromatin environment in which it lies, as well as other features such as potential chromatin-binding protein sites, higher-order chromatin structures, and permissiveness towards transcription could be determined. In order to further this end, this study has undertaken to examine the distributions of several different histone modifications in undifferentiated and differentiated mouse cells.

6.2 GLOBAL DISTRIBUTIONS OF HISTONE MODIFICATIONS AT METAPHASE

6.2.1 Distribution and inheritance of H3K4me1 and H3K4me2 at metaphase

The results shown in chapter 3 describe the distributions of histone modifications on histones H3 and H4 across mouse metaphase chromosomes. Maintenance of histone marks is a process that has not yet been fully elucidated or explained. It is evident; however, that although a general reduction in H3K4me1 can be seen, H3K4me1 is maintained throughout differentiation at many loci during metaphase from ES cells to MEFs (**figure 3.1**). This mark is thought to be generally permissive of transcription, with a particular affinity to enhancer motifs (Heintzman *et al.*, 2007, Heintzman *et al.*, 2009). The low metabolic activity and reduced accessibility of enzymes to chromatin characteristic of metaphase could have implied that maintenance of histone modifications may be compromised. However, persistence of histone modifications at metaphase has already been demonstrated in humans, and it is shown to be true here in mice as well (Terrenoire *et al.*, 2010). The intransigence of the H3K4me1 mark at some loci

through differentiation could mean that gene enhancers, which correlate with increased H3K4me1, maintain their chromatin signature (**figure 3.1**). This is supported by a recent study showing that H3K4me1 occupies the same proportion of the genome both in ES cells and ectodermal cells, although the authors do suggest that H3K4me1 is enriched at specific loci after differentiation, which could explain the apparent depletion of H3K4me1 in other genomic areas, relative to the increase in H3K4me1 at enhancer-rich sequences (Gifford *et al.*, 2013). The supposedly permissive environment conferred by H3K4 methylation could be conserved at enhancers throughout the developmental process, leaving control of the on-off switch of transcription to chromatin motifs closer to the TSS, such as H3K4me3 (Guenther *et al.*, 2007, Lloret-Llinares *et al.*, 2012). This theory is given added weight by the results showing a reduction in H3K4me2 prevalence in MEF cells at metaphase (**figure 3.1**). H3K4me2 is more directly correlated with transcription activity than H3K4me1, and in this context seems more likely to be influenced by the changing expression profile of the cell population (Heintzman *et al.*, 2007, Zhang *et al.*, 2012).

The distributions for both marks H3K4me1 and me2 show that centromeres and the Y chromosome are nominally free of these modifications (**figure 3.1**). The association between actively transcribed chromatin and methylated H3K4 would suggest that these marks would not be widespread in this region (Pekowska *et al.*, 2011). These results show that the centromeres and Y chromosome are entirely free of methylated H3K4, confirming that heterochromatin has a unique epigenetic makeup, and that it is not affected by histone methylation at H3K4 (**figures 3.1 and 3.2**). One of the striking features of these pictures for mono and dimethyl H3K4 is the high contrast between stained and unstained areas, producing

discrete bands of FITC fluorescence. This indicates that these marks are targeted to specific regions of chromatin, and are not maintained in other areas, perhaps due to a low level of underlying transcriptional activity.

6.2.2 Distribution of H3K4me3 through differentiation

H3K4me3 has for a long time been considered a marker for active transcription (Kouzarides, 2007, Karlic *et al.*, 2010). This modification is widespread in the promoters of genes, and marks developmental loci, with H3K27me3, as a bivalent epigenetic signature (Bernstein *et al.*, 2006, Sachs *et al.*, 2013). Images shown here reveal the genome-wide reduction in H3K4me3 upon differentiation of OS25 cells (**figure 3.2**). Some of the strongly fluorescent bands seen in undifferentiated OS25s are retained after differentiation. However, for the majority of bands, fluorescence is reduced and bands appear smaller and thinner. This could indicate that a large-scale demethylation of H3K4me3 has occurred, whereby the cell responds to the stimuli involved in the differentiation process by switching off, or de-poising, large sections of chromatin through demethylation of H3K4me3. The sections that remain fluorescent, or rich in H3K4me3, may be the genes specific to the neural lineage, which this method of differentiation favours (Bami *et al.*, 2011). This is supported by data in **figure 4.10**, which shows a specific increase in H3K4me3 at *Hox* loci after differentiation, but not at intergenic sequences. This could lead, with more knowledge about the distribution of these modifications, to identification of genomic regions based on epigenetic signature. For example, the telomeric end of chromosome 4 contains several genes related to neural development, and so it is possible that the strong H3K4me3 band that can be seen in **figure 3.2** on the chromosome just below the Y chromosome is

indicative of chromosome 4 (NCBI, 2013). Indeed, **figure 3.7** confirms that the telomeric end of chromosome 4 (this time clearly identified by FISH) is indeed labelled with H3K4me3 in undifferentiated cells. However, this pattern of widespread loss of H3K4me3 is not evident for the cultured MEF cells. The inherent differences between MEF cells and differentiated OS25 ES cells may provide an explanation for this finding. MEF cells are embryonic fibroblast cells that can be maintained in culture, and so have a stable phenotype. OS25 cells in this study have been differentiating for seven days, and are undergoing rapid changes in gene expression and cell identity (Glover *et al.*, 2006). This entails major shifts in the epigenetic landscape. A mechanism can be imagined whereby a widespread demethylation event occurs at non-developmental and intergenic loci during development, causing pluripotency genes to be down-regulated and for cells to quickly differentiate (**figure 3.2**). A more targeted and specific epigenetic identity is established by the altering of histone modifications and other epigenetic marks (**figure 4.10**). This could be evident here, whereby the MEF cells have already gone through the loss of pluripotency and have regained a stable epigenetic environment. This may even resemble the pattern it has lost for H3K4me3, but there will be crucial differences in other epigenetic modifications that identify the cell as differentiated.

6.2.3 Inheritance of repressive chromatin marks through differentiation

The other side of the bivalent coin is the repressive histone modification H3K27me3. This is known to be closely associated with H3K4me3 at developmental loci in undifferentiated cells (Bernstein *et al.*, 2006, Sachs *et al.*, 2013). The photos shown here corroborate this finding, as H3K27me3 is present

across the majority of the genome, therefore intersecting with many of the bands already observed for H3K4me3 (**figure 3.2**). This modification is stable throughout development up to and including the MEF stage. Taken together with the data for H3K4me3, it seems more likely that H3K27me3 is more heritable than H3K4me3 for the majority of the genome, and that H3K4me3 is lost (Kaneda *et al.*, 2011). Whether H3K4me3 is lost simply in the intergenic or non-coding regions cannot be determined from immunofluorescence, but these results do point to the preferential genome-wide preservation of some marks such as H3K27me3 over H3K4me2/me3 through the process of differentiation (**figures 3.1 and 3.2**). At specific developmental loci, H3K27me3 decreases after differentiation and H3K4me3 increases (**Figure 6.1**), but it seems here that for the majority of the genome, this is not the case (Bernstein *et al.*, 2006). Although the two marks are components of the bivalent mark, they are not solely associated with that function. H3K27me3 acts independently of H3K4me3 to form broad areas of inactive chromatin, as well as interacting with other modifications to regulate chromatin in other mechanisms (Pauler *et al.*, 2009, Schwartz *et al.*, 2010). This could explain how H3K27me3 is more widespread throughout the genome than H3K4me3.

Histone methylation on H3K9 is also maintained through differentiation (**fig 3.3**). Both H3K4me1 and –me2 retain their genome-wide distributions from ES cells through to differentiating and differentiated cells, displaying a general level of heritability throughout many cell cycles. There does however seem to be a redistribution of H3K9me3 between d0 and d7 OS25s (**figure 3.4**). This mark is abundant at centromeric chromatin, but by d7 the chromosome arms have become denuded of H3K9me3. There are proven links between H3K9me3 and centric DNA, however the ES cell epigenetic state shown in these pictures

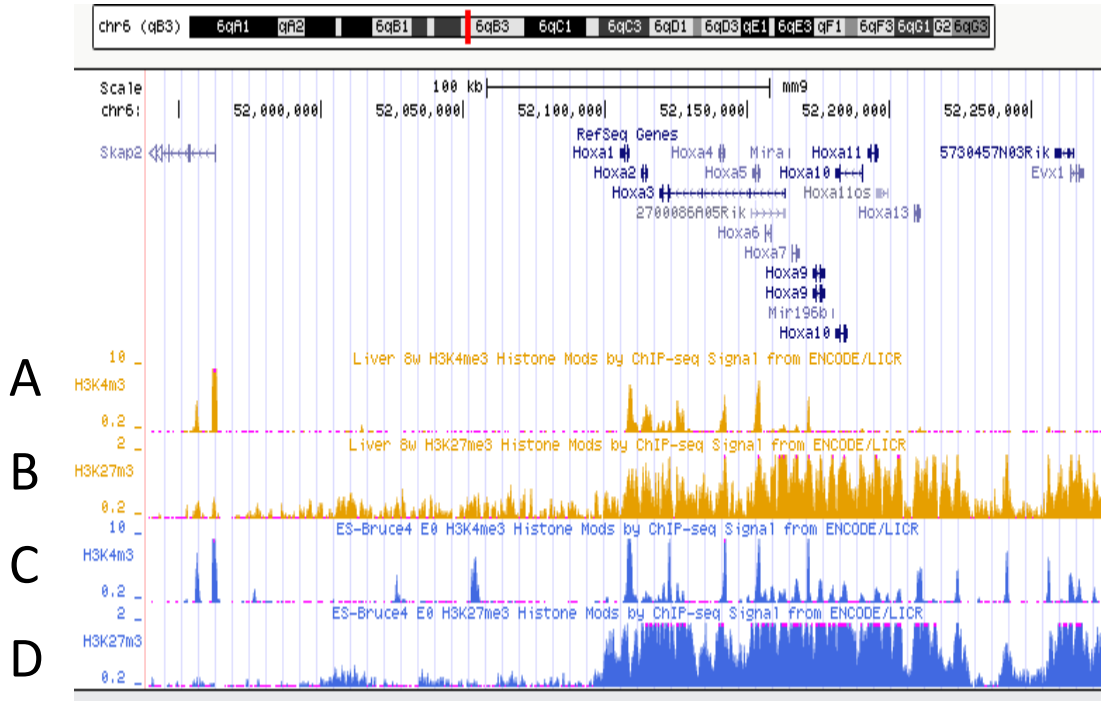


Figure 6.1 ChIP-seq data from ENCODE and UCSC mouse genome browser. H3K4me3 and H3K27me3 signals derived from ChIP experiments on undifferentiated ES cells (blue) and differentiated liver cells (yellow).

A. H3K4me3 ChIP-seq enrichment at Hoxa cluster in liver cells
B. H3K27me3 ChIP-seq enrichment at Hoxa cluster in liver cells
C. H3K4me3 ChIP-seq enrichment at Hoxa cluster in inner mass (ES) cells
D. H3K27me3 ChIP-seq enrichment at Hoxa cluster in inner mass (ES) cells

suggests that there is an additional widespread role for H3K9me3 in undifferentiated cells (Puschendorf *et al.*, 2008, Alder *et al.*, 2010). In a similar fashion to H3K4me3, the MEF photo in **figure 3.4** shows some H3K9me3 present on chromosome arms in contrast to the image for d7 OS25 cells. Again, the quickly changing phenotype of the differentiating cells may provide an explanation for this (Glover *et al.*, 2006).

6.2.4 Shift in epigenomic distribution of H4K20me3 after differentiation

The most evident differences between undifferentiated and differentiated cell histone modification distributions can be seen for the H4K20me3 mark (**figure 3.4**). This has often been linked to heterochromatic and centromeric DNA in past studies; however, this association is shown here to be dependent on the developmental stage of the cell. At the ES cell stage it is clear that centromeres are free of H4K20me3. The chromosome arms display areas of high and low FITC fluorescence to indicate that this modification is distributed preferentially in certain sections, but these sections are present across the entire karyotype. On differentiation, even after only seven days, this distribution is totally inverted so that chromosome arms are now free of H4K20me3, but centromeres are now labelled with this modification (**figure 3.4**). A reorganisation of chromatin in this manner could indicate that H4K20me3 has a role in the maintenance of pluripotency, as the widespread distribution of this mark is evident in ES cells but not in differentiated cells. The compartmentalisation of H4K20me3 into centromeres after differentiation may represent a mechanism for marking the changes in gene expression and phenotype of the cell after loss of pluripotency.

Interestingly, the Y chromosome is clearly labelled on undifferentiated spreads for H4K20me3, but indistinguishable (i.e. unlabelled) from autosomes in differentiated samples (**figure 3.4**). This shows in essence the same pattern as for the other chromosomes, in that the distribution of H4K20me3 is reversed after differentiation. However, the H4K20me3 mark is present on the mainly heterochromatic Y chromosome in ES cells, but not within heterochromatic centromeres. After differentiation the opposite is true, signifying that the trimethyl H4K20 mark can be used to distinguish centric from Y chromosome heterochromatin in ES cells and confirming the observation that centric chromatin has a unique epigenetic signature that is not shared by other, similar, heterochromatic regions (Haldar *et al.*, 2011, Honda *et al.*, 2012).

6.2.5 Distribution and transmission of acetylated histone residues through differentiation

Acetylated histones are associated with promoting transcription of underlying DNA (Allfrey *et al.*, 1964, Marushige, 1976, Agalioti *et al.*, 2002). These modifications are widespread in ES cells in **figure 3.5**, showing a uniform distribution across the chromosome arms in metaphase spreads from the ES cell stage through to d7. This shows strong maintenance of all these acetyl marks through differentiation, even when other methyl marks have been lost (**figures 3.1 and 3.5**). Images for MEF histone acetylation however, show a decrease in the abundance of all acetyl marks. The fact that all acetyl marks are lost at the same time, even acetyl marks on different histone molecules, indicates that a general mechanism of deacetylation is occurring at the majority of genomic loci. The acetylation signature of ES cells may have been altered by the MEF stage, although it is still possible

that acetylation remains at key areas of the genome, most likely at gene-rich regions, and is lost at non-coding sections of chromatin and at pluripotency-related genes or other genes no longer required by the cell. One theory resulting from these images could be that ES cells contain acetylated chromatin at transcriptionally inactive sequences, whereas by the time cells have differentiated into MEFs, the chromatin signature at these silent sites has become more deacetylated and therefore restrictive, although expression is not affected (**figures 4.3 and 4.11**). Certainly, this data suggests that there are several acetylated residues that undergo a large scale redistribution event at some point between day 7 of differentiation from ES cells and the differentiated MEF cell stage.

6.3 ALIGNMENT OF THE GENOME AND EPIGENOME

6.3.1 Linking the epigenome to the genome for H3K4me3 and H3K27me3

The metaphase chromosome immunofluorescence data already discussed has a lot to say about the distributions of histone modifications at this discrete portion of the cell cycle in various embryonic cell types. In order to put this data into a proper context, I compared these results to already available resources derived from publicly available sequence data (NCBI, 2013). This requires a link to be established between immunofluorescence and DNA sequence, which is provided by FISH identification of chromosomes already labelled with histone modification specific antibodies (**figure 3.6**). The epigenome karyotypes for H3K4me3 and H3K27me3 have been presented here, in a similar fashion to that already published for human chromosomes (Terrenoire *et al.*, 2010). The characteristic banding pattern seen previously for H3K4me3 is again evident, and can be used

to construct ideograms based not on G banding patterns, but on immunofluorescence patterns (**figure 3.7**). This creates clear distinctions between areas on chromosomes, and suggests that these modifications are specific and directed towards specific genetic features, which are themselves concentrated preferentially on some areas of the chromosome over others, creating bands of high H3K4me3 content (**figures 3.2 and 3.7**).

The implications of these observations can be explored by aligning the data with NCBI sequence data for mouse chromosomes (**figures 3.9 and 3.10**). These were organised into 10Mb windows in order to reflect the resolution of the chromosome immunofluorescence procedure. The strong correlation between H3K4me3 and gene density, CpG islands and repeats show that there is compelling evidence to support the theory that H3K4me3 is preferentially targeted to genes, and is less abundant in areas of low gene density. This supports previous research that found strong links between gene promoters and H3K4me3 deposition (Kouzarides, 2007, Karlic *et al.*, 2010, Tian *et al.*, 2011).

The distribution of H3K27me3 at first glance seems to be more widespread across the genome than that for H3K4me3 (**figure 3.2**). However, after analysis in the same way as above, the distributions for these two marks are actually quite similar (**figures 3.11, 3.12, 3.13**). The peaks in H3K4me3 abundance overlap with peaks of H3K27me3 abundance, signifying that these two marks are often found together in ES cells, agreeing with previous studies (Bernstein *et al.*, 2006, Jiang *et al.*, 2011, Sachs *et al.*, 2013). Although it seems in **figure 3.2** that the repressive mark is more widespread, FISH data suggests that the areas of specific H3K27me3 increases do coincide with H3K4me3 peaks, and may be rich in developmental loci, as is the case for chromosome 6 at 50-60Mb (**figure 3.13**).

6.3.2 Co-methylation of H3K4me3 and H3K27me3 throughout the mouse ES cell genome

This study has provided new insight into the behaviour of histone modifications in metaphase ES cells. The relationship between H3K4me3 and gene-rich regions of chromatin has been confirmed, and also that this relationship remains even into metaphase (**figure 3.2**). A reproducible karyotype has been produced that has made the sorting of H3K4me3 bands into specific and recognised regions of the genome possible (**figure 3.7**). This epigenome karyotype may also be able to predict gene density, CpG island density, and repeat density, as it has been shown that these features have a close relationship with H3K4me3 enrichment (**figures 3.9 and 3.10**). H3K27me3 has been shown to coincide in many cases with H3K4me3-rich chromosomal areas, and both are enriched at the chromosomal locus of the *Hoxa* cluster (**figures 3.7 and 3.11**). This karotype denotes H3K27me3-rich bands as being less well-defined than those for H3K4me3, due to a high general genomic background level of H3K27me3 as shown in photographs taken after immunofluorescence, as well as in ChIP seq data (**figures 3.2, 3.10 and 6.1**). There is evidence for specific deposition or maintenance of both H3K4me3 and H3K27me3 at the same loci, as shown by the analysed data of chromosomes in **figures 3.13 - 3.15**.

6.4 HISTONE MODIFICATIONS IN THE HOXA CLUSTER

6.4.1 Histone modifications at interphase and G_2/M at the Hoxa cluster

Results obtained from immunofluorescence studies have provided information about the distributions of histone modifications through differentiation and how they are aligned with DNA sequence features. To investigate further the nature and distributions of histone marks with regard to genetic features, I used chromatin immunoprecipitation to focus more closely on the relationships between histone modifications and genetic features through differentiation and the cell cycle (O' Neill and Turner, 2003). In order to prove or disprove any links between histone modifications and gene expression, I assayed expression of ES cells between day 0 and day 7 of a differentiation time-course by RNA extraction and subsequent qPCR. One of the central focuses of these experiments was the chromatin environment around the *Hoxa* cluster, and its expression in differentiating cells. As shown in **figure 4.3**, the expression of several *Hoxa* genes change rapidly upon cell differentiation as a response to retinoic acid, as previously reported (Chen and Gudas, 1996, Bami *et al.*, 2011). Interestingly in this study, *Hoxa1*, 4 and 5 have all been up-regulated after d2, whereas *Hoxa2* and 3 have remained silent. This contradicts the classical *Hox* cluster theory that genes are activated in chromosomal order from 1 to 13, although other studies have also suggested non-colinearity for other *Hox* clusters (Carroll, 1995, Soshnikova and Duboule, 2009). The expressions of pluripotency-related factors such as *Pou5f1* and *Nanog* have also fallen after at least 5 days of differentiation, confirming the loss of pluripotency for d5, d6 and d7 cells (**figure 4.4**). The expression of colcemid-stalled cells was also measured, allowing comparison

between G_2/M and asynchronous populations. This leads on from the chromosome immunofluorescence study that showed histone modifications persist at metaphase in mouse cells. Interestingly, results from **figure 4.3** show that there are no major differences in *Hoxa* expression between these two populations, indicating that expression after synthesis is equivalent to expression before synthesis and therefore that the *Hoxa* genes are not cell-cycle regulated.

6.4.2 Histone acetylation changes rapidly through the cell cycle in ES cells

The expression results led to the decision to use d5 differentiated cells as the sample for differentiated cells for N-ChIP experiments. A comparison was also made between asynchronous cells and colcemid-stalled cells to gauge whether histone marks were maintained at the G_2/M stage, or lost relative to an asynchronous population. Firstly, it is evident that on the whole, there are few differences between asynchronous cells and metaphase cells in terms of histone modifications (**figures 4.5 – 4.8**). *Hoxa5* has slightly higher enrichments of H3K4me3 and H3K27me3 in the metaphase population, however the greatest differences between asynchronous and G_2/M histone modifications can be seen in results for H3K9ac (**figure 4.7**). This shows a lack of transmission of the acetyl mark from S phase to G_2. It has been demonstrated that this is a cell-cycle specific mechanism whereby H3K9 is deacetylated and subsequently tri-methylated (but not di-methylated, as seen in **figure 4.7**) in preparation for mitosis (Park *et al.*, 2011). Rather than being lost after synthesis, it is likely that H3K9ac is intentionally and specifically removed from histone tails at G_2/M in preparation for chromatin condensation (Park *et al.*, 2011). The images obtained from immunofluorescence using antibodies for H3K9ac in **figure 3.5** show abundant H3K9ac at metaphase,

however it is not possible to compare this to pre-mitotic chromatin using the immunofluorescence technique. These observations highlight the extra layers of information that are extracted by focusing, or 'zooming in' the investigation into the *Hoxa* cluster by the use of chromatin immunoprecipitation.

6.5 BIVALENCY AT HOXA THROUGH DIFFERENTIATION

6.5.1 The bivalent mark is dynamically regulated at the early stages of differentiation at the Hoxa cluster

The bivalent mark has been found to occupy promoters of genes involved in the control of differentiation (Bernstein *et al.*, 2006, Jiang *et al.*, 2011, Vastenhouw and Schier, 2012). However, it is not clear how H3K4me3 and H3K27me3 levels change with respect to each other throughout differentiation. *Hox* genes represent an ideal model to explore this, as they are silent in ES cells, but show up-regulation at the onset of differentiation (**figure 4.3**) (Dolle and Duboule, 1989). The pattern of enrichment of these histone modifications at the ES cell stage gives a revealing picture of how the bivalent mark contributes to the overall state of the chromatin, and how it relates to expression of the underlying genes (**figures 4.3 and 4.6**). All *Hox* genes are silent at d0; however an enrichment of H3K4me3 is present at *Hoxa1* and *Hoxa4*, with a commensurate lack of H3K27me3 enrichment (**figure 4.6**). This shows that the permissive chromatin environment inferred by a strong bias towards the activating bivalent mark H3K4me3 is established before the expression of *Hoxa1* at d3 onwards. This shows a chromatin state predictive of future gene expression, and that two of the three most highly expressed *Hox* genes at d5 can be distinguished in this instance from the silent genes, based on

undifferentiated epigenetic signature, or in other words by determining part of their histone codes (**figure 4.6**).

6.5.2 Bivalency as a key regulator of developmental expression

Bivalency is thought to provide a poised state of chromatin whereby a gene's expression can be up-regulated after differentiation by the loss of the repressive half of the mark H3K27me3 (Bernstein *et al.*, 2006). This data, however, supports a more nuanced view of the bivalent mark. Both marks are enriched across the entire *Hoxa* cluster, supporting the theory that the two histone modifications are present at developmental loci in ES cells (**figure 4.6**) (Sachs *et al.*, 2013). The relative levels of the two marks seem to be indicative of future gene expression; so rather than a digital mechanism whereby the loss of the H3K27me3 modification results in an increase in expression, a bias towards the deposition of H3K4me3 and removal of H3K27me3 results over time in a switching-on of the associated gene. Graphs comparing d0 and d5 ChIP results for the two bivalent marks support this theory (**figure 4.10**). After five days of differentiation, the CCE/R cell population has switched on *Hoxa1* and *Hoxa4* expression (**figure 4.3**). The profiles of the bivalent marks at d5 reveal that these loci have been affected by increased H3K4me3 and decreased H3K27me3 enrichment. This may indicate a mechanism of histone modification controlled gene expression whereby a build-up of H3K4me3 and a depletion of H3K27me3 eventually cause a threshold to be breached and a switch in gene expression to occur. Alternatively, a progressive deposition of H3K4me3 may cause increased signalling to chromatin binding proteins, chromatin remodellers and the transcription machinery to overcome the negative influence of H3K27me3 and PRC2 at the gene promoter (Margueron and

Reinberg, 2011, Jia *et al.*, 2012). Soshnikova and Duboule noted a wave of demethylation of H3K27me3 as ES cells differentiated, starting from *Hox*d1 and moving onwards through the *Hox*d cluster, in a way analogous to the colinearity observed for *Hox* gene expression (Soshnikova and Duboule, 2009). For *Hoxa*, however, it seems from these that H3K27 demethylation is still specific for each *Hoxa* gene, rather than a non-specific wave of demethylation progressing in one direction (**figure 4.10**). *Hoxa2* and *Hoxa3* are still noticeably more methylated at H3K27me3 than *Hoxa4* at d5, and so follow the underlying expression of the genes rather than their position in the *Hoxa* cluster. The results for H3K4me3, however, show an increase in the mark on all *Hoxa* gene promoters, in a similar fashion to that observed for *Hox*d (**figure 4.10**) (Soshnikova and Duboule, 2009).

6.5.3 N-ChIP of differentiating ES cells provides a fresh perspective on the dynamics of bivalent histone modifications

Recent studies using the method of ChIP-sequencing have also shed light on the distributions of histone modifications in the region of the *Hoxa* cluster. UCSC genome browser data shown in **figure 6.1** shows the profiles of both bivalent histone modifications in both ES cells and differentiated liver cells (Rosenbloom *et al.*, 2013). As reflected in the data shown in this study, a decrease in H3K27me3 methylation occurs after differentiation at gene promoters (**figure 4.10**). Similarly to the N-ChIP data above, the H3K27me3 mark is still present after differentiation, although at a lower enrichment than before. However, the ChIP-seq data also shows that H3K4me3 enrichment is lower in terminally differentiated cells with respect to ES cells (Rosenbloom *et al.*, 2013). Findings from this study clearly show an increase in H3K4me3 after differentiation to mirror the decrease in

H3K27me3 (**figure 4.10**). This observation indicates that the *Hoxa* cluster in developing cells goes through a number of stages of changes to its surrounding chromatin through the course of differentiation, as suggested by immunofluorescence images (**figure 3.2**). The likely mechanism for the changes to the bivalent mark would involve a rise in trimethylation at H3K4 after the loss of pluripotency in order to promote the transcription of the underlying *Hox* genes. The fall in H3K27me3 could also facilitate this change. The H3K4me3 mark is then eroded either specifically by KDM5B, or non-specifically due to loss of H3K4me3 transmission through differentiation (Radman-Livaja *et al.*, 2010, Lloret-Llinares *et al.*, 2012). Both bivalent markers are still found at *Hoxa* loci after differentiation, although depleted, so that the epigenetic marker of developmental loci is maintained. This could form the basis of a heritable epigenetic signature, whilst still being plastic enough to respond to changing cellular environments (Turner, 2000, Hemberger *et al.*, 2009). These two roles of histone modifications may not be mutually exclusive, and could provide an answer to the question of how chromatin can control or assist in several aspects of the proper maintenance of the cell's function as it is, and to indicate future expression profiles.

It is probable that the loss of pluripotency has led to a specific increase in recruitment of MLL complexes to *Hoxa* through intermediate transcription factors that are up-regulated as a response to the fall in pluripotency factors such as Oct4 (Lubitz *et al.*, 2007). The increase in H3K4me3 is specific to the *Hoxa* cluster, as upstream and downstream sites are not enriched, although the increase in H3K4me3 seems to be similar for each gene within the cluster (**figure 4.10**). The discovery of asymmetrically bivalent nucleosomes points to a possible mechanism whereby two chromatin binding proteins with opposing effects on gene

transcription can recognise modified residues at each copy of histone H3 on the same nucleosome (Voigt *et al.*, 2012). The evidence here suggests that as more and more nucleosomes gain H3K4me3 and lose H3K27me3, a switch occurs whereby the level of H3K27me3 around the gene promoter is no longer enough to repress transcription of *Hoxa* genes and expression is up-regulated (**figures 4.3 and 4.10**). This can be thought of in a similar way to a transistor in a computer chip, which switches on only after a certain voltage threshold is reached (**figure 6.2**).

6.5.4 Boundary genes confirm bivalency as a marker of developmental loci

This study has not only focused closely on the *Hoxa* cluster, but also on the genes and intergenic areas surrounding it (**figure 2.3**). This allows us to discern whether changes in histone modification patterns are specific to *Hoxa* genes, any gene, or just a large section of chromatin regardless of the underlying features. Both bivalent marks in d0 and d5 cells show depletion at intergenic sequences, as well as at the upstream *Skap2* gene (**figure 4.10**). The *Evx1* gene promoter does show a bivalent motif, as it is a developmentally regulated gene expressed during gastrulation with a role similar to that of *Hox* genes (Dush and Martin, 1992). *Evx1* remains enriched in H3K27me3 even at d5, reflecting the lack of gene expression found throughout the time course shown (**figures 4.4 and 4.10**). *Skap2*, on the other hand, is expressed at a fairly constant level throughout the seven days of differentiation thanks to its close association with the function of actin, a vital housekeeping protein (Shimamura *et al.*, 2013). Interestingly, despite bordering the *Hoxa* cluster, the promoter of *Skap2* shows very little H3K4me3 and H3K27me3 enrichment, and in fact appears to be depleted in every histone

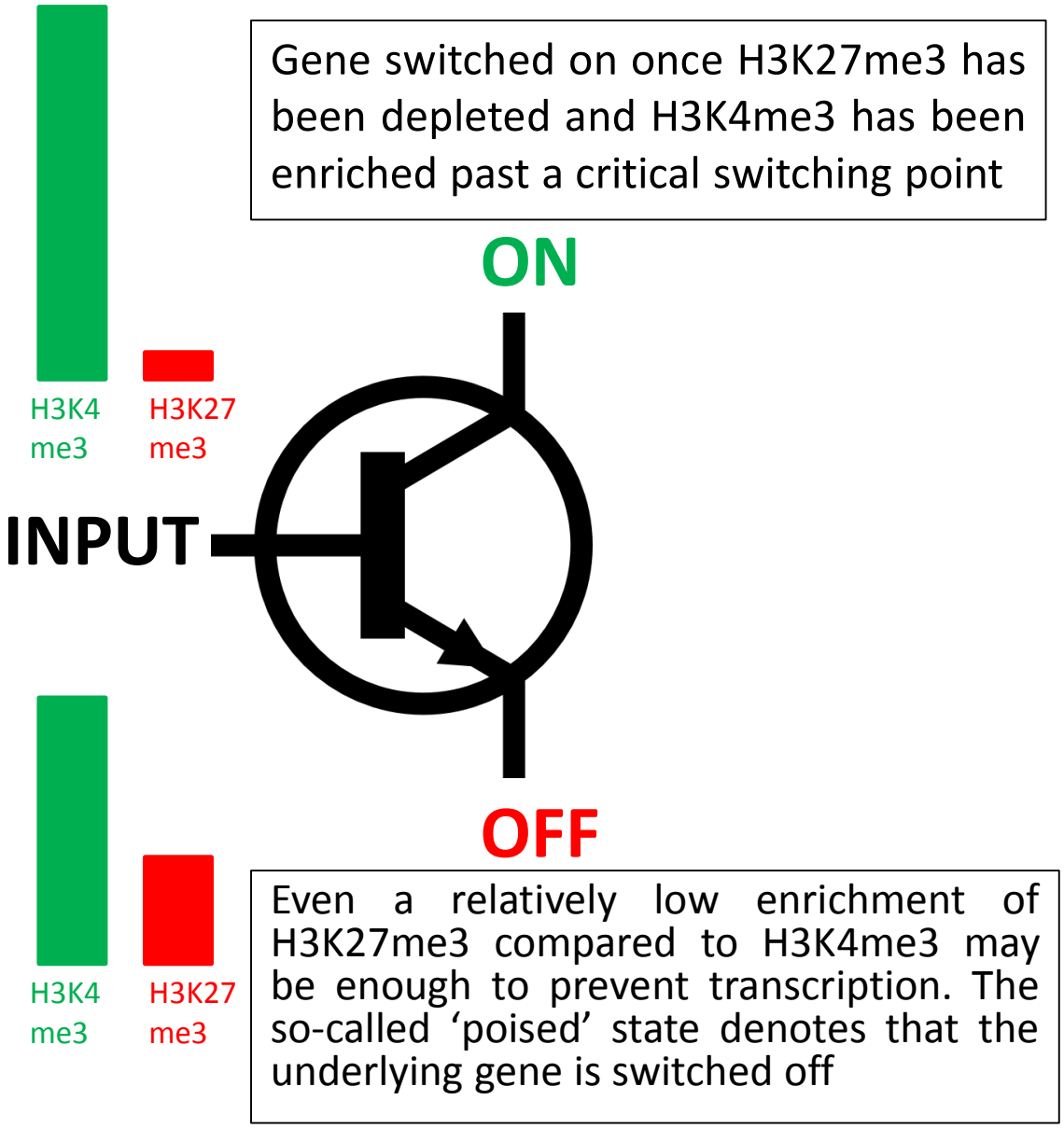


Figure 6.2 Transistor model of bivalency at developmental gene promoters. The switching on and off of genes through differentiation may be accomplished by bivalent histone modifications. The mechanism for this may resemble an electronic transistor, where the input voltage corresponds to how permissive the H3K4me3/H3K27me3 chromatin signature is. A poised state ensures that the transcription switch remains off, whereas if enough voltage (increased H3K4me3, decreased H3K27me3) is applied, the gene is switched on and transcription commences

modification tested in this study (**figures 4.6, 4.7, 4.8**). The high level of expression of *Skap2* is independent of the action of any of these histone modifications, although as it is likely to be constitutively expressed in all cell types, there may be no need for this locus to have any epigenetic regulation. These observations provide evidence supporting the theory that bivalent signatures are specific to developmental loci such as *Hoxa* and *Evx1*, and do not extend to surrounding chromatin (Azuara *et al.*, 2006).

6.5.5 Dynamics of histone modifications at non-developmental loci

The N-ChIP results for the housekeeping gene *Gapdh* and the pluripotency related genes *Pou5f1* and *Nanog* also provide insight into the specific nature of histone modification distributions. The pluripotency genes show a decrease in H3K4me3 abundance from d0 to d5, reflecting the drop in gene expression (**figures 4.4 and 4.9**). This is similar to results for *Hoxa1*, which show an increase in H3K4me3 related to an increase in expression (**figures 4.3 and 4.10**). However, the pluripotency genes are not marked by a bivalent motif as confirmed by N-ChIP data showing depletion of H3K27me3 before and after differentiation (**figure 4.9**). Although H3K27me3 displays a higher enrichment at d5 at these loci, this is below the general background level of H3K27me3 present in the unbound ChIP sample. *Gapdh*, as a housekeeping gene, shows little change in enrichment of either H3K4me3 or H3K27me3 during differentiation (**figure 4.9**).

6.6 BIVALENCY AMONG HOX CLUSTERS IN PREVIOUS STUDIES

6.6.1 ChIP-seq data mining in literature mouse studies

Data obtained from ChIP-sequencing studies from previous studies are presented in chapter 5. Shown in **figure 5.3 and 5.5** are graphs that demonstrate the wide variability between studies of differentiated embryonic mouse cells. This results from the difficulty in comparing studies that do not use cells from exactly the same stage in the differentiation process, from varying lineages or differentiated in diverse directions. For this reason, I have concentrated the analysis on the stage of development that is the most well-defined and well-studied; the d0 ES cell. For this type of cell, there have been several studies that analyse H3K4me3 and H3K27me3, the two components of the bivalent mark. These data have also been compared to my own results from undifferentiated ES cells (**figure 5.2 and 5.4**). These results show a good agreement between previous results and my data for the *Hoxa* cluster, suggesting that the bivalent pattern that I have observed throughout the cluster is representative of the true epigenetic state of ES cells. This adds weight to the hypothesis that at developmental loci, the bivalent mark is intimately involved with the epigenetic regulation of the future initiation of Hox gene expression during embryonic development. This ties in with the earlier expression data, confirming that *Hoxa*1 and *Hoxa*4, which show particularly strong H3K4me3 enrichment, are also expressed more strongly once induced by differentiation (**figures 4.3 and 5.6**). These graphs show *Hoxa*5 as having a similar expression induction profile after differentiation to *Hoxa*4, but a different bivalent signature at d0, characterised in both my data and in literature data as having lower H3K4me3 and higher H3K27me3 than *Hoxa*1 and *Hoxa*4. This could

indicate that *Hoxa*5 has a different epigenetic control pathway to the other genes, or it may imply that *Hoxa5* could share a regulatory pathway with *Hoxa4*, although the genomic distance between these two genes is in fact fairly substantial when compared to other *Hoxa* genes (**figure 2.5**).

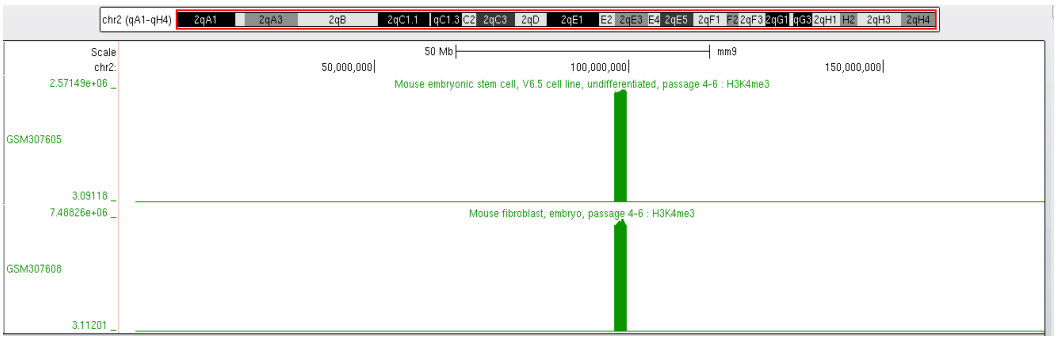
Once again, the specificity of the bivalent mark to Hox loci is suggested by these results, as other genes, even those involved with differentiation such as *Shh* and *Foxd1,* or those involved with maintaining pluripotency such as *Nanog* and *Pou5f1*, do not show the same bivalent signature as that found at all four Hox clusters (**figures 5.6 - 5.8**). These clusters show various patterns of distributions of the bivalent histone modifications, which could indicate a possible mechanism by which the epigenetic signature of the cluster affects the timing of expression activation *in vivo*. Judging by the changes in the enrichments of bivalent marks after initiation of differentiation, a gradual build-up of H3K4me3 as described in **figure 6.2** could occur in *Hoxb*, *Hoxc* and *Hoxd* as well, although at different times, to direct the transcription machinery as to the appropriate time to express these factors. For example, although *Hoxd* shows, like *Hoxa*, the most transcriptionally favourable bivalent state at *Hoxd1* and *Hoxd3*, *Hoxb* shows more favourable loci to be at the distal end of the cluster, and *Hoxc* shows no clear positional patterning. This may indicate that *Hoxd* is regulated in a similar way to *Hoxa*, and that *Hoxb* and *Hoxc* are not, or it could mean that proximal genes are switched on early in *Hoxd*, and later in *Hoxb*, for example (Zha *et al*., 2012). The timing of Hox gene expression is critical to the proper development of the embryo, so the epigenetic control of the on switch must be carefully controlled. An additive effect mechanism could be envisaged, whereby consecutive developmental stages are sensed by pathways involving various histone modifying enzymes and

H3K4me3 is added bit by bit. Once the embryo has reached the point where Hox expression is required, the level of H3K4me3 reaches the threshold for activation of expression and transcription proceeds.

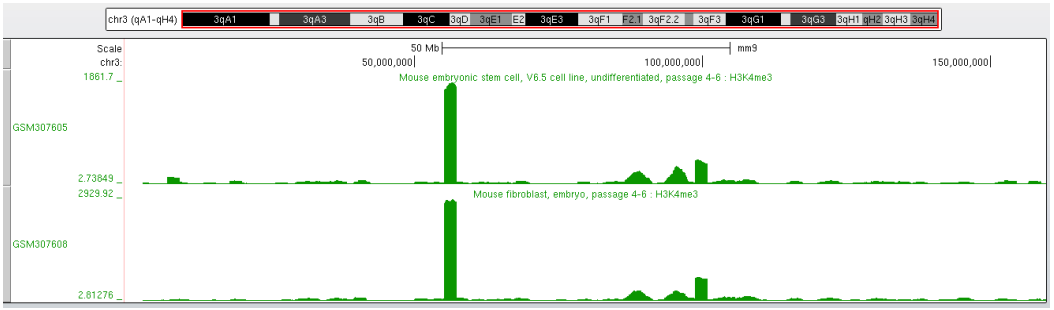
6.6.2 Genome-wide application of ChIP-seq data

ChIP-sequencing is designed to analyse the entire genome of the studied cells, therefore it is appropriate to determine whether the genome-wide karyotype data presented in chapter 3 is consistent or conflicting with these studies. **Figure 5.9** shows that patterns of H3K4me3 enrichment correspond to patterns of fluorescence found in **figures 3.11-3.17**. The differences of the techniques can be seen clearly in these data, as ChIP-seq experiments give more fine detail on the peaks and troughs of H3K4me3 density than immunofluorescence can. However, ChIP-seq experiments require a large population of cells in order to provide results, whereas each individual genome karyotype can be seen as part of the immunofluorescence procedure. Immunostained cells are also photographed at a very distinct and discrete stage of the cell cycle, whereas ChIP-seq cell populations are at all and any parts of the cell cycle, which can vary widely in their histone modification profiles. This can be shown clearly in **figure 6.3**, which shows UCSC browser data for chromosomes 2,3 and 4 in the same way as **figure 5.9**. The ChIP-seq results have several spikes in H3K4me3 enrichment that obscure the rest of the data, making it difficult to see the overall picture of H3K4me3 enrichment across the whole chromosome. This is not such a problem for immunofluorescence, with **figure 3.11** showing clear patterns of H3K4me3 density across all three of these chromosomes. There are also regions of DNA that cannot currently be analysed using sequencing, as evidenced by the gaps in data in

A



B



C

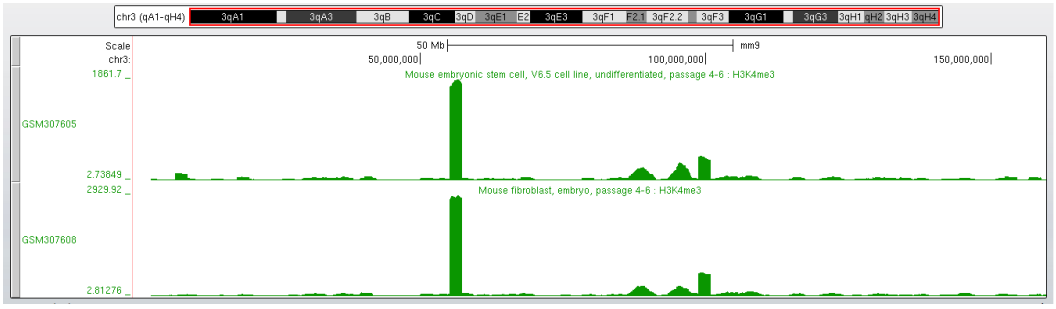


Figure 6.3 UCSC browser data showing ChIP-seq results for ES cells (upper – accession GSM307605) and embryonic fibroblasts (lower – accession GSM307608). The data bars indicate the level of enrichment of H3K4me3 throughout **A -** mouse chromosome 2, **B** – chromosome 3, **C** – chromosome 4. Chromosome ideograms and scale bar is shown above

figure 5.9 in chromosomes 1 and 7. This is due to long stretches of repeats in DNA which cannot be resolved using current sequencing techniques. Although both techniques have their strengths, they also both have weaknesses. Immunofluorescence is an excellent technique to use in order to look at single-cell histone modification profiles in a whole-genome at-once view, also maintaining a uniformity of cell-cycle phase between samples. The major drawback with this technique is its lack of resolution, which makes it impossible to distinguish histone modification patterns at the gene level. ChIP, and ChIP-seq specifically, do allow this resolution, and can offer valuable insights into the epigenetic profiles of cell populations at various stages of development. However, the construction of a whole-genome karyotype based on ChIP-seq data presents difficulties, as it must take a number of variables into consideration, such as cell to cell variation, cell cycle phase variation and sequencing difficulties. For these reasons, a wide-ranging immunofluorescent approach to determine whole genome histone modification profiles, coupled with a ChIP-based study to investigate the intricacies of histone modification distributions in differentiating cells, is an appropriate method of determining locations and effects of histone modifications in ES cells.

6.7 ACETYLATION THROUGH DIFFERENTIATION

6.7.1 Widespread hyperacetylation of H3K9 at interphase in ES cells

Other histone modifications assayed during this study did not show such a strong association with *Hoxa* gene transcription levels as the bivalent marks. The methyl marks H3K9me2 and H4K20me3 were not greatly enriched at the *Hoxa* cluster,

which is expected as they are markers of silent and heterochromatic regions of the genome (**figure 4.7 and 4.8**) (Schotta *et al.*, 2004, Lienert *et al.*, 2011). H3K27ac appears to be present at a constant level across the cluster and beyond, and does not show enrichment of greater than 2.5-fold anywhere (**figure 4.8**). H3K9ac, on the other hand, is enriched in asynchronous cells at most loci across the cluster and even within the intergenic stretch of chromatin upstream of *Hoxa1* (**figure 4.7**). This is also evident in immunofluorescence pictures for ES cells, which show widespread enrichment of H3K9ac in stem cells (**figure 3.5**). The enrichment of H3K9ac even in housekeeping genes and pluripotency related genes suggests that non-specific genome-wide acetylation is instigated during G_1/S phase in ES cells (**figures 4.5 and 4.7**). The difference between asynchronous cells and colcemid-stalled cells could represent the difference between inherited H3K9ac and transient H3K9ac.

6.7.2 Evidence of specific acetylation at Hoxa cluster after differentiation

There does not seem to be a preferential acetylation of the highly expressed *Hoxa1* and *Hoxa4* in the asynchronous population, although this pattern does re-emerge when looking at the data for G_2/M cells (**figure 4.7**). At this stage, the acetylation of H3K9 is highest at *Hoxa1* and *Hoxa4*, although the distinction between acetylation at these loci and at silent genes is less obvious than for H3K4me3 and H3K27me3. This may infer that histone H3 is heavily acetylated in ES cells, but on entering mitosis, these residues are methylated to facilitate chromatin compaction (Park *et al.*, 2011). This is confirmed by data shown in **figure 4.5**, showing a general enrichment of H3K9ac during interphase, even in housekeeping genes such as *Gapdh*. However, **figure 4.8** shows that H3K9me2

displays no changes in distribution within the *Hoxa* cluster through the cell cycle, perhaps indicating that H3K9 is methylated directly to H3K9me3.

The lower level of acetylation present at G_2/M phases may uncover the underlying chromatin signature that prepares the developmental loci for transcription soon after differentiation. This is borne out by the data presented in **figure 4.11**, which shows that acetylation of H3K9 has reduced in interphase to the same level as that at G_2/M, and is present across the *Hoxa* cluster. *Hoxa1* still shows the highest level of acetylation enrichment, although this is now only marginally higher than those seen for silent genes such as *Hoxa2* and *Hoxa10* (**figure 4.11**). The expression profiles for the *Hoxa* genes show that after the up-regulation of expression at d3, transcript levels decrease sharply after another 3 days of differentiation (**figure 4.3**). This could have implications for the transmission of histone modifications, as the H3K9ac mark could be specifically enriched at loci including *Hoxa1* before the onset of transcription. **Figures 4.10 and 4.11** suggest that H3K9ac is then no longer preferentially deposited at *Hoxa1* after differentiation, whereas H3K4me3 is maintained and even increased to provide a heritable marker of the chromatin. This mechanism shows the distinction between transient H3K9ac sites that facilitate a loosening of chromatin with those laying down an epigenetic signature.

6.8 HOXA EPIGENETIC SIGNATURES IN THE CONTEXT OF CHROMOSOME 6

The *Hoxa* cluster is of particular interest as it comprises a broad area of related developmental loci to enable the examination of distributions of histone modifications in this environment (Larson and Yuan, 2010). The chromosomal

data displayed in **figure 3.13** show that the region between 50 and 60 megabases from the centromere on chromosome 6 (containing *Hoxa*) is more heavily methylated at both H3K4me3 and H3K27me3 than for any other 10Mb block from 0 to 50 megabases. This gives an indication that bivalent modifications are targeted towards this region of the genome. Using N-ChIP to look more closely at the region has allowed us to determine the precise relationship between these two marks at *Hoxa* promoters and surrounding chromatin. This more focused viewpoint has allowed the observation that these marks are targeted specifically at *Hoxa* promoters and are not enriched at intergenic loci, or at non-developmental loci nearby (**figure 4.10**). The combination of these two techniques can on the one hand show us a picture of the entire epigenome at once, laying bare all the areas of increased methylation or acetylation of a residue. The patterns discovered here can be explored further with the use of chromatin immunoprecipitation to produce a detailed and specific view of a region of interest highlighted by the wider scan derived from immunofluorescence data. The use of *Hoxa* in this instance gives us a clear perspective on the dynamics of these histone modifications through differentiation at a locus that is known to be epigenetically switched on at some point during development (Soshnikova and Duboule, 2009). It is shown here that the behaviour of bivalent modifications may signify and predict future transcription levels (**figures 4.3 and 4.10**). The presence of both at equilibrium prevents transcription but indicates future expression due to its poised nature (Bernstein *et al.*, 2006). The subsequent reduction in H3K27me3 informs us of the loss of repression and the switching on of several *Hoxa* loci.

6.9 CONTROL OF DIFFERENTIATION BY THE BIVALENT MARK

6.9.1 The epigenetic code – Implications for the bivalent mark

These results also have significance for the examination of the histone code hypothesis (Strahl and Allis, 2000, Turner, 2000). This states that the modifications present on chromatin can inform the transcription machinery if, or maybe even when, underlying genes will be expressed. The results shown in **figure 4.10** suggest that the bivalent mark has a key role in this, by predicting the order of expression of the *Hoxa* cluster. *Hoxa1* shows a low enrichment of H3K27me3 at d0, and this gets even lower at d5 to become depleted below background H3K27me3 levels (**figure 4.10**). *Hoxa1* is also the first gene to become expressed and has the highest relative level of expression of all *Hoxa* genes (**figure 4.3**). *Hoxa4* has the second lowest level of H3K27me3 at d0, and the second highest H3K4me3 enrichment behind *Hoxa1*. This gene is also highly expressed after differentiation. *Hoxa5* has a higher than average level of H3K4me3 and lower than average H3K27me3, and is also expressed at d5 (**figures 4.3 and 4.10**). These three *Hoxa* genes are more highly expressed than the other *Hoxa* genes, and looking at the bivalent landscape at d0, one could infer a model to predict the expression of other genes through differentiation. Certainly the lack of H3K4me3 at *Hoxa2*, 3 and *Hoxa6*-11 would suggest that they are not likely to be expressed, or are not yet poised for expression. Therefore, these genes display a low level of transcription relative to *Hoxa1* (**figure 4.3**). The changes in H3K9ac at G_2/M after differentiation reflect the data for H3K4me3 in that there is a general increase in acetylation across the entire *Hoxa* gene cluster (**figure 4.11**). Looking at the changes in heritable H3K9ac at G_2/M from d0 to d5,

the pattern of acetylation follows the pattern of H3K4me3 methylation in that *Hoxa1* and *Hoxa4* show the highest levels of acetylation. The link between H3K9ac and transcription is not as clear as that for H3K4me3, however, as there are several *Hoxa* genes including *Hoxa10* that show high levels of H3K9ac but not much expression (**figures 4.3 and 4.11**). It may be that general *Hoxa* acetylation facilitates the methylation of H3K4me3 in order to drive a switch in gene expression, in accordance with other studies (Ha *et al.*, 2011).

6.9.2 A transcriptional switch for developmental loci

The theory of the bivalent mark being acutely involved in marking developmental loci is given added weight by these findings (Bernstein *et al.*, 2006). However, the results shown here have added an extra dimension to the argument behind the theory. It has been suggested that the bivalent mark denotes a 'poised' state whereby upon differentiation one half of the bivalent mark is lost and the gene is switched on (Bernstein *et al.*, 2006). However, close examination of the N-ChIP results shows that bivalent histone modifications could denote several different states of chromatin in different circumstances.

The *Hoxa* cluster is a developmental locus, and would be expected to be marked with bivalency in ES cells (Kashyap *et al.*, 2011). Indeed, for many of the *Hox* genes that is the case (**figure 4.6**). However, even within the relatively small genomic region of the *Hoxa* cluster, there are several different bivalent signatures present. *Hoxa1* shows a clear dominance of H3K4me3 over H3K27me3, which reveals its nature as being rapidly activated when responding to differentiation. *Hoxa2*, on the other hand, displays a reversal of these marks, its low abundance of H3K4me3 indicating that it is not poised for activation, although the increase in

the activating mark shown after 5 days of differentiation indicates that expression of *Hoxa2* is likely to occur at some point soon after the 7 days assayed here (**figures 4.3 and 4.10**). *Hoxa7*, on the other hand, shows a low level of H3K4me3 even after differentiation and strong (although reduced) enrichment of H3K27me3 at d5. This gene, according to the theory presented here, would be the last to be expressed. It may be the case that the bivalent mark signifies an analogue rather than a binary system, whereby the respective levels of the two marks can denote the timings where the underlying genes are activated. An assay of the bivalent mark at a given point in the process of differentiation could determine whether a gene is 'poised' or about to be expressed (*Hoxa1* at d0), is currently being expressed (*Hoxa1* at d5), is unlikely to be expressed soon (*Hoxa7* at d0 or d5), or is not developmentally regulated (*Skap2* at d0 and d5) (**figures 4.3, 4.4 and 4.10**). Since the levels of H3K4me3 and H3K27me3 are dependent on a number of factors and a number of other histone modifications, the control of differentiation likely rests in the hands of crosstalk between chromatin modifications, transcription factors and chromatin binding proteins, like other cellular processes (Van Attikum and Gasser, 2009, Zippo *et al.*, 2009). This can provide the responsiveness, specificity and plasticity needed to direct the progress of differentiation in a proper and carefully regulated manner.

6.10 CONCLUSIONS

The research into the locations and functions of histone modifications have in recent years grown into a vast wealth of data that indicate an ever more complex and elegant system for controlling many aspects of cellular regulation. These

results have given insight into where marks are heritable throughout the cell cycle, and throughout many cell cycles in the process of differentiation in mice. Heritable histone marks have been shown to persist through metaphase by chromosome immunofluorescence and through G_2/M by N-ChIP. Both methyl and acetyl modifications can be persistent, again shown in both immunofluorescence and ChIP experiments, although plasticity of histone modifications has been shown for other marks. H4K20me3 shows remarkable changes in its genomic distribution, being distributed throughout the chromosome arms in ES cells, but confined in differentiated cells to heterochromatic centromeres. More subtle changes were also seen in the H3K4me3 and H3K9me3 marks on chromosomes through differentiation, indicating some involvement in the differentiation process. The link between bivalent modifications and differentiation led to the definition of a mouse epigenomic karyotype being produced for both H3K4me3 and H3K27me3. This represents the sum total of the genomic organisation of these marks at metaphase, and can be aligned with genetic sequence data to produce an epigenetic map of these two marks and to discover the connections and relationships between histone modifications and their underlying DNA. The H3K4me3 karyotype data shows a close association between H3K4me3 and gene-rich regions, and could predict highly expressed regions of the genome. The H3K4me3 data correlated well with previous ChIP-seq studies in mice. H3K27me3 has been shown to loosely follow the same distributions as H3K4me3, although chromosome immunofluorescence has suggested that trimethyl H3K27 is a more widespread feature across the genome than H3K4me3.

Dynamic changes in histone modifications were also seen at more specific gene loci in results for N-ChIP, revealing a deeply nuanced relationship between

H3K4me3 and H3K27me3, the two bivalent marks. Data obtained here for the *Hoxa* cluster in d0 cells are mainly consistent with previous ChIP-seq studies. The state of the bivalent motif appears to be predictive of the expression of *Hoxa* genes during the differentiation process, and may form the basis for epigenetic control of differentiation. The trend throughout the differentiation process at *Hoxa* is the demethylation of H3K27me3 and methylation of H3K4me3 at all *Hoxa* promoters. The extent to which these processes occur is determined by how poised the genes are towards expression. The first genes to switch on are more heavily methylated at H3K4, and less methylated at H3K27me3. The genes retaining a relatively high level of H3K27me3 and a relatively low level of H3K4me3 after differentiation show no signs of being expressed. This relationship between H3K4me3 and H3K27me3 is deeply involved in the active or repressed status of chromatin, and this feature will be a major player in the elucidation of the histone code. Further investigation is required into how important a role in the determination of expression status the bivalent marker has. I have postulated the basis of a predictive model based on the status of the bivalent mark at two points during the differentiation process. If differentiation were continued for a more sustained period of time, it could be determined whether those predictions ring true, or whether other modifications and factors are influencing the switching on of developmental genes during differentiation. Investigation of other developmentally regulated genes, and not just *Hox* genes, would also confirm if this bivalent transistor switch is widely applicable to developmental loci. Extending the comparison with ChIP-seq data by determining the expression changes for *Hoxb, Hoxc and Hoxd* would also provide valuable comparison data for these other Hox gene clusters.

7. BIBLIOGRAPHY

Abraham, B. J., Cui, K., Tang, Q. & Zhao, K. 2013. Dynamic regulation of epigenomic landscapes during hematopoiesis. *BMC Genomics,* 14**,** 193.

Abranches, E., Bekman, E. & Henrique, D. 2013. Generation and characterization of a novel mouse embryonic stem cell line with a dynamic reporter of Nanog expression. *PLoS One,* 8**,** e59928.

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.

Adams-Cioaba, M. A. & Min, J. 2009. Structure and function of histone methylation binding proteins. *Biochem Cell Biol,* 87**,** 93-105.

Agalioti, T., Chen, G. & Thanos, D. 2002. Deciphering the transcriptional histone acetylation code for a human gene. *Cell,* 111**,** 381-92.

Alder, O., Lavial, F., Helness, A., Brookes, E., Pinho, S., Chandrashekran, A., Arnaud, P., Pombo, A., O' Neill, L. & Azuara, V. 2010. Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment. *Development,* 137**,** 2483-92.

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.

Andang, M., Hjerling-Leffler, J., Moliner, A., Lundgren, T. K., Castelo-Branco, G., Nanou, E., Pozas, E., Bryja, V., Halliez, S., Nishimaru, H., Wilbertz, J., Arenas, E., Koltzenburg, M., Charnay, P., El Manira, A., Ibanez, C. F. & Ernfors, P. 2008. Histone H2AX-dependent GABA(A) receptor regulation of stem cell proliferation. *Nature,* 451**,** 460-4.

Antequera, F., Boyes, J. & Bird, A. 1990. High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell,* 62**,** 503-14.

Arai, Y., Pulvers, J. N., Haffner, C., Schilling, B., Nusslein, I., Calegari, F. & Huttner, W. B. 2011. Neural stem and progenitor cells shorten S-phase on commitment to neuron production. *Nat Commun,* 2**,** 154.

Arnaudo, A. M. & Garcia, B. A. 2013. Proteomic characterization of novel histone post-translational modifications. *Epigenetics Chromatin,* 6**,** 24.

Ashraf, M. & Godward, M. B. 1980. The nucleolus in telophase, interphase and prophase. *J Cell Sci,* 41**,** 321-9.

Athanasiadou, R., De Sousa, D., Myant, K., Merusi, C., Stancheva, I. & Bird, A. 2010. Targeting of de novo DNA methylation throughout the Oct-4 gene regulatory region in differentiating embryonic stem cells. *PLoS One,* 5**,** e9937.

Augui, S., Nora, E. P. & Heard, E. 2011. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet,* 12**,** 429-42.

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.

Bakayev, V. V., Bakayeva, T. G. & Varshavsky, A. J. 1977. Nucleosomes and subnucleosomes: heterogeneity and composition. *Cell,* 11**,** 619-29.

Bami, M., Episkopou, V., Gavalas, A. & Gouti, M. 2011. Directed neural differentiation of mouse embryonic stem cells is a sensitive system for the identification of novel Hox gene effectors. *PLoS One,* 6**,** e20197.

Bannister, A. J. & Kouzarides, T. 2011. Regulation of chromatin by histone modifications. *Cell Res,* 21**,** 381-95.

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.

Barta, T., Vinarsky, V., Holubcova, Z., Dolezalova, D., Verner, J., Pospisilova, S., Dvorak, P. & Hampl, A. 2010. Human embryonic stem cells are capable of executing G1/S checkpoint activation. *Stem Cells,* 28**,** 1143-52.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M. & Kouzarides, T. 2010. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell,* 143**,** 470-84.

Becker, K. A., Stein, J. L., Lian, J. B., Van Wijnen, A. J. & Stein, G. S. 2007. Establishment of histone gene regulation and cell cycle checkpoint control in human embryonic stem cells. *J Cell Physiol,* 210**,** 517-26.

Becker, K. A., Stein, J. L., Lian, J. B., Van Wijnen, A. J. & Stein, G. S. 2010. Human embryonic stem cells are pre-mitotically committed to self-renewal and acquire a lengthened G1 phase upon lineage programming. *J Cell Physiol,* 222**,** 103-10.

Belch, Y., Yang, J., Liu, Y., Malkaram, S. A., Liu, R., Riethoven, J. J. & Ladunga, I. 2010. Weakly positioned nucleosomes enhance the transcriptional competency of chromatin. *PLoS One,* 5**,** e12984.

Berger, S. L. 2010. Cell signaling and transcriptional regulation via histone phosphorylation. *Cold Spring Harb Symp Quant Biol,* 75**,** 23-6.

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.

Bhaumik, S. R., Smith, E. & Shilatifard, A. 2007. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol,* 14**,** 1008-16.

Bian, Q. & Belmont, A. S. 2012. Revisiting higher-order and large-scale chromatin organization. *Curr Opin Cell Biol,* 24**,** 359-66.

Billon, N., Jolicoeur, C., Ying, Q. L., Smith, A. & Raff, M. 2002. Normal timing of oligodendrocyte development from genetically engineered, lineage-selectable mouse ES cells. *J Cell Sci,* 115**,** 3657-65.

Bird, A. P. 1986. CpG-rich islands and the function of DNA methylation. *Nature,* 321**,** 209-13.

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., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W. K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C. L., Ruan, Y., Struhl, K.,

Gerstein, M., Antonarakis, S. E., Fu, Y., Green, E. D., Karaoz, U., Siepel, A., Taylor, J., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Cooper, G. M., Asimenos, G., Dewey, C. N., Hou, M., Nikolaev, S., Montoya-Burgos, J. I., Loytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N. R., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Sidow, A., Trinklein, N. D., Zhang, Z. D., Barrera, L., Stuart, R., King, D. C., Ameur, A., Enroth, S., Bieda, M. C., Kim, J., Bhinge, A. A., Jiang, N., Liu, J., Yao, F., Vega, V. B., Lee, C. W., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J. C., Couttet, P., Bruce, A. W., Dovey, O. M., Ellis, P. D., Langford, C. F., Nix, D. A., Euskirchen, G., Hartman, S., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T. H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Zhang, X., Xu, M., Haidar, J. N., Yu, Y., Ruan, Y., Iyer, V. R., Green, R. D., Wadelius, C., Farnham, P. J., Ren, B., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Armengol, L., Bird, C. P., De Bakker, P. I., Kern, A. D., Lopez-Bigas, N., Martin, J. D., Stranger, B. E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrimsdottir, I. B., Huppert, J., Zody, M. C., Abecasis, G. R., Estivill, X., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V. V., Maskeri, B., Mcdowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B. & De Jong, P. J. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature,* 447**,** 799-816.

Bishop, R. P. & Young, I. T. 1977. The automated classification of mitotic phase for human chromosome spreads. *J Histochem Cytochem,* 25**,** 730-40.

Blacketer, M. J., Feely, S. J. & Shogren-Knaak, M. A. 2010. Nucleosome interactions and stability in an ordered nucleosome array model system. *J Biol Chem,* 285**,** 34597-607.

Bloushtain-Qimron, N., Yao, J., Shipitsin, M., Maruyama, R. & Polyak, K. 2009. Epigenetic patterns of embryonic and adult stem cells. *Cell Cycle,* 8**,** 809-17.

Bode, J., Gomez-Lira, M. M. & Schroter, H. 1983. Nucleosomal particles open as the histone core becomes hyperacetylated. *Eur J Biochem,* 130**,** 437-45.

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

Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S. & Jacobsen, S. E. 2007. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science,* 317**,** 1760-4.

Bowen, C., Ju, J. H., Lee, J. H., Paull, T. T. & Gelmann, E. P. 2013. Functional Activation of ATM by the Prostate Cancer Suppressor NKX3.1. *Cell Rep,* 4**,** 516-29.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R. & Young, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell,* 122**,** 947-56.

Branzei, D. & Foiani, M. 2006. The Rad53 signal transduction pathway: Replication fork stabilization, DNA repair, and adaptation. *Exp Cell Res,* 312**,** 2654-9.

Breiling, A., O' Neill, L. P., D'eliseo, D., Turner, B. M. & Orlando, V. 2004. Epigenome changes in active and inactive polycomb-group-controlled regions. *EMBO Rep,* 5**,** 976-82.

Bronner, C. 2011. Control of DNMT1 abundance in epigenetic inheritance by acetylation, ubiquitylation, and the histone code. *Sci Signal,* 4**,** pe3.

Buckley, B. A., Burkhart, K. B., Gu, S. G., Spracklin, G., Kershner, A., Fritz, H., Kimble, J., Fire, A. & Kennedy, S. 2012. A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature,* 489**,** 447-51.

Calder, A., Roth-Albin, I., Bhatia, S., Pilquil, C., Lee, J. H., Bhatia, M., Levadoux-Martin, M., Mcnicol, J., Russell, J., Collins, T. & Draper, J. S. 2013. Lengthened G1 phase indicates differentiation status in human embryonic stem cells. *Stem Cells Dev,* 22**,** 279-95.

Candido, E. P., Reeves, R. & Davie, J. R. 1978. Sodium butyrate inhibits histone deacetylation in cultured cells. *Cell,* 14**,** 105-13.

Cannuyer, J., Loriot, A., Parvizi, G. K. & De Smet, C. 2013. Epigenetic hierarchy within the MAGEA1 cancer-germline gene: promoter DNA methylation dictates local histone modifications. *PLoS One,* 8**,** e58743.

Carroll, S. B. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature,* 376**,** 479-85.

Cartwright, I. M., Genet, M. D. & Kato, T. A. 2013. A simple and rapid fluorescence in situ hybridization microwave protocol for reliable dicentric chromosome analysis. *J Radiat Res,* 54**,** 344-8.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. & Smith, A. 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell,* 113**,** 643-55.

Chen, A. C. & Gudas, L. J. 1996. An analysis of retinoic acid-induced gene expression and metabolism in AB1 embryonic stem cells. *J Biol Chem,* 271**,** 14971-80.

Chen, K., Hsu, L. T., Wu, C. Y., Chang, S. Y., Huang, H. T. & Chen, W. 2013a. CBARA1 plays a role in stemness and proliferation of human embryonic stem cells. *PLoS One,* 8**,** e63653.

Chen, T., Du, J. & Lu, G. 2012. Cell growth arrest and apoptosis induced by Oct4 or Nanog knockdown in mouse embryonic stem cells: a possible role of Trp53. *Mol Biol Rep,* 39**,** 1855-61.

Chen, Y., Lan, Q., Zheng, T., Zhao, N., Holford, T. R., Lerro, C., Dai, M., Huang, H., Liang, J., Ma, S., Leaderer, B., Boyle, P., Chanock, S., Rothman, N. & Zhang, Y. 2013b. Polymorphisms in JAK/STAT signaling pathway genes and risk of non-Hodgkin lymphoma. *Leuk Res,* 37**,** 1120-4.

Cheng, D., Guo, Y., Li, Z., Liu, Y., Gao, X., Gao, Y., Cheng, X., Hu, J. & Wang, H. 2012. Porcine induced pluripotent stem cells require LIF and maintain their developmental potential in early stage of embryos. *PLoS One,* 7**,** e51778.

Chicha, L., Feki, A., Boni, A., Irion, O., Hovatta, O. & Jaconi, M. 2011. Human Pluripotent Stem Cells Differentiated in Fully Defined Medium Generate Hematopoietic CD34 and CD34 Progenitors with Distinct Characteristics. *PLoS One,* 6**,** e14733.

Choi, Y. J. & Anders, L. 2013. Signaling through cyclin D-dependent kinases. *Oncogene*.

Cohen, B. B. & Deane, D. L. 1976. Template activity of unfixed metaphase chromosomes. *J Cell Sci,* 20**,** 215-9.

Coulombe, Y., Lemieux, M., Moreau, J., Aubin, J., Joksimovic, M., Berube-Simard, F. A., Tabaries, S., Boucherat, O., Guillou, F., Larochelle, C., Tuggle, C. K. & Jeannotte, L. 2010. Multiple promoters and alternative splicing: *Hoxa*5 transcriptional complexity in the mouse embryo. *PLoS One,* 5**,** e10600.

Cross, S. H. & Bird, A. P. 1995. CpG islands and genes. *Curr Opin Genet Dev,* 5**,** 309-14.

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.

Dai, L., Liu, Y., Liu, J., Wen, X., Xu, Z., Wang, Z., Sun, H., Tang, S., Maguire, A. R., Quan, J., Zhang, H. & Ye, T. 2013. A novel cyclinE/cyclinA-CDK inhibitor targets p27(Kip1) degradation, cell cycle progression and cell survival: implications in cancer therapy. *Cancer Lett,* 333**,** 103-12.

Dani, C., Chambers, I., Johnstone, S., Robertson, M., Ebrahimi, B., Saito, M., Taga, T., Li, M., Burdon, T., Nichols, J. & Smith, A. 1998. Paracrine induction of stem cell renewal by LIF-deficient cells: a new ES cell regulatory pathway. *Dev Biol,* 203**,** 149-62.

Das, C., Lucia, M. S., Hansen, K. C. & Tyler, J. K. 2009. CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature,* 459**,** 113-7.

De Arce, M. A., Law, E., Martin, L. & Masterson, J. G. 1982. A case of inverted insertion assessed by R and G banding. *J Med Genet,* 19**,** 148-51.

De La Serna, I. L., Ohkawa, Y. & Imbalzano, A. N. 2006. Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat Rev Genet,* 7**,** 461-73.

Deckbar, D., Jeggo, P. A. & Lobrich, M. 2011. Understanding the limitations of radiation-induced cell cycle checkpoints. *Crit Rev Biochem Mol Biol,* 46**,** 271-83.

Di Leonardo, A., Linke, S. P., Clarkin, K. & Wahl, G. M. 1994. DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts. *Genes Dev,* 8**,** 2540-51.

Ding, L., Pan, R., Huang, X., Wang, J. X., Shen, Y. T., Xu, L., Zhang, Y., Liu, Y., He, X. Q., Yang, X. J., Qi, Z. Q. & Wang, H. L. 2012. Changes in histone acetylation during oocyte meiotic maturation in the diabetic mouse. *Theriogenology,* 78**,** 784-92.

Dolle, P. & Duboule, D. 1989. Two gene members of the murine HOX-5 complex show regional and cell-type specific expression in developing limbs and gonads. *Embo j,* 8**,** 1507-15.

Duan, Q., Chen, H., Costa, M. & Dai, W. 2008. Phosphorylation of H3S10 blocks the access of H3K9 by specific antibodies and histone methyltransferase. Implication in regulating chromatin dynamics and epigenetic inheritance during mitosis. *J Biol Chem,* 283**,** 33585-90.

Dunlap, D., Yokoyama, R., Ling, H., Sun, H. Y., Mcgill, K., Cugusi, S. & Lucchesi, J. C. 2012. Distinct contributions of MSL complex subunits to the transcriptional enhancement responsible for dosage compensation in Drosophila. *Nucleic Acids Res,* 40**,** 11281-91.

Durston, A. J., Jansen, H. J., In Der Rieden, P. & Hooiveld, M. H. 2011. Hox collinearity - a new perspective. *Int J Dev Biol,* 55**,** 899-908.

Dush, M. K. & Martin, G. R. 1992. Analysis of mouse Evx genes: Evx-1 displays graded expression in the primitive streak. *Dev Biol,* 151**,** 273-87.

Ekwall, K., Olsson, T., Turner, B. M., Cranston, G. & Allshire, R. C. 1997. Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. *Cell,* 91**,** 1021-32.

Elliott, G. O., Murphy, K. J., Hayes, J. J. & Thiriet, C. 2013. Replication-independent nucleosome exchange is enhanced by local and specific acetylation of histone H4. *Nucleic Acids Res,* 41**,** 2228-38.

Euskirchen, G., Auerbach, R. K. & Snyder, M. 2012. SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions. *J Biol Chem,* 287**,** 30897-905.

Evans, J. W., Chang, J. A., Giaccia, A. J., Pinkel, D. & Brown, J. M. 1991. The use of fluorescence in situ hybridisation combined with premature chromosome condensation for the identification of chromosome damage. *Br J Cancer,* 63**,** 517-21.

Evans, M. J. & Kaufman, M. H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature,* 292**,** 154-6.

Fachinetti, D., Diego Folco, H., Nechemia-Arbely, Y., Valente, L. P., Nguyen, K., Wong, A. J., Zhu, Q., Holland, A. J., Desai, A., Jansen, L. E. & Cleveland, D. W. 2013. A two-step mechanism for epigenetic specification of centromere identity and function. *Nat Cell Biol,* 15**,** 1056-66.

Fanti, L. & Pimpinelli, S. 2008. HP1: a functionally multifaceted protein. *Curr Opin Genet Dev,* 18**,** 169-74.

Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H. & Allis, C. D. 2005. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature,* 438**,** 1116-22.

Fodor, B. D., Shukeir, N., Reuter, G. & Jenuwein, T. 2010. Mammalian Su(var) genes in chromatin control. *Annu Rev Cell Dev Biol,* 26**,** 471-501.

Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E. & Mattevi, A. 2005. Human histone demethylase LSD1 reads the histone code. *J Biol Chem,* 280**,** 41360-5.

Francis, N. J. 2009. Mechanisms of epigenetic inheritance: copying of polycomb repressed chromatin. *Cell Cycle,* 8**,** 3513-8.

Francis, N. J., Follmer, N. E., Simon, M. D., Aghia, G. & Butler, J. D. 2009. Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell,* 137**,** 110-22.

Fujii-Yamamoto, H., Kim, J. M., Arai, K. & Masai, H. 2005. Cell cycle and developmental regulations of replication factors in mouse embryonic stem cells. *J Biol Chem,* 280**,** 12976-87.

Fujioka, M., Lear, B. C., Landgraf, M., Yusibova, G. L., Zhou, J., Riley, K. M., Patel, N. H. & Jaynes, J. B. 2003. Even-skipped, acting as a repressor, regulates axonal projections in Drosophila. *Development,* 130**,** 5385-400.

Furey, T. S. 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet,* 13**,** 840-52.

Furnari, B., Blasina, A., Boddy, M. N., Mcgowan, C. H. & Russell, P. 1999. Cdc25 inhibited in vivo and in vitro by checkpoint kinases Cds1 and Chk1. *Mol Biol Cell,* 10**,** 833-45.

Fussner, E., Ching, R. W. & Bazett-Jones, D. P. 2011. Living without 30nm chromatin fibers. *Trends Biochem Sci,* 36**,** 1-6.

Gao, F., Wei, Z., An, W., Wang, K. & Lu, W. 2013. The interactomes of POU5F1 and SOX2 enhancers in human embryonic stem cells. *Sci Rep,* 3**,** 1588.

Gifford, C. A., Ziller, M. J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A. K., Kelley, D. R., Shishkin, A. A., Issner, R., Zhang, X., Coyne, M., Fostel, J. L., Holmes, L., Meldrim, J., Guttman, M., Epstein, C., Park, H., Kohlbacher, O., Rinn, J., Gnirke, A., Lander, E. S., Bernstein, B. E. & Meissner, A. 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell,* 153**,** 1149-63.

Glover, C. H., Marin, M., Eaves, C. J., Helgason, C. D., Piret, J. M. & Bryan, J. 2006. Meta-analysis of differentiating mouse embryonic stem cell gene expression kinetics reveals early change of a small gene set. *PLoS Comput Biol,* 2**,** e158.

Gonzalo, S., Garcia-Cao, M., Fraga, M. F., Schotta, G., Peters, A. H., Cotter, S. E., Eguia, R., Dean, D. C., Esteller, M., Jenuwein, T. & Blasco, M. A. 2005. Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. *Nat Cell Biol,* 7**,** 420-8.

Goyal, R., Reinhardt, R. & Jeltsch, A. 2006. Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic Acids Res,* 34**,** 1182-8.

Groth, A., Rocha, W., Verreault, A. & Almouzni, G. 2007. Chromatin challenges during DNA replication and repair. *Cell,* 128**,** 721-33.

Grunstein, M. 1997. Histone acetylation in chromatin structure and transcription. *Nature,* 389**,** 349-52.

Gruss, C., Wu, J., Koller, T. & Sogo, J. M. 1993. Disruption of the nucleosomes at the replication fork. *Embo j,* 12**,** 4533-45.

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.

Guerrero-Bosagna, C., Settles, M., Lucker, B. & Skinner, M. K. 2010. Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. *PLoS One,* 5.

Guo, Z., Deshpande, R. & Paull, T. T. 2010. ATM activation in the presence of oxidative stress. *Cell Cycle,* 9**,** 4805-11.

Ha, M., Ng, D. W., Li, W. H. & Chen, Z. J. 2011. Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Res,* 21**,** 590-8.

Hake, S. B. & Allis, C. D. 2006. Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". *Proc Natl Acad Sci U S A,* 103**,** 6428-35.

Haldar, S., Saini, A., Nanda, J. S., Saini, S. & Singh, J. 2011. Role of Swi6/HP1 self-association-mediated recruitment of Clr4/Suv39 in establishment and maintenance of heterochromatin in fission yeast. *J Biol Chem,* 286**,** 9308-20.

Hanna, J., Saha, K., Pando, B., Van Zon, J., Lengner, C. J., Creyghton, M. P., Van Oudenaarden, A. & Jaenisch, R. 2009. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature,* 462**,** 595-601.

Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., Rappsilber, J., Lerdrup, M. & Helin, K. 2008. A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol,* 10**,** 1291-300.

Hathaway, N. A., Bell, O., Hodges, C., Miller, E. L., Neel, D. S. & Crabtree, G. R. 2012. Dynamics and memory of heterochromatin in living cells. *Cell,* 149**,** 1447-60.

Hattori, N., Niwa, T., Kimura, K., Helin, K. & Ushijima, T. 2013. Visualization of multivalent histone modification in a single cell reveals highly concerted epigenetic changes on differentiation of embryonic stem cells. *Nucleic Acids Res,* 41**,** 7231-9.

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.

Hemberger, M., Dean, W. & Reik, W. 2009. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat Rev Mol Cell Biol,* 10**,** 526-37.

Heng, H. H., Goetze, S., Ye, C. J., Liu, G., Stevens, J. B., Bremer, S. W., Wykes, S. M., Bode, J. & Krawetz, S. A. 2004. Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J Cell Sci,* 117**,** 999-1008.

Hirota, T., Lipp, J. J., Toh, B. H. & Peters, J. M. 2005. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature,* 438**,** 1176-80.

Honda, S., Lewis, Z. A., Shimada, K., Fischle, W., Sack, R. & Selker, E. U. 2012. Heterochromatin protein 1 forms distinct complexes to direct histone deacetylation and DNA methylation. *Nat Struct Mol Biol,* 19**,** 471-7, s1.

Hong, L., Schroth, G. P., Matthews, H. R., Yau, P. & Bradbury, E. M. 1993. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *J Biol Chem,* 268**,** 305-14.

Huang, C., Xu, M. & Zhu, B. 2013. Epigenetic inheritance mediated by histone lysine methylation: maintaining transcriptional states without the precise restoration of marks? *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20110332.

Huangfu D. and Anderson K. V. 2006. Signaling from Smo to Ci/Gli: conservation and divergence of Hedgehog pathways from Drosophila to vertebrates. *Development* 133, 3-14

Hubscher, U., Maga, G. & Spadari, S. 2002. Eukaryotic DNA polymerases. *Annu Rev Biochem,* 71**,** 133-63.

Hurst, L. D. & Eyre-Walker, A. 2000. Evolutionary genomics: reading the bands. *Bioessays,* 22**,** 105-7.

Iavarone, A. & Massague, J. 1997. Repression of the CDK activator Cdc25A and cell-cycle arrest by cytokine TGF-beta in cells lacking the CDK inhibitor p15. *Nature,* 387**,** 417-22.

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R. & Bird, A. P. 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet,* 6.

Iwamoto, K., Tashima, Y., Hamada, H., Eguchi, Y. & Okamoto, M. 2008. Mathematical modeling and sensitivity analysis of G1/S phase in the cell cycle including the DNA-damage signal transduction pathway. *Biosystems,* 94**,** 109-17.

Jackson, V., Shires, A., Chalkley, R. & Granner, D. K. 1975. Studies on highly metabolically active acetylation and phosphorylation of histones. *J Biol Chem,* 250**,** 4856-63.

Jansen, L. E., Black, B. E., Foltz, D. R. & Cleveland, D. W. 2007. Propagation of centromeric chromatin requires exit from mitosis. *J Cell Biol,* 176**,** 795-805.

Jasencakova, Z., Scharf, A. N., Ask, K., Corpet, A., Imhof, A., Almouzni, G. & Groth, A. 2010. Replication stress interferes with histone recycling and predeposition marking of new histones. *Mol Cell,* 37**,** 736-43.

Jauch, R., Aksoy, I., Hutchins, A. P., Ng, C. K., Tian, X. F., Chen, J., Palasingam, P., Robson, P., Stanton, L. W. & Kolatkar, P. R. 2011. Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. *Stem Cells,* 29**,** 940-51.

Jhappan, C., Yusufzai, T. M., Anderson, S., Anver, M. R. & Merlino, G. 2000. The p53 response to DNA damage in vivo is independent of DNA-dependent protein kinase. *Mol Cell Biol,* 20**,** 4075-83.

Jia, J., Zheng, X., Hu, G., Cui, K., Zhang, J., Zhang, A., Jiang, H., Lu, B., Yates, J., 3rd, Liu, C., Zhao, K. & Zheng, Y. 2012. Regulation of pluripotency and self- renewal of ESCs through epigenetic-threshold modulation and mRNA pruning. *Cell,* 151**,** 576-89.

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.

Johnson, C. A., O' Neill, L. P., Mitchell, A. & Turner, B. M. 1998. Distinctive patterns of histone H4 acetylation are associated with defined sequence elements within both heterochromatic and euchromatic regions of the human genome. *Nucleic Acids Res,* 26**,** 994-1001.

Jones, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet,* 13**,** 484-92.

Jorgensen, S., Schotta, G. & Sorensen, C. S. 2013. Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res,* 41**,** 2797-806.

Kaneda, A., Fujita, T., Anai, M., Yamamoto, S., Nagae, G., Morikawa, M., Tsuji, S., Oshima, M., Miyazono, K. & Aburatani, H. 2011. Activation of Bmp2-Smad1 signal and its regulation by coordinated alteration of H3K27 trimethylation in Ras-induced senescence. *PLoS Genet,* 7**,** e1002359.

Karlic, R., Chung, H. R., Lasserre, J., Vlahovicek, K. & Vingron, M. 2010. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A,* 107**,** 2926-31.

Kashyap, V., Gudas, L. J., Brenet, F., Funk, P., Viale, A. & Scandura, J. M. 2011. Epigenomic reorganization of the clustered Hox genes in embryonic stem cells induced by retinoic acid. *J Biol Chem,* 286**,** 3250-60.

Kaufmann, W. K. & Paules, R. S. 1996. DNA damage and cell cycle checkpoints. *Faseb j,* 10**,** 238-47.

Kawai, H., Yamashita, T., Ohta, Y., Deguchi, K., Nagotani, S., Zhang, X., Ikeda, Y., Matsuura, T. & Abe, K. 2010. Tridermal tumorigenesis of induced pluripotent stem cells transplanted in ischemic brain. *J Cereb Blood Flow Metab,* 30**,** 1487-93.

Keohane, A. M., O' Neill L, P., Belyaev, N. D., Lavender, J. S. & Turner, B. M. 1996. X-Inactivation and histone H4 acetylation in embryonic stem cells. *Dev Biol,* 180**,** 618-30.

Khammanivong, A., Wang, C., Sorenson, B. S., Ross, K. F. & Herzberg, M. C. 2013. S100A8/A9 (Calprotectin) Negatively Regulates G2/M Cell Cycle Progression and Growth of Squamous Cell Carcinoma. *PLoS One,* 8**,** e69395.

Khan, A. A., Abel, P. D., Chaudhary, K. S., Gulzar, Z., Stamp, G. W. & Lalani, E. N. 2003. Inverse correlation between high level expression of cyclin E and proliferation index in transitional cell carcinoma of the bladder. *Mol Pathol,* 56**,** 353-61.

Khromov, T., Pantakani, D. V., Nolte, J., Wolf, M., Dressel, R., Engel, W. & Zechner, U. 2011. Global and gene-specific histone modification profiles of mouse multipotent adult germline stem cells. *Mol Hum Reprod,* 17**,** 166-74.

Kleinfeld, R. G. & Sisken, J. E. 1966. Morphological and kinetic aspects of mitotic arrest by and recovery from colcemid. *J Cell Biol,* 31**,** 369-79.

Konrad, C. G. 1963. Protein synthesis and RNA synthesis during mitosis in animal cells. *J Cell Biol,* 19**,** 267-77.

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

Kowenz-Leutz, E., Pless, O., Dittmar, G., Knoblich, M. & Leutz, A. 2010. Crosstalk between C/EBPbeta phosphorylation, arginine methylation, and SWI/SNF/Mediator implies an indexing transcription factor code. *Embo j,* 29**,** 1105-15.

Kubicek, S. & Jenuwein, T. 2004. A crack in histone lysine methylation. *Cell,* 119**,** 903-6.

Kuntz, K. & O' Connell, M. J. 2009. The G(2) DNA damage checkpoint: could this ancient regulator be the Achilles heel of cancer? *Cancer Biol Ther,* 8**,** 1433-9.

Kuo, M. H., Brownell, J. E., Sobel, R. E., Ranalli, T. A., Cook, R. G., Edmondson, D. G., Roth, S. Y. & Allis, C. D. 1996. Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature,* 383**,** 269-72.

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.

Lafontaine, J. G. & Chouinard, L. A. 1963. A correlated light and electron microscope study of the nucleolar material during mitosis in Vicia faba. *J Cell Biol,* 17**,** 167-201.

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.

Larson, J. L. & Yuan, G. C. 2010. Epigenetic domains found in mouse embryonic stem cells via a hidden Markov model. *BMC Bioinformatics,* 11**,** 557.

Le Martelot, G., Canella, D., Symul, L., Migliavacca, E., Gilardi, F., Liechti, R., Martin, O., Harshman, K., Delorenzi, M., Desvergne, B., Herr, W., Deplancke, B., Schibler, U., Rougemont, J., Guex, N., Hernandez, N. & Naef, F. 2012. Genome-wide RNA polymerase II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles. *PLoS Biol,* 10**,** e1001442.

Leffak, I. M., Grainger, R. & Weintraub, H. 1977. Conservative assembly and segregation of nucleosomal histones. *Cell,* 12**,** 837-45.

Lehnertz, B., Ueda, Y., Derijck, A. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T. & Peters, A. H. 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol,* 13**,** 1192-200.

Lengner, C. J., Camargo, F. D., Hochedlinger, K., Welstead, G. G., Zaidi, S., Gokhale, S., Scholer, H. R., Tomilin, A. & Jaenisch, R. 2007. Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell,* 1**,** 403-15.

Lewis, E. B. 1978. A gene complex controlling segmentation in Drosophila. *Nature,* 276**,** 565-70.

Li, G. & Zhou, L. 2013. Genome-wide identification of chromatin transitional regions reveals diverse mechanisms defining the boundary of facultative heterochromatin. *PLoS One,* 8**,** e67156.

Li, Y., Sun, L., Zhang, Y., Wang, D., Wang, F., Liang, J., Gui, B. & Shang, Y. 2011. The histone modifications governing TFF1 transcription mediated by estrogen receptor. *J Biol Chem,* 286**,** 13925-36.

Lienert, F., Mohn, F., Tiwari, V. K., Baubec, T., Roloff, T. C., Gaidatzis, D., Stadler, M. B. & Schubeler, D. 2011. Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS Genet,* 7**,** e1002090.

Lim, T. T., Geisen, C., Hesse, M., Fleischmann, B. K., Zimmermann, K. & Pfeifer, A. 2013. Lentiviral vector mediated thymidine kinase expression in pluripotent stem cells enables removal of tumorigenic cells. *PLoS One,* 8**,** e70543.

Lindahl Allen, M., Koch, C. M., Clelland, G. K., Dunham, I. & Antoniou, M. 2009. DNA methylation-histone modification relationships across the desmin locus in human primary cells. *BMC Mol Biol,* 10**,** 51.

Liu, J. Y., Ji, M. F., Wang, X. R., Luo, R. L., Ren, X., Liu, M. & Wang, Q. K. 2006. Detection of human chromosomal abnormalities using a new technique combining 4',6-diamidino-2-phenyl-indole staining and image analysis. *Clin Genet,* 69**,** 65-71.

Lloret-Llinares, M., Perez-Lluch, S., Rossell, D., Moran, T., Ponsa-Cobas, J., Auer, H., Corominas, M. & Azorin, F. 2012. dKDM5/LID regulates H3K4me3 dynamics at the transcription-start site (TSS) of actively transcribed developmental genes. *Nucleic Acids Res,* 40**,** 9493-505.

Lohse, B., Helgstrand, C., Kristensen, J. B., Leurs, U., Cloos, P. A., Kristensen, J. L. & Clausen, R. P. 2013. Posttranslational modifications of the histone 3 tail and their impact on the activity of histone lysine demethylases in vitro. *PLoS One,* 8**,** e67653.

Loyola, A., Bonaldi, T., Roche, D., Imhof, A. & Almouzni, G. 2006. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Mol Cell,* 24**,** 309-16.

Lubitz, S., Glaser, S., Schaft, J., Stewart, A. F. & Anastassiadis, K. 2007. Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Mol Biol Cell,* 18**,** 2356-66.

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.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., Plath, K. & Hochedlinger, K. 2007. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell,* 1**,** 55-70.

Mailand, N., Falck, J., Lukas, C., Syljuasen, R. G., Welcker, M., Bartek, J. & Lukas, J. 2000. Rapid destruction of human Cdc25A in response to DNA damage. *Science,* 288**,** 1425-9.

Margueron, R. & Reinberg, D. 2011. The Polycomb complex PRC2 and its mark in life. *Nature,* 469**,** 343-9.

Marks, H., Chow, J. C., Denissov, S., Francoijs, K. J., Brockdorff, N., Heard, E. & Stunnenberg, H. G. 2009. High-resolution analysis of epigenetic changes associated with X inactivation. *Genome Res,* 19**,** 1361-73.

Martin, C. & Zhang, Y. 2005. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol,* 6**,** 838-49.

Martin, G. R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A,* 78**,** 7634-8.

Martino, D., Loke, Y. J., Gordon, L., Ollikainen, M., Cruickshank, M. N., Saffery, R. & Craig, J. M. 2013. Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol,* 14**,** R42.

Marushige, K. 1976. Activation of chromatin by acetylation of histone side chains. *Proc Natl Acad Sci U S A,* 73**,** 3937-41.

Marzi, I., Cipolleschi, M. G., D'amico, M., Stivarou, T., Rovida, E., Vinci, M. C., Pandolfi, S., Dello Sbarba, P., Stecca, B. & Olivotto, M. 2013. The involvement of a Nanog, Klf4 and c-Myc transcriptional circuitry in the intertwining between neoplastic progression and reprogramming. *Cell Cycle,* 12**,** 353-64.

Masumoto, H., Nakato, R., Kanemaki, M., Shirahige, K. & Hachinohe, M. 2011. The inheritance of histone modifications depends upon the location in the chromosome in Saccharomyces cerevisiae. *PLoS One,* 6**,** e28980.

Matsuoka, Y., Takechi, S., Nakayama, T. & Yoneda, Y. 1994. Exogenous histone H1 injection into mitotic cells disrupts synchronous progression of mitotic events by delaying chromosome decondensation. *J Cell Sci,* 107 (Pt 3)**,** 693-701.

Matthews, A. G., Kuo, A. J., Ramon-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D. N., Walter, K. L., Utz, P. J., Shi, Y., Kutateladze, T. G., Yang, W., Gozani, O. & Oettinger, M. A. 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature,* 450**,** 1106-10.

Mattout, A., Biran, A. & Meshorer, E. 2011. Global epigenetic changes during somatic cell reprogramming to iPS cells. *J Mol Cell Biol,* 3**,** 341-50.

Mcmanus, K. J., Biron, V. L., Heit, R., Underhill, D. A. & Hendzel, M. J. 2006. Dynamic changes in histone H3 lysine 9 methylations: identification of a mitosis-specific function for dynamic methylation in chromosome congression and segregation. *J Biol Chem,* 281**,** 8888-97.

Meneghini, M. D., Wu, M. & Madhani, H. D. 2003. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell,* 112**,** 725-36.

Meng, X., Laidler, L. L., Kosmacek, E. A., Yang, S., Xiong, Z., Zhu, D., Wang, X., Dai, D., Zhang, Y., Wang, X., Brachova, P., Albitar, L., Liu, D., Ianzini, F., Mackey, M. A. & Leslie, K. K. 2013. Induction of mitotic cell death by overriding G2/M checkpoint in endometrial cancer cells with non-functional p53. *Gynecol Oncol,* 128**,** 461-9.

Meregalli, M. F., A and Torrente Y 2011. Stem Cell Therapy for Neuromuscular Diseases. *In:* GHOLAMREZANEZHAD, D. A. (ed.) *Stem Cells in Clinic and Research.* InTech.

Meselson, M. & Stahl, F. W. 1958. The replication of DNA in *Escherichia coli*. *Proc Natl Acad Sci U S A,* 44**,** 671-82.

Messier, V., Zenklusen, D. & Michnick, S. W. 2013. A nutrient-responsive pathway that determines M phase timing through control of B-cyclin mRNA stability. *Cell,* 153**,** 1080-93.

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.

Mirsky, A. E. 1971. The structure of chromatin. *Proc Natl Acad Sci U S A,* 68**,** 2945-8.

Moody, S. A., Klein, S. L., Karpinski, B. A., Maynard, T. M., & LaMantia, A.-S. 2013. On becoming neural: what the embryo can tell us about differentiating neural stem cells. *American Journal of Stem Cells*, *2*(2), 74–94.

Morgan, D. O. 1997. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol,* 13**,** 261-91.

Morillon, A., Karabetsou, N., Nair, A. & Mellor, J. 2005. Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. *Mol Cell,* 18**,** 723-34.

Mudryj, M., Devoto, S. H., Hiebert, S. W., Hunter, T., Pines, J. & Nevins, J. R. 1991. Cell cycle regulation of the E2F transcription factor involves an interaction with cyclin A. *Cell,* 65**,** 1243-53.

Muntion, S., Sanchez-Guijo, F. M., Carrancio, S., Villaron, E., Lopez, O., Diez-Campelo, M., San Miguel, J. F. & Del Canizo, M. C. 2012. Optimisation of mesenchymal stromal cells karyotyping analysis: implications for clinical use. *Transfus Med,* 22**,** 122-7.

Murzina, N., Verreault, A., Laue, E. & Stillman, B. 1999. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. *Mol Cell,* 4**,** 529-40.

Nagamachi, C. Y., Pieczarka, J. C., O' Brien, P. C., Pinto, J. A., Malcher, S. M., Pereira, A. L., Rissino, J., Mendes-Oliveira, A. C., Rossi, R. V. & Ferguson-Smith, M. A. 2013. FISH with whole chromosome and telomeric probes demonstrates huge karyotypic reorganization with ITS between two species of Oryzomyini (Sigmodontinae, Rodentia): Hylaeamys megacephalus probes on Cerradomys langguthi karyotype. *Chromosome Res,* 21**,** 107-19.

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., Hause, G., Aurich, A. C., Dorn, R., Jenuwein, T. & Reuter, G. 2005. Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis. *Embo j,* 24**,** 1418-29.

Navas, T. A., Sanchez, Y. & Elledge, S. J. 1996. RAD9 and DNA polymerase epsilon form parallel sensory branches for transducing the DNA damage checkpoint signal in Saccharomyces cerevisiae. *Genes Dev,* 10**,** 2632-43.

Ncbi. 2013. *NCBI MapViewer* [Online]. NCBI. Available: http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=10090 [Accessed May 2013].

Nekrasov, M., Amrichova, J., Parker, B. J., Soboleva, T. A., Jack, C., Williams, R., Huttley, G. A. & Tremethick, D. J. 2012. Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics. *Nat Struct Mol Biol,* 19**,** 1076-83.

Ng, P. M. & Lufkin, T. 2011. Embryonic stem cells: protein interaction networks. *Biomol Concepts,* 2**,** 13-25.

Nightingale, K. P., Baumann, M., Eberharter, A., Mamais, A., Becker, P. B. & Boyes, J. 2007. Acetylation increases access of remodelling complexes to their nucleosome targets to enhance initiation of V(D)J recombination. *Nucleic Acids Res,* 35**,** 6311-21.

Nightingale, K. P., O' Neill, L. P. & Turner, B. M. 2006. Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. *Curr Opin Genet Dev,* 16**,** 125-36.

Nishiyama, A., Sharov, A. A., Piao, Y., Amano, M., Amano, T., Hoang, H. G., Binder, B. Y., Tapnio, R., Bassey, U., Malinou, J. N., Correa-Cerro, L. S., Yu, H., Xin, L., Meyers, E., Zalzman, M., Nakatake, Y., Stagg, C., Sharova, L., Qian, Y., Dudekula, D., Sheer, S., Cadet, J. S., Hirata, T., Yang, H. T., Goldberg, I., Evans, M. K., Longo, D. L., Schlessinger, D. & Ko, M. S. 2013. Systematic repression of transcription factors reveals limited patterns of gene expression changes in ES cells. *Sci Rep,* 3**,** 1390.

Niwa, H. 2007. How is pluripotency determined and maintained? *Development,* 134**,** 635-46.

O' Donnell, M., Langston, L. & Stillman, B. 2013. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol,* 5.

O' Neill, L. P., Randall, T. E., Lavender, J., Spotswood, H. T., Lee, J. T. & Turner, B. M. 2003. X-linked genes in female embryonic stem cells carry an epigenetic mark prior to the onset of X inactivation. *Hum Mol Genet,* 12**,** 1783-90.

O' Neill, L. P., Spotswood, H. T., Fernando, M. & Turner, B. M. 2008. Differential loss of histone H3 isoforms mono-, di- and tri-methylated at lysine 4 during X-inactivation in female embryonic stem cells. *Biol Chem,* 389**,** 365-70.

O' Neill, L. P. & Turner, B. M. 2003. Immunoprecipitation of native chromatin: NChIP. *Methods,* 31**,** 76-82.

Oda, M., Oxley, D., Dean, W. & Reik, W. 2013. Regulation of lineage specific DNA hypomethylation in mouse trophectoderm. *PLoS One,* 8**,** e68846.

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.

Olinski, R., Jurgowiak, M. & Zaremba, T. 2010. Uracil in DNA--its biological significance. *Mutat Res,* 705**,** 239-45.

Oliver, S. S., Musselman, C. A., Srinivasan, R., Svaren, J. P., Kutateladze, T. G. & Denu, J. M. 2012. Multivalent recognition of histone tails by the PHD fingers of CHD5. *Biochemistry,* 51**,** 6534-44.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P. J., Chambers, I. & Wilson, V. 2012. The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development,* 139**,** 2288-98.

Pai, A. A., Bell, J. T., Marioni, J. C., Pritchard, J. K. & Gilad, Y. 2011. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet,* 7**,** e1001316.

Palacios, A., Munoz, I. G., Pantoja-Uceda, D., Marcaida, M. J., Torres, D., Martin-Garcia, J. M., Luque, I., Montoya, G. & Blanco, F. J. 2008. Molecular basis of histone H3K4me3 recognition by ING4. *J Biol Chem,* 283**,** 15956-64.

Papamokos, G. V., Tziatzos, G., Papageorgiou, D. G., Georgatos, S. D., Politou, A. S. & Kaxiras, E. 2012. Structural role of RKS motifs in chromatin interactions: a molecular dynamics study of HP1 bound to a variably modified histone tail. *Biophys J,* 102**,** 1926-33.

Park, J. A., Kim, A. J., Kang, Y., Jung, Y. J., Kim, H. K. & Kim, K. C. 2011. Deacetylation and methylation at histone H3 lysine 9 (H3K9) coordinate chromosome condensation during cell cycle progression. *Mol Cells,* 31**,** 343-9.

Parsons, G. G. & Spencer, C. A. 1997. Mitotic repression of RNA polymerase II transcription is accompanied by release of transcription elongation complexes. *Mol Cell Biol,* 17**,** 5791-802.

Pasini, D., Malatesta, M., Jung, H. R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O. N. & Helin, K. 2010. Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res,* 38**,** 4958-69.

Patzlaff, J. S., Terrenoire, E., Turner, B. M., Earnshaw, W. C. & Paulson, J. R. 2010. Acetylation of core histones in response to HDAC inhibitors is diminished in mitotic HeLa cells. *Exp Cell Res,* 316**,** 2123-35.

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.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J. C., Ferrier, P. & Spicuglia, S. 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *Embo j,* 30**,** 4198-210.

Pestana, A., Sudilovsky, O. & Pitot, H. C. 1971. Characterization of a cytoplasmic histone-coenzyme A acetyltransferase. *FEBS Lett,* 19**,** 83-86.

Peters, A. H. & Schubeler, D. 2005. Methylation of histones: playing memory with DNA. *Curr Opin Cell Biol,* 17**,** 230-8.

Piao, S., Lee, S. J., Xu, Y., Gwak, J., Oh, S., Park, B. J. & Ha, N. C. 2011. CK1epsilon targets Cdc25A for ubiquitin-mediated proteolysis under normal conditions and in response to checkpoint activation. *Cell Cycle,* 10**,** 531-7.

Plesca, D., Crosby, M. E., Gupta, D. & Almasan, A. 2007. E2F4 function in G2: maintaining G2-arrest to prevent mitotic entry with damaged DNA. *Cell Cycle,* 6**,** 1147-52.

Pohl, T. J., Brewer, B. J. & Raghuraman, M. K. 2012. Functional centromeres determine the activation time of pericentric origins of DNA replication in Saccharomyces cerevisiae. *PLoS Genet,* 8**,** e1002677.

Poirier, M. G. & Marko, J. F. 2002. Mitotic chromosomes are chromatin networks without a mechanically contiguous protein scaffold. *Proc Natl Acad Sci U S A,* 99**,** 15393-7.

Pradhan, S., Bacolla, A., Wells, R. D. & Roberts, R. J. 1999. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J Biol Chem,* 274**,** 33002-10.

Probst, A. V., Dunleavy, E. & Almouzni, G. 2009. Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol,* 10**,** 192-206.

Puschendorf, M., Terranova, R., Boutsma, E., Mao, X., Isono, K., Brykczynska, U., Kolb, C., Otte, A. P., Koseki, H., Orkin, S. H., Van Lohuizen, M. & Peters, A. H. 2008. PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat Genet,* 40**,** 411-20.

Qi, H. H., Sarkissian, M., Hu, G. Q., Wang, Z., Bhattacharjee, A., Gordon, D. B., Gonzales, M., Lan, F., Ongusaha, P. P., Huarte, M., Yaghi, N. K., Lim, H., Garcia, B. A., Brizuela, L., Zhao, K., Roberts, T. M. & Shi, Y. 2010. Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature,* 466**,** 503-7.

Radman-Livaja, M., Liu, C. L., Friedman, N., Schreiber, S. L. & Rando, O. J. 2010. Replication and active demethylation represent partially overlapping mechanisms for erasure of H3K4me3 in budding yeast. *PLoS Genet,* 6**,** e1000837.

Rea, S., Eisenhaber, F., O' Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. & Jenuwein, T. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature,* 406**,** 593-9.

Reik, W., Dean, W. & Walter, J. 2001. Epigenetic reprogramming in mammalian development. *Science,* 293**,** 1089-93.

Rens, W., Wallduck, M. S., Lovell, F. L., Ferguson-Smith, M. A. & Ferguson-Smith, A. C. 2010. Epigenetic modifications on X chromosomes in marsupial and monotreme mammals and implications for evolution of dosage compensation. *Proc Natl Acad Sci U S A,* 107**,** 17657-62.

Rhee, J. K., Kim, K., Chae, H., Evans, J., Yan, P., Zhang, B. T., Gray, J., Spellman, P., Huang, T. H., Nephew, K. P. & Kim, S. 2013. Integrated analysis of genome-wide DNA methylation and gene expression profiles in molecular subtypes of breast cancer. *Nucleic Acids Res*.

Ribeiro-Mason, K., Boulesteix, C., Fleurot, R., Aguirre-Lavin, T., Adenot, P., Gall, L., Debey, P. & Beaujean, N. 2012. H3S10 phosphorylation marks constitutive heterochromatin during interphase in early mouse embryos until the 4-cell stage. *J Reprod Dev,* 58**,** 467-75.

Rice, J. C. & Allis, C. D. 2001. Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr Opin Cell Biol,* 13**,** 263-73.

Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R., Learned, K., Kirkup, V. M., Wong, M. C., Maddren, M., Fang, R., Heitner, S. G., Lee, B. T., Barber, G. P., Harte, R. A., Diekhans, M., Long, J. C., Wilder, S. P., Zweig, A. S., Karolchik, D., Kuhn, R. M., Haussler, D. & Kent, W. J. 2013. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res,* 41**,** D56-63.

Rossetto, D., Truman, A. W., Kron, S. J. & Cote, J. 2010. Epigenetic modifications in double-strand break DNA damage signaling and repair. *Clin Cancer Res,* 16**,** 4543-52.

Rossow, P. W., Riddle, V. G. & Pardee, A. B. 1979. Synthesis of labile, serum-dependent protein in early G1 controls animal cell growth. *Proc Natl Acad Sci U S A,* 76**,** 4446-50.

Roudier, F., Ahmed, I., Berard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., Giraut, L., Despres, B., Drevensek, S., Barneche, F., Derozier, S., Brunaud, V., Aubourg, S., Schnittger, A., Bowler, C., Martin-Magniette, M. L., Robin, S., Caboche, M. & Colot, V. 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *Embo j,* 30**,** 1928-38.

Royle, S. J. 2012. The role of clathrin in mitotic spindle organisation. *J Cell Sci,* 125**,** 19-28.

Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schafer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T. & Reuter, G. 2007. Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol Cell,* 26**,** 103-15.

Ruiz, S., Panopoulos, A. D., Herrerias, A., Bissig, K. D., Lutz, M., Berggren, W. T., Verma, I. M. & Izpisua Belmonte, J. C. 2011. A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr Biol,* 21**,** 45-52.

Sachs, M., Onodera, C., Blaschke, K., Ebata, K. T., Song, J. S. & Ramalho-Santos, M. 2013. Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. *Cell Rep,* 3**,** 1777-84.

Santenard, A. & Torres-Padilla, M. E. 2009. Epigenetic reprogramming in mammalian reproduction: contribution from histone variants. *Epigenetics,* 4**,** 80-4.

Santocanale, C. & Diffley, J. F. 1998. A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication. *Nature,* 395**,** 615-8.

Santos, J., Pereira, C. F., Di-Gregorio, A., Spruce, T., Alder, O., Rodriguez, T., Azuara, V., Merkenschlager, M. & Fisher, A. G. 2010. Differences in the epigenetic and reprogramming properties of pluripotent and extra-embryonic stem cells implicate chromatin remodelling as an important early event in the developing mouse embryo. *Epigenetics Chromatin,* 3**,** 1.

Scharf, A. N., Barth, T. K. & Imhof, A. 2009. Establishment of histone modifications after chromatin assembly. *Nucleic Acids Res,* 37**,** 5032-40.

Scholer, H. R., Balling, R., Hatzopoulos, A. K., Suzuki, N. & Gruss, P. 1989. Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *EMBO J,* 8**,** 2551-7.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D. & Jenuwein, T. 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev,* 18**,** 1251-62.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., Van Lohuizen, M., Tanay, A. & Cavalli, G. 2009. Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biol,* 7**,** e13.

Schwartz, Y. B., Kahn, T. G., & Pirrotta, V. 2005. Characteristic Low Density and Shear Sensitivity of Cross-Linked Chromatin Containing Polycomb Complexes . *Molecular and Cellular Biology*, *25*(1), 432–439. doi:10.1128/MCB.25.1.432-439.2005

Schwartz, Y. B., Kahn, T. G., Stenberg, P., Ohno, K., Bourgon, R. & Pirrotta, V. 2010. Alternative epigenetic chromatin states of polycomb target genes. *PLoS Genet,* 6**,** e1000805.

Sclafani, R. A. & Holzen, T. M. 2007. Cell cycle regulation of DNA replication. *Annu Rev Genet,* 41**,** 237-80.

Serandour, A. A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R., Salbert, G. & Eeckhoute, J. 2011. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res,* 21**,** 555-65.

Sha, K. B. L. A. 2009. The chromatin signature of pluripotent cells. *StemBook [Internet].* Cambridge (MA): Harvard Stem Cell Institute.

Shaw, M. L., Williams, E. J., Hawes, S. & Saffery, R. 2009. Characterisation of histone variant distribution in human embryonic stem cells by transfection of in vitro transcribed mRNA. *Mol Reprod Dev,* 76**,** 1128-42.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A. & Shi, Y. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell,* 119**,** 941-53.

Shibata, A., Barton, O., Noon, A. T., Dahm, K., Deckbar, D., Goodarzi, A. A., Lobrich, M. & Jeggo, P. A. 2010. Role of ATM and the damage response mediator proteins 53BP1 and MDC1 in the maintenance of G(2)/M checkpoint arrest. *Mol Cell Biol,* 30**,** 3371-83.

Shimamura, S., Sasaki, K. & Tanaka, M. 2013. The Src substrate SKAP2 regulates actin assembly by interacting with WAVE2 and cortactin proteins. *J Biol Chem,* 288**,** 1171-83.

Shin, D. M., Liu, R., Wu, W., Waigel, S. J., Zacharias, W., Ratajczak, M. Z. & Kucia, M. 2012. Global gene expression analysis of very small embryonic-like stem cells reveals that the Ezh2-dependent bivalent domain mechanism contributes to their pluripotent state. *Stem Cells Dev,* 21**,** 1639-52.

Singer, M. F. 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell,* 28**,** 433-4.

Singh, J., Freeling, M. & Lisch, D. 2008. A position effect on the heritability of epigenetic silencing. *PLoS Genet,* 4**,** e1000216.

Singh, N., Basnet, H., Wiltshire, T. D., Mohammad, D. H., Thompson, J. R., Heroux, A., Botuyan, M. V., Yaffe, M. B., Couch, F. J., Rosenfeld, M. G. & Mer, G. 2012. Dual recognition of phosphoserine and phosphotyrosine in histone variant H2A.X by DNA damage response protein MCPH1. *Proc Natl Acad Sci U S A,* 109**,** 14381-6.

Soshnikova, N. & Duboule, D. 2009. Epigenetic temporal control of mouse Hox genes in vivo. *Science,* 324**,** 1320-3.

Steiner, M., Clark, B., Tang, J. Z., Zhu, T. & Lobie, P. E. 2012. 14-3-3sigma mediates G2-M arrest produced by 5-aza-2'-deoxycytidine and possesses a tumor suppressor role in endometrial carcinoma cells. *Gynecol Oncol,* 127**,** 231-40.

Sterneckert, J., Hoing, S. & Scholer, H. R. 2012. Concise review: Oct4 and more: the reprogramming expressway. *Stem Cells,* 30**,** 15-21.

Strahl, B. D. & Allis, C. D. 2000. The language of covalent histone modifications. *Nature,* 403**,** 41-5.

Sugiyama, K., Sugiura, K., Hara, T., Sugimoto, K., Shima, H., Honda, K., Furukawa, K., Yamashita, S. & Urano, T. 2002. Aurora-B associated protein phosphatases as negative regulators of kinase activation. *Oncogene,* 21**,** 3103-11.

Sullivan, K. F., Hechenberger, M. & Masri, K. 1994. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J Cell Biol,* 127**,** 581-92.

Svotelis, A., Gevry, N. & Gaudreau, L. 2009. Regulation of gene expression and cellular proliferation by histone H2A.Z. *Biochem Cell Biol,* 87**,** 179-88.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. & Yamanaka, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131**,** 861-72.

Takahashi, K. & Yamanaka, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126**,** 663-76.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C. & Kennison, J. A. 1992. brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell,* 68**,** 561-72.

Tarasov, K. V., Tarasova, Y. S., Tam, W. L., Riordon, D. R., Elliott, S. T., Kania, G., Li, J., Yamanaka, S., Crider, D. G., Testa, G., Li, R. A., Lim, B., Stewart, C. L., Liu, Y., Van Eyk, J. E., Wersto, R.

P., Wobus, A. M. & Boheler, K. R. 2008. B-MYB is essential for normal cell cycle progression and chromosomal stability of embryonic stem cells. *PLoS One,* 3**,** e2478.

Taylor, J. H. 1973. Replication of DNA in mammalian chromosomes: isolation of replicating segments. *Proc Natl Acad Sci U S A,* 70**,** 1083-7.

Terrenoire, E., Mcronald, F., Halsall, J. A., Page, P., Illingworth, R. S., Taylor, A. M., Davison, V., O' Neill, L. P. & Turner, B. M. 2010. Immunostaining of modified histones defines high-level features of the human metaphase epigenome. *Genome Biol,* 11**,** R110.

Theunissen, T. W. & Silva, J. C. 2011. Switching on pluripotency: a perspective on the biological requirement of Nanog. *Philos Trans R Soc Lond B Biol Sci,* 366**,** 2222-9.

Tian, Y., Jia, Z., Wang, J., Huang, Z., Tang, J., Zheng, Y., Tang, Y., Wang, Q., Tian, Z., Yang, D., Zhang, Y., Fu, X., Song, J., Liu, S., Van Velkinburgh, J. C., Wu, Y. & Ni, B. 2011. Global mapping of H3K4me1 and H3K4me3 reveals the chromatin state-based cell type-specific gene regulation in human Treg cells. *PLoS One,* 6**,** e27770.

Tichy, E. D. 2011. Mechanisms maintaining genomic integrity in embryonic stem cells and induced pluripotent stem cells. *Exp Biol Med (Maywood),* 236**,** 987-96.

Trojer, P. & Reinberg, D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell,* 28**,** 1-13.

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.

Tucker, K. L., Talbot, D., Lee, M. A., Leonhardt, H. & Jaenisch, R. 1996. Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. *Proc Natl Acad Sci U S A,* 93**,** 12920-5.

Turner, B. M. 1998. Histone acetylation as an epigenetic determinant of long-term transcriptional competence. *Cell Mol Life Sci,* 54**,** 21-31.

Turner, B. M. 2000. Histone acetylation and an epigenetic code. *Bioessays,* 22**,** 836-45.

Turner, B. M. 2007. Defining an epigenetic code. *Nat Cell Biol,* 9**,** 2-6.

Turner, B. M. & O' Neill, L. P. 1995. Histone acetylation in chromatin and chromosomes. *Semin Cell Biol,* 6**,** 229-36.

Unoki, M., Brunet, J. & Mousli, M. 2009. Drug discovery targeting epigenetic codes: the great potential of UHRF1, which links DNA methylation and histone modifications, as a drug target in cancers and toxoplasmosis. *Biochem Pharmacol,* 78**,** 1279-88.

Unoki, M., Nishidate, T. & Nakamura, Y. 2004. ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene,* 23**,** 7601-10.

Van Attikum, H. & Gasser, S. M. 2009. Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol,* 19**,** 207-17.

Vastenhouw, N. L. & Schier, A. F. 2012. Bivalent histone modifications in early embryogenesis. *Curr Opin Cell Biol,* 24**,** 374-86.

Vavouri, T. & Lehner, B. 2012. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biol,* 13**,** R110.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., Mckusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., Mccawley, S., Mcintosh, T., Mcmullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M.,

Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., Mcdaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. & Zhu, X. 2001. The sequence of the human genome. *Science,* 291**,** 1304-51.

Vermilyea, M. D., O' Neill, L. P. & Turner, B. M. 2009. Transcription-independent heritability of induced histone modifications in the mouse preimplantation embryo. *PLoS One,* 4**,** e6086.

Voigt, P., Leroy, G., Drury, W. J., 3rd, Zee, B. M., Son, J., Beck, D. B., Young, N. L., Garcia, B. A. & Reinberg, D. 2012. Asymmetrically modified nucleosomes. *Cell,* 151**,** 181-93.

Voigt, P., Tee, W. W. & Reinberg, D. 2013. A double take on bivalent promoters. *Genes Dev,* 27**,** 1318-38.

Wang, Y. & Blelloch, R. 2009. Cell cycle regulation by MicroRNAs in embryonic stem cells. *Cancer Res,* 69**,** 4093-6.

Wang, Y., Melton, C., Li, Y. P., Shenoy, A., Zhang, X. X., Subramanyam, D. & Blelloch, R. 2013. miR-294/miR-302 Promotes Proliferation, Suppresses G1-S Restriction Point, and Inhibits ESC Differentiation through Separable Mechanisms. *Cell Rep,* 4**,** 99-109.

Wang, Y., Yates, F., Naveiras, O., Ernst, P. & Daley, G. Q. 2005. Embryonic stem cell-derived hematopoietic stem cells. *Proc Natl Acad Sci U S A,* 102**,** 19081-6.

Waterborg, J. H. 1998. Dynamics of histone acetylation in Chlamydomonas reinhardtii. *J Biol Chem,* 273**,** 27602-9.

Weinert, T. A. & Hartwell, L. H. 1990. Characterization of RAD9 of Saccharomyces cerevisiae and evidence that its function acts posttranslationally in cell cycle arrest after DNA damage. *Mol Cell Biol,* 10**,** 6554-64.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. 2009. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet,* 41**,** 246-50.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. & Young, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell,* 153**,** 307-19.

Xu, M. & Zhu, B. 2010. Nucleosome assembly and epigenetic inheritance. *Protein Cell,* 1**,** 820-9.

Xu, W., Cho, H., Kadam, S., Banayo, E. M., Anderson, S., Yates, J. R., 3rd, Emerson, B. M. & Evans, R. M. 2004. A methylation-mediator complex in hormone signaling. *Genes Dev,* 18**,** 144-56.

Xu, Y., Ayrapetov, M. K., Xu, C., Gursoy-Yuzugullu, O., Hu, Y. & Price, B. D. 2012. Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair. *Mol Cell,* 48**,** 723-33.

Yamagata, K., Suetsugu, R. & Wakayama, T. 2009. Assessment of chromosomal integrity using a novel live-cell imaging technique in mouse embryos produced by intracytoplasmic sperm injection. *Hum Reprod,* 24**,** 2490-9.

Yamanaka, K., Balboula, A. Z., Sakatani, M. & Takahashi, M. 2010. Gene silencing of DNA methyltransferases by RNA interference in bovine fibroblast cells. *J Reprod Dev,* 56**,** 60-7.

Yanagida, A., Ito, K., Chikada, H., Nakauchi, H. & Kamiya, A. 2013. An In Vitro Expansion System for Generation of Human iPS Cell-Derived Hepatic Progenitor-Like Cells Exhibiting a Bipotent Differentiation Potential. *PLoS One,* 8**,** e67541.

Yi, H. & Richards, E. J. 2009. Gene duplication and hypermutation of the pathogen Resistance gene SNC1 in the Arabidopsis bal variant. *Genetics,* 183**,** 1227-34.

Zee, B. M., Levin, R. S., Dimaggio, P. A. & Garcia, B. A. 2010. Global turnover of histone post-translational modifications and variants in human cells. *Epigenetics Chromatin,* 3**,** 22.

Zha, Y., Ding, E., Yang, L., Mao, L., Wang, X., McCarthy, B.A., Huang, S., Ding, H-F. 2012. Functional Dissection of HOXD Cluster Genes in Regulation of Neuroblastoma Cell Proliferation and Differentiation. *PLoS ONE,* 7(8): e40728.

Zhang, J., Parvin, J. & Huang, K. 2012. Redistribution of H3K4me2 on neural tissue specific genes during mouse brain development. *BMC Genomics,* 13 Suppl 8**,** S5.

Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M. & Jacobsen, S. E. 2009. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. *Genome Biol,* 10**,** R62.

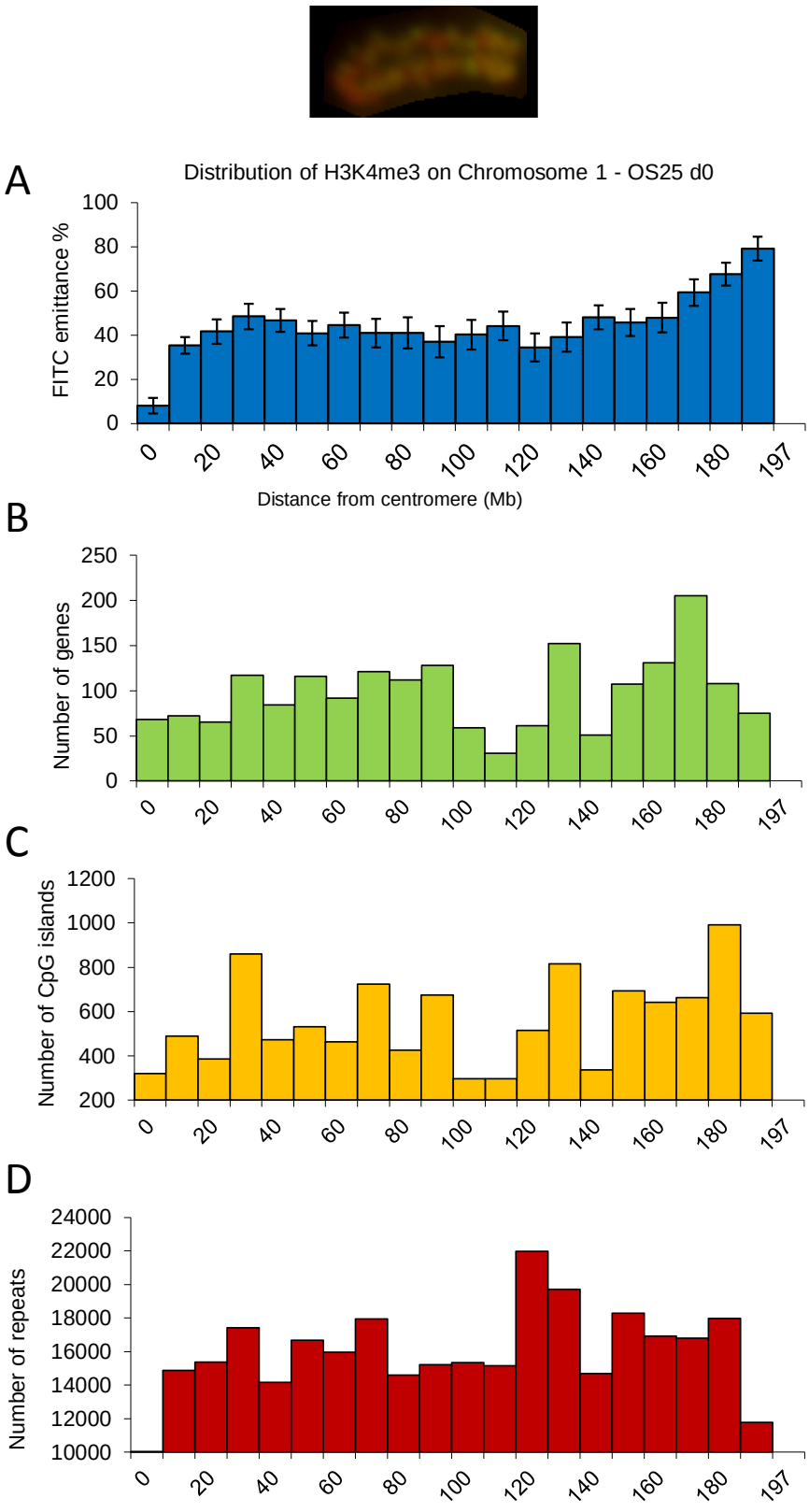
Zheng, S. & Pan, Y. X. 2010. Histone modifications, not DNA methylation, cause transcriptional repression of p16 (CDKN2A) in the mammary glands of offspring of protein-restricted rats. *J Nutr Biochem*.

Zhou, Q., Chipperfield, H., Melton, D. A. & Wong, W. H. 2007. A gene regulatory network in mouse embryonic stem cells. *Proc Natl Acad Sci U S A,* 104**,** 16438-43.

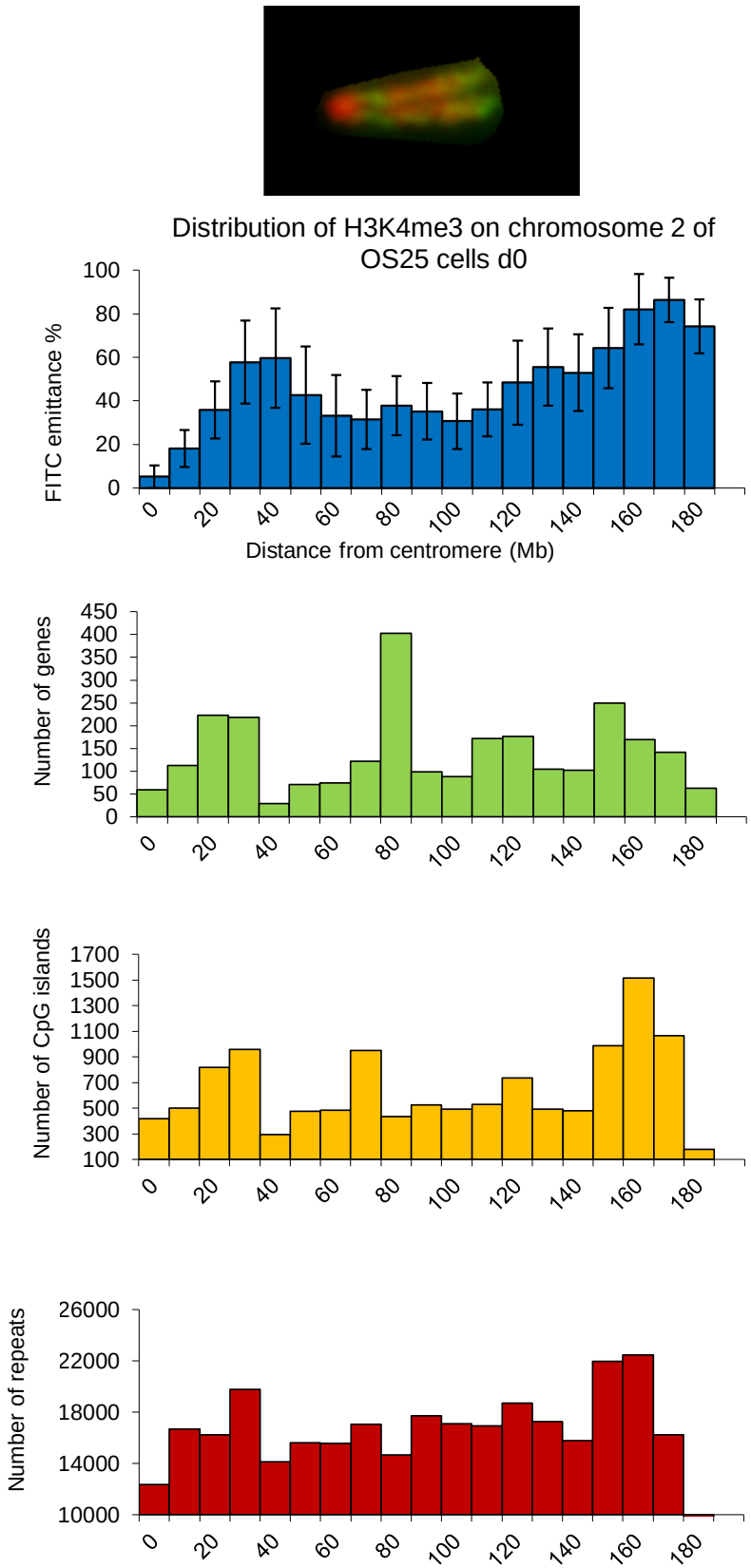
Zhu, S., Wang, G., Liu, B. & Wang, Y. 2013. Modeling exon expression using histone modifications. *PLoS One,* 8**,** e67448.

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A. & Oliviero, S. 2009. Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell,* 138**,** 1122-36.

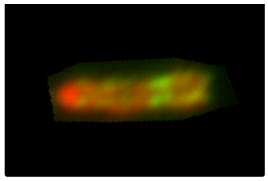
8. APPENDIX

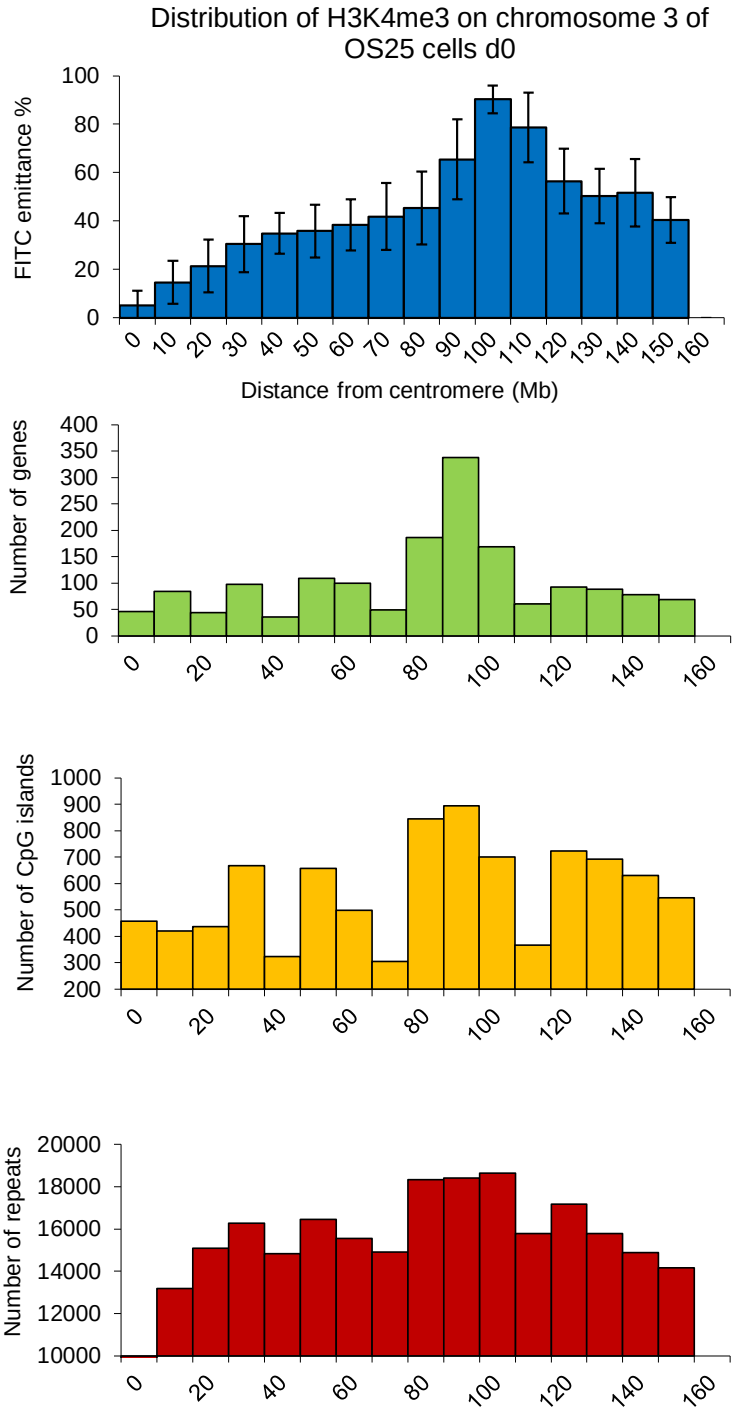


Appendix figure 1 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 1 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates

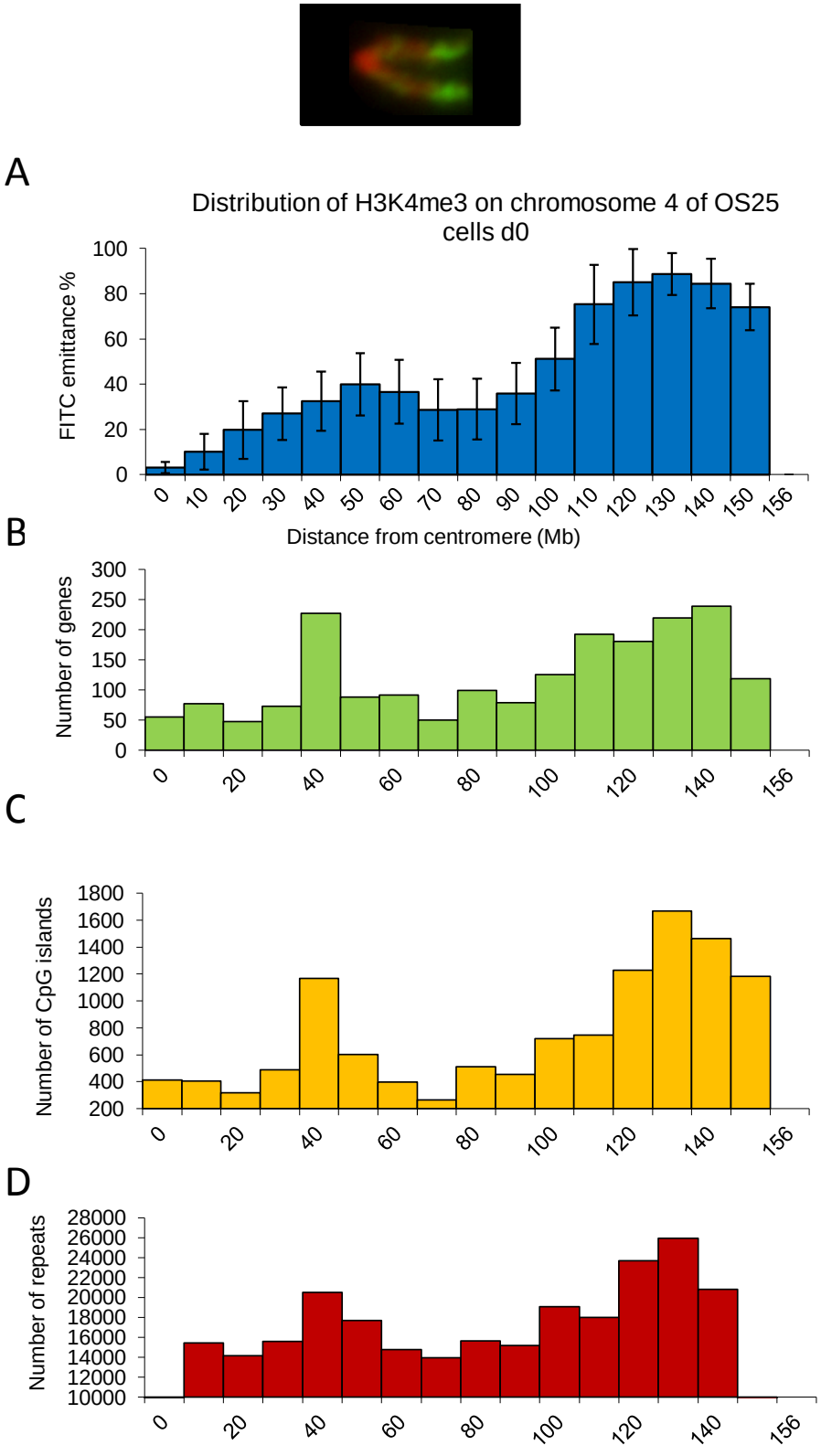


Appendix figure 2 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 2 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates

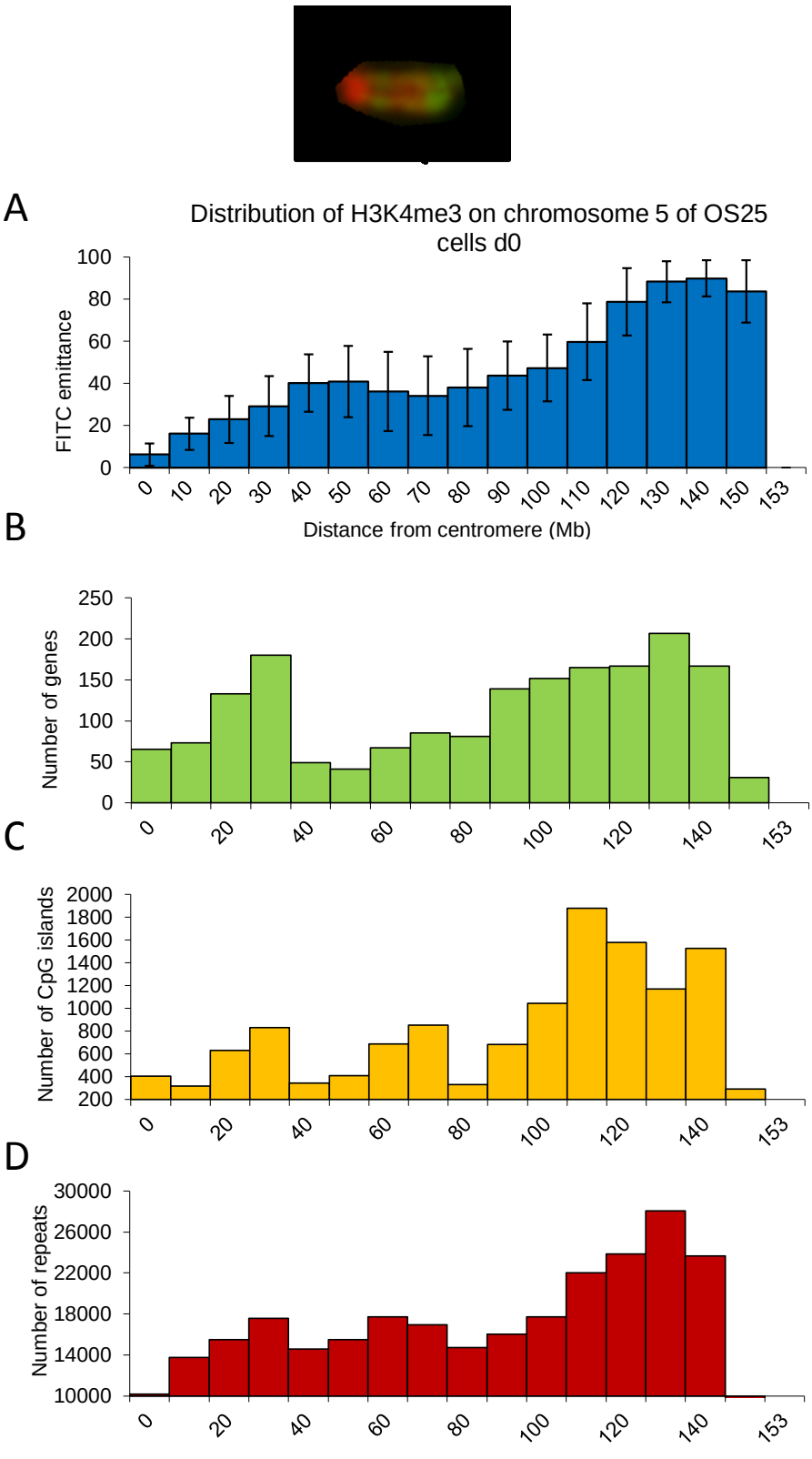




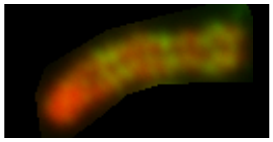
Appendix figure 3 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 3 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates

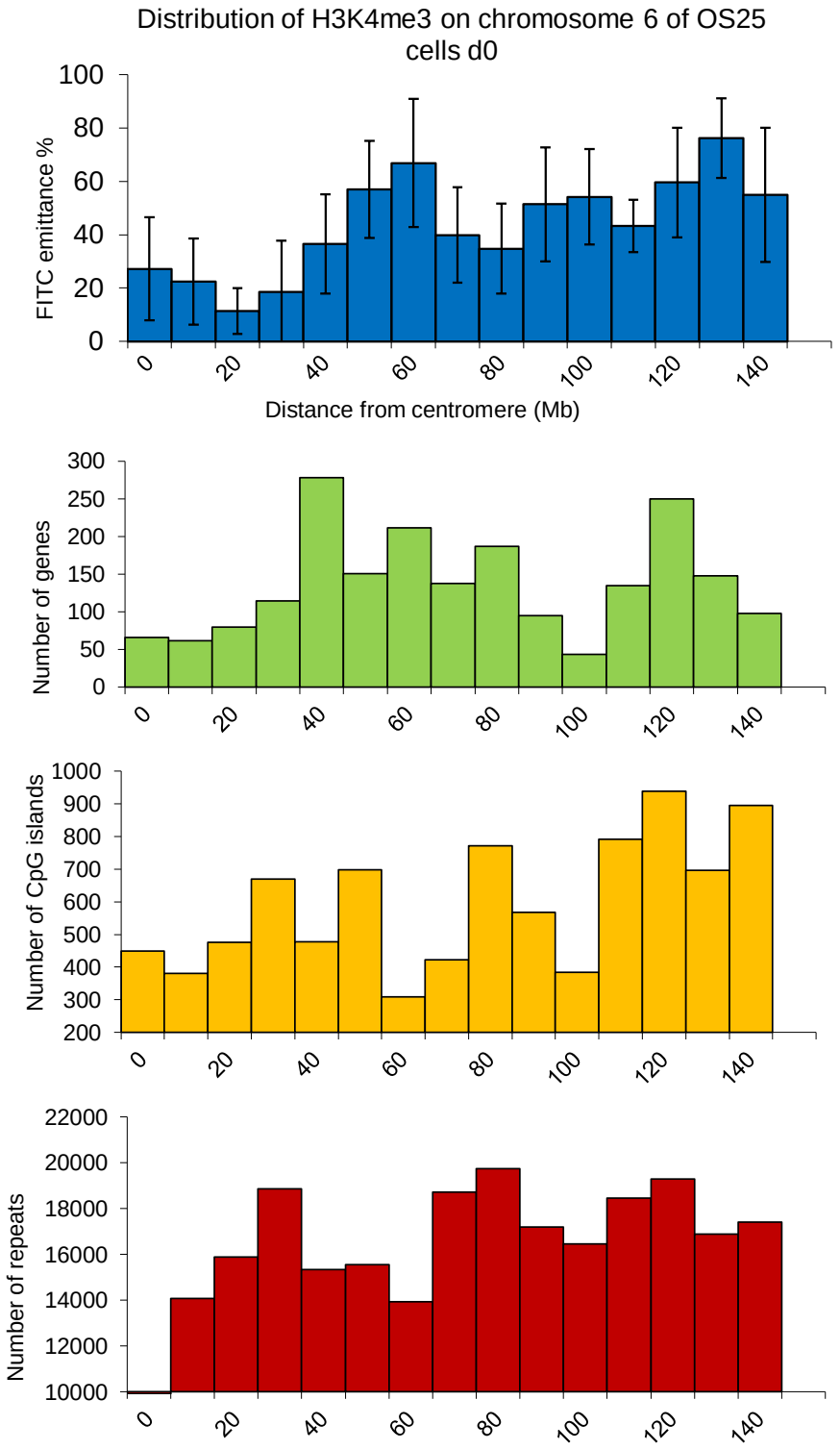


Appendix figure 4 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 4 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates

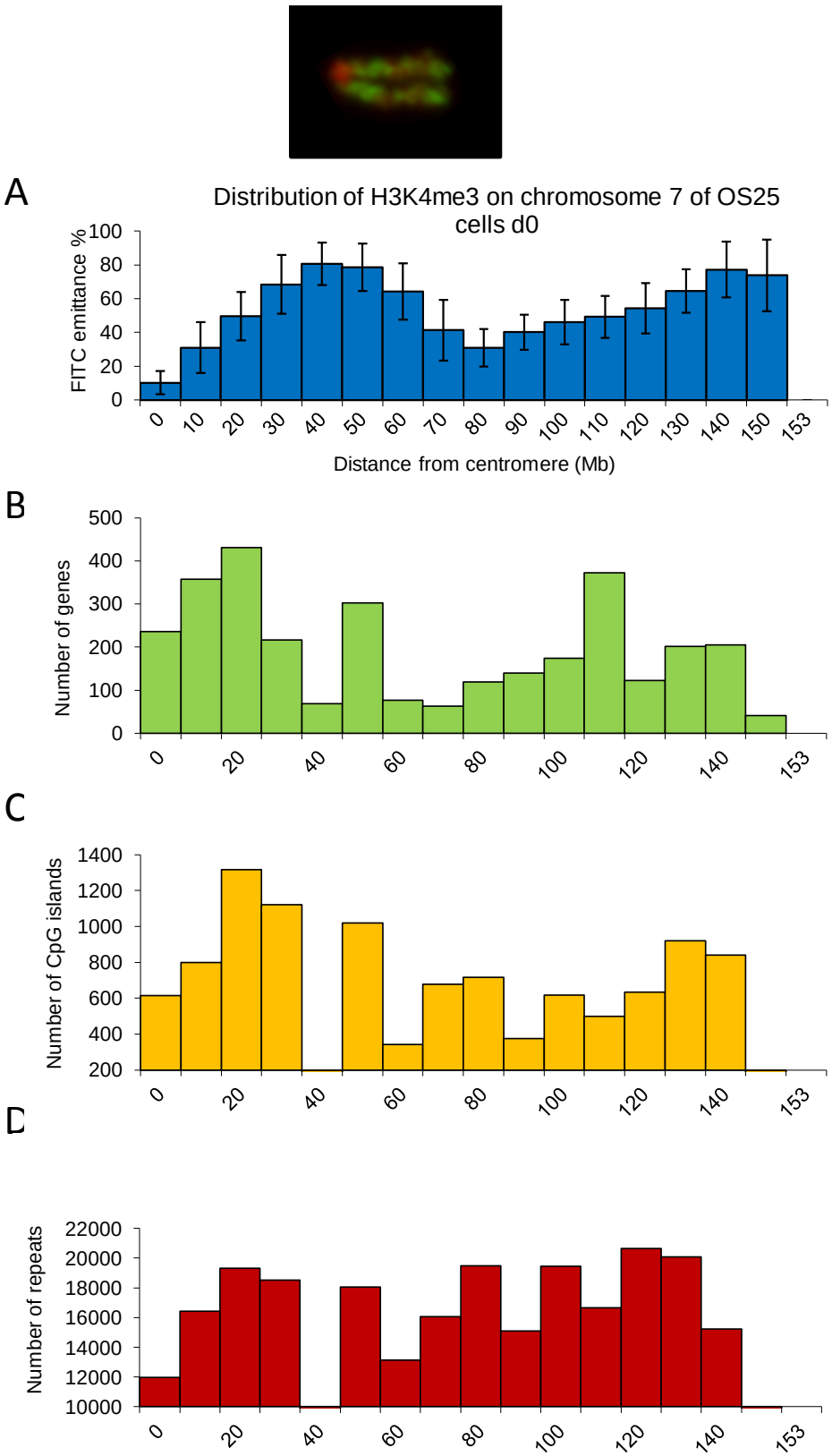


Appendix figure 5 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 5 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates

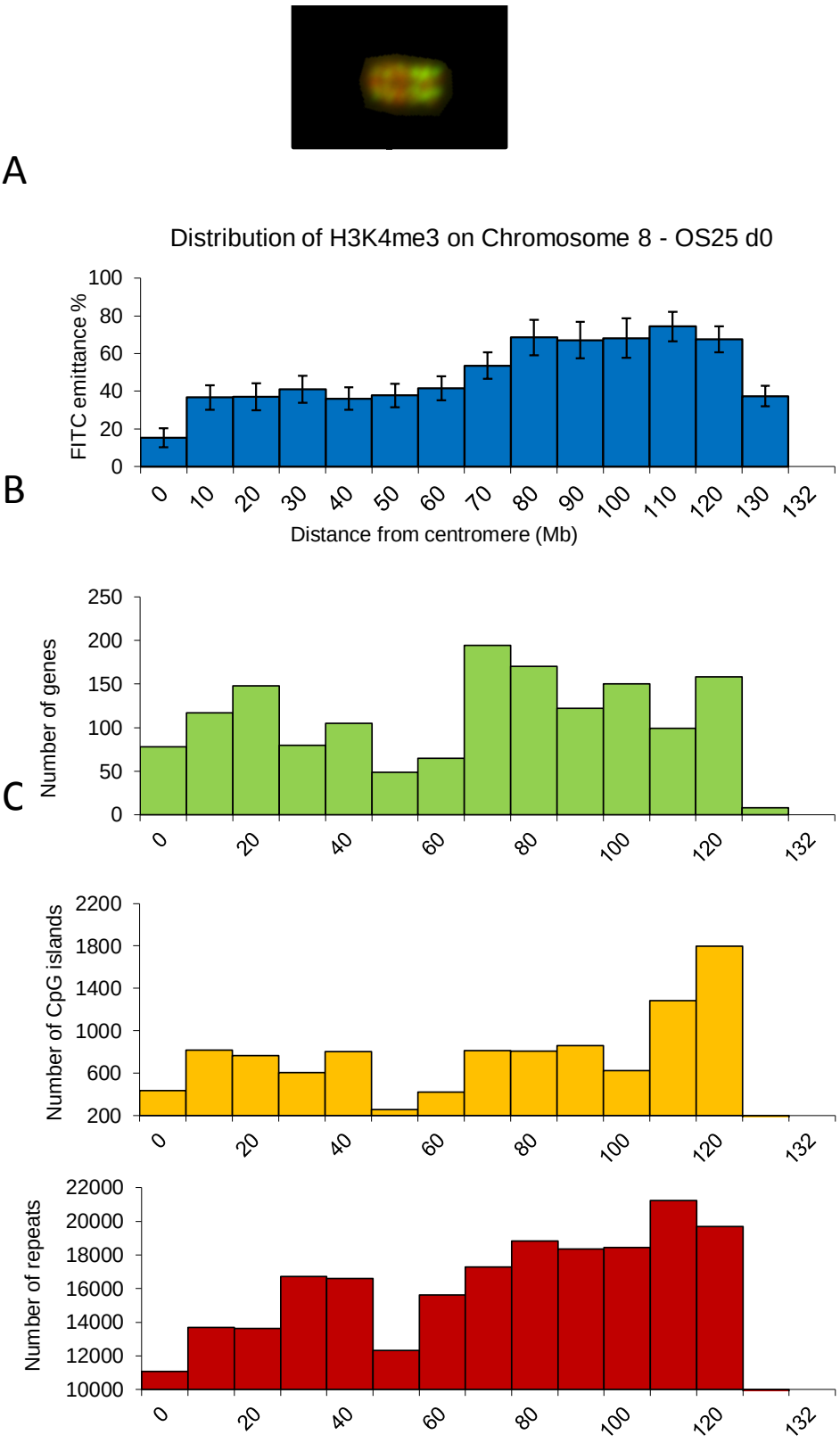




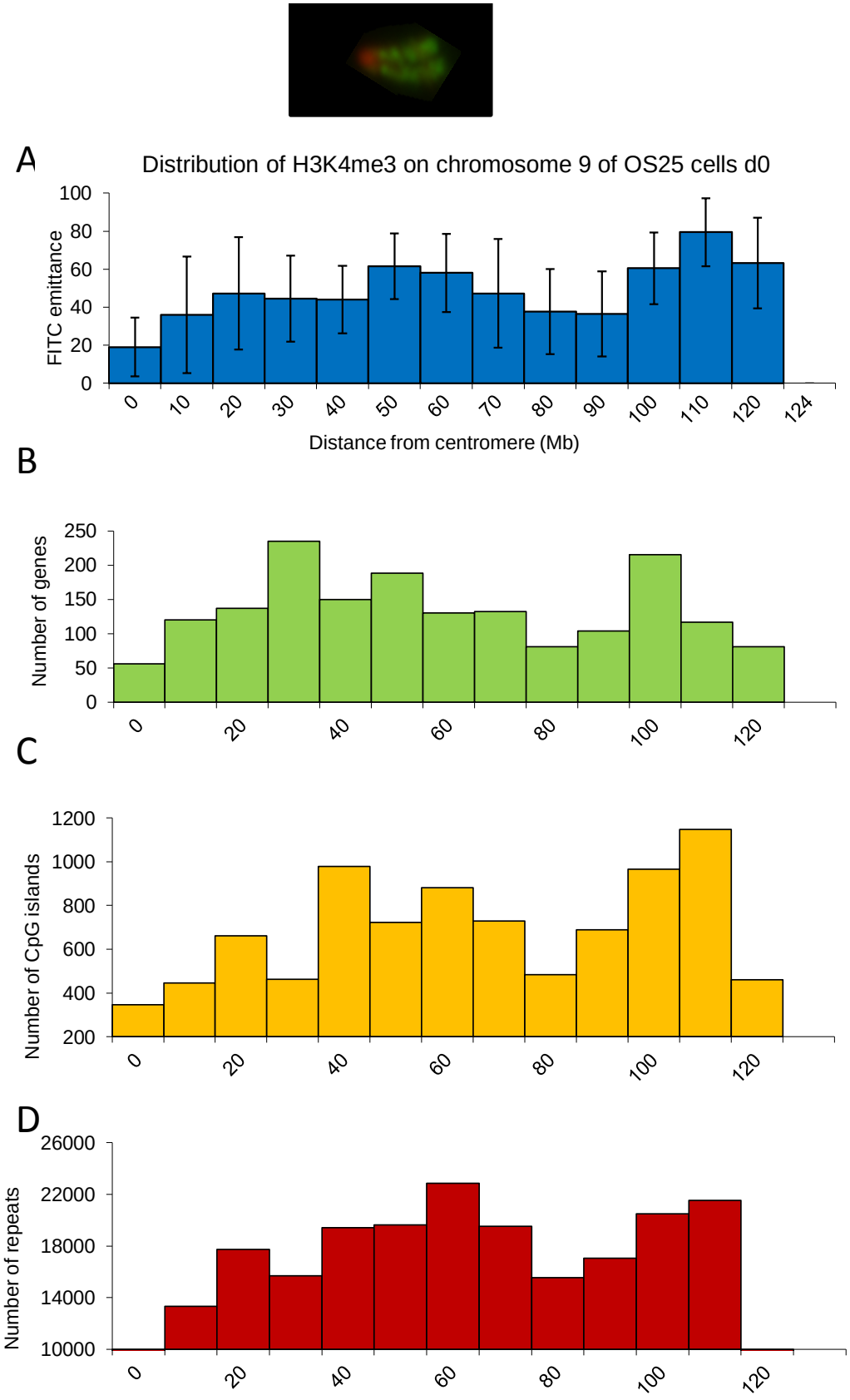
Appendix figure 6 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 6 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



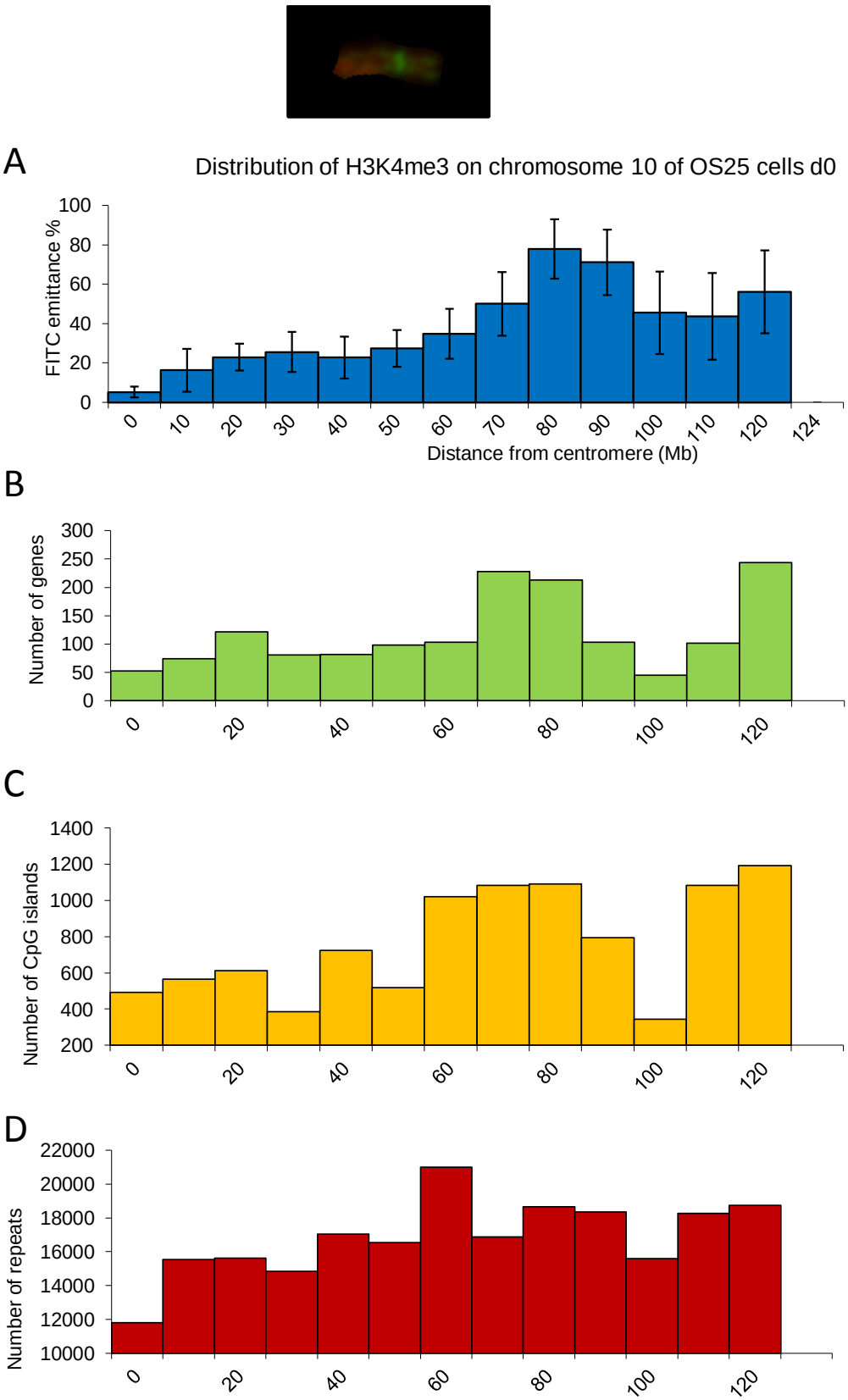
Appendix figure 7 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 7 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



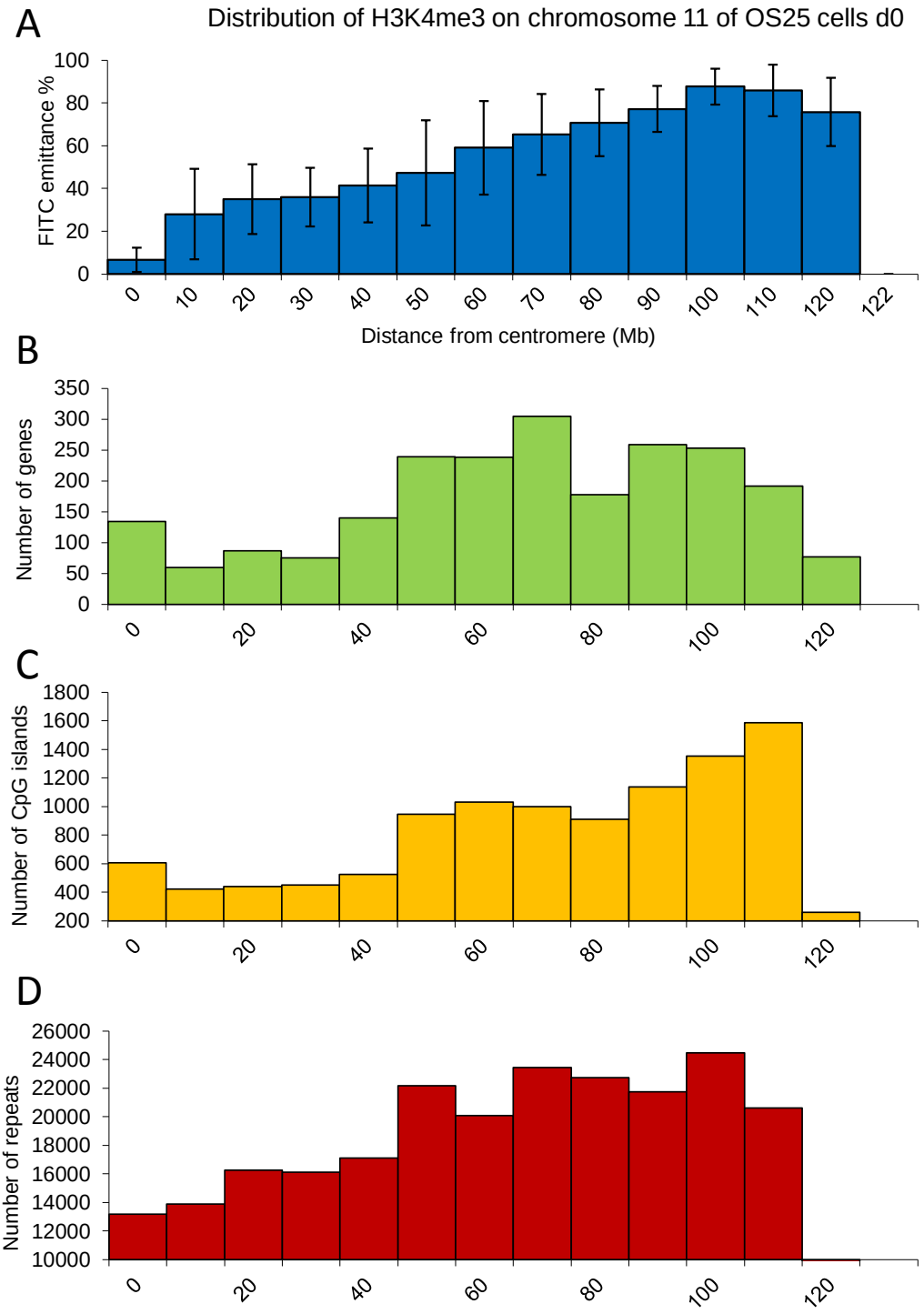
Appendix figure 8 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 8 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



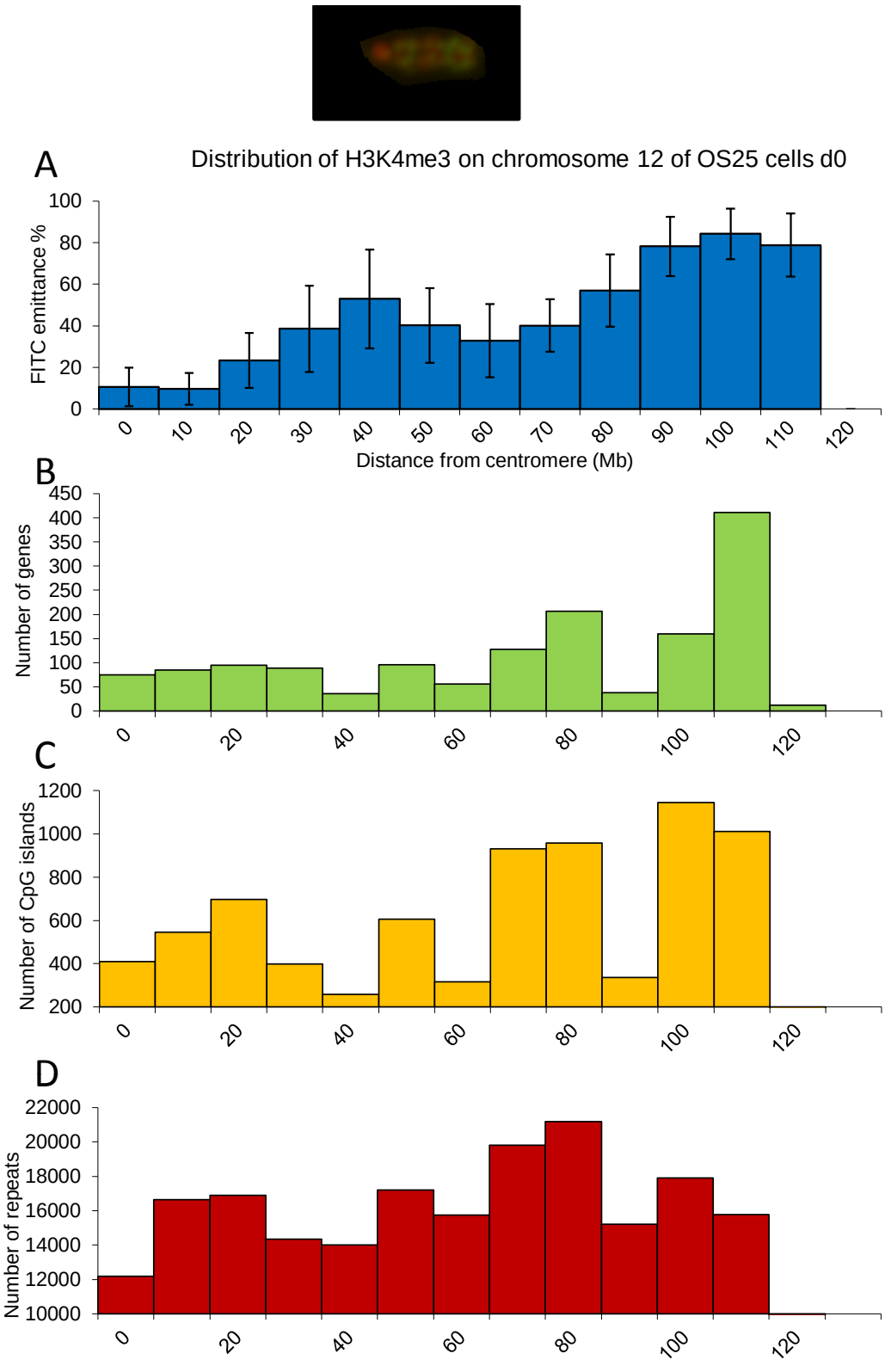
Appendix figure 9 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 9 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



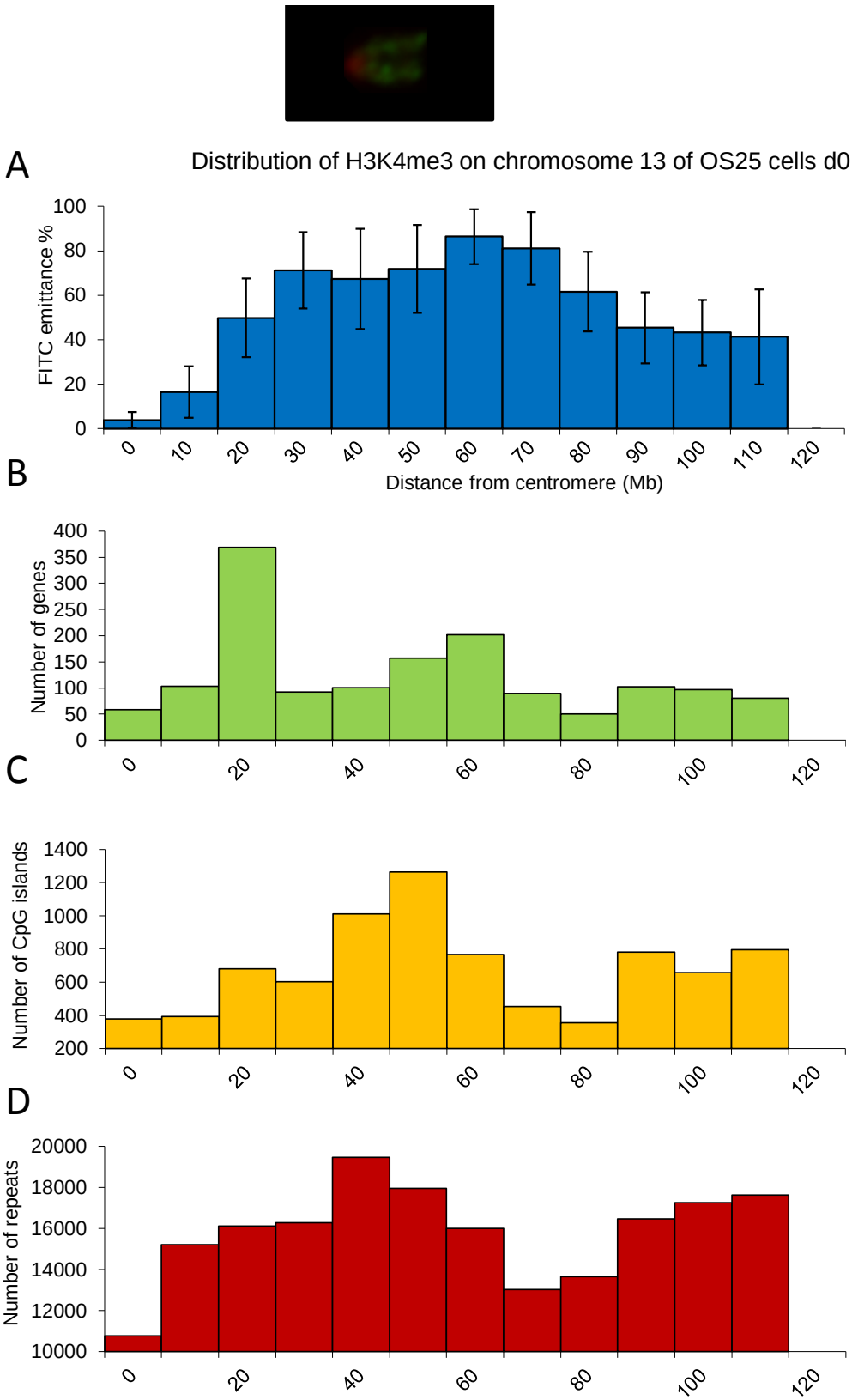
Appendix figure 10 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 10 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



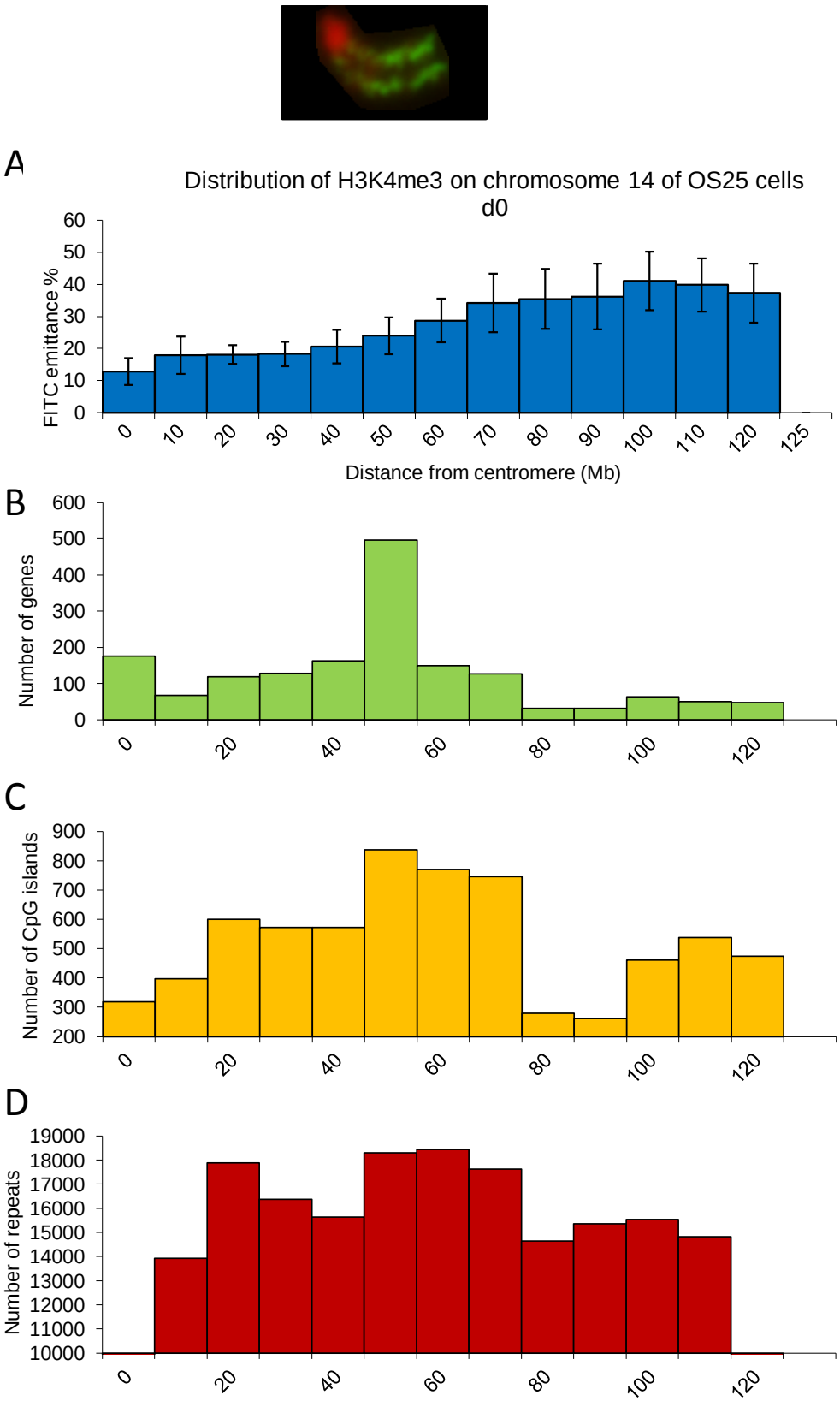
Appendix figure 11 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 11 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



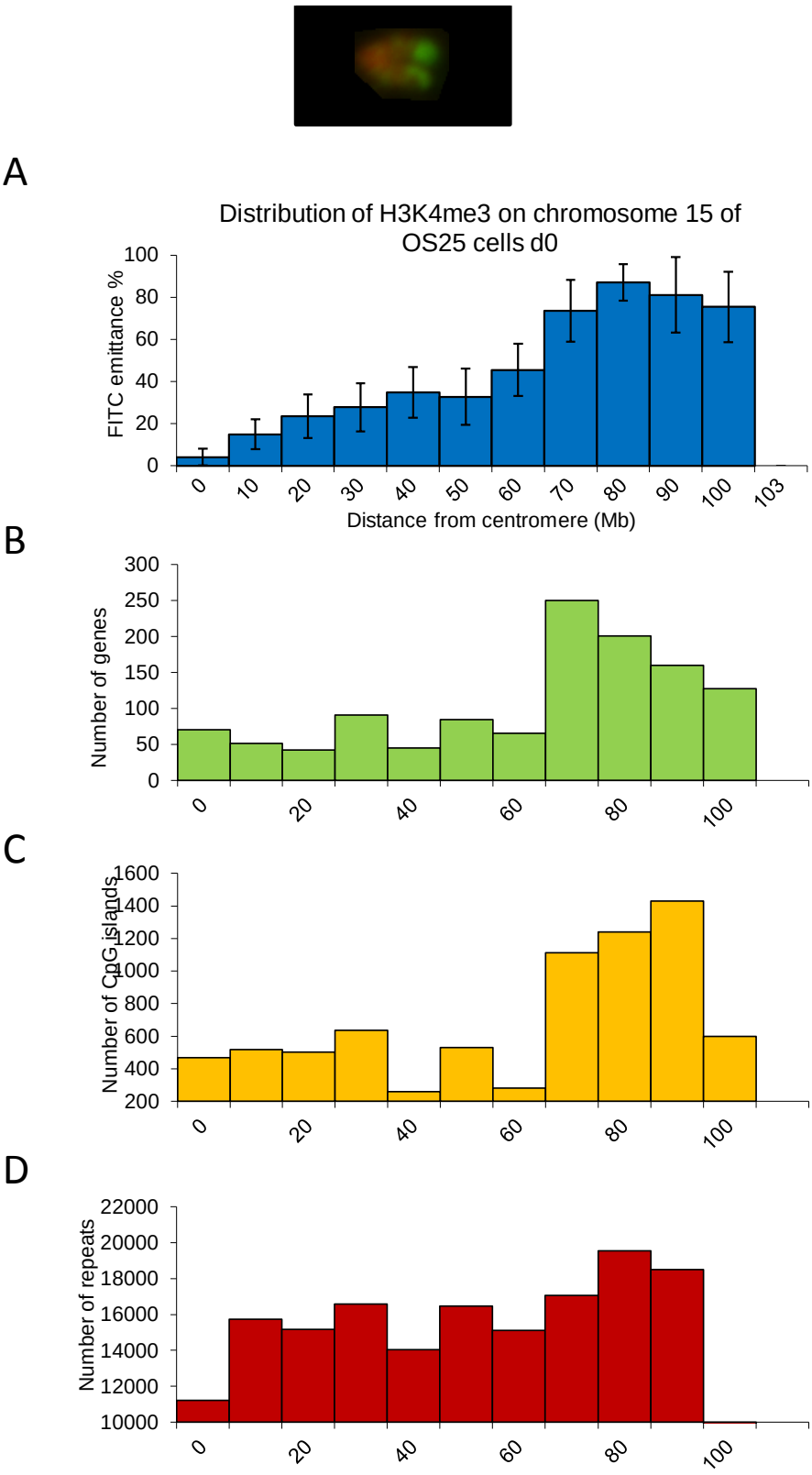
Appendix figure 12 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 12 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



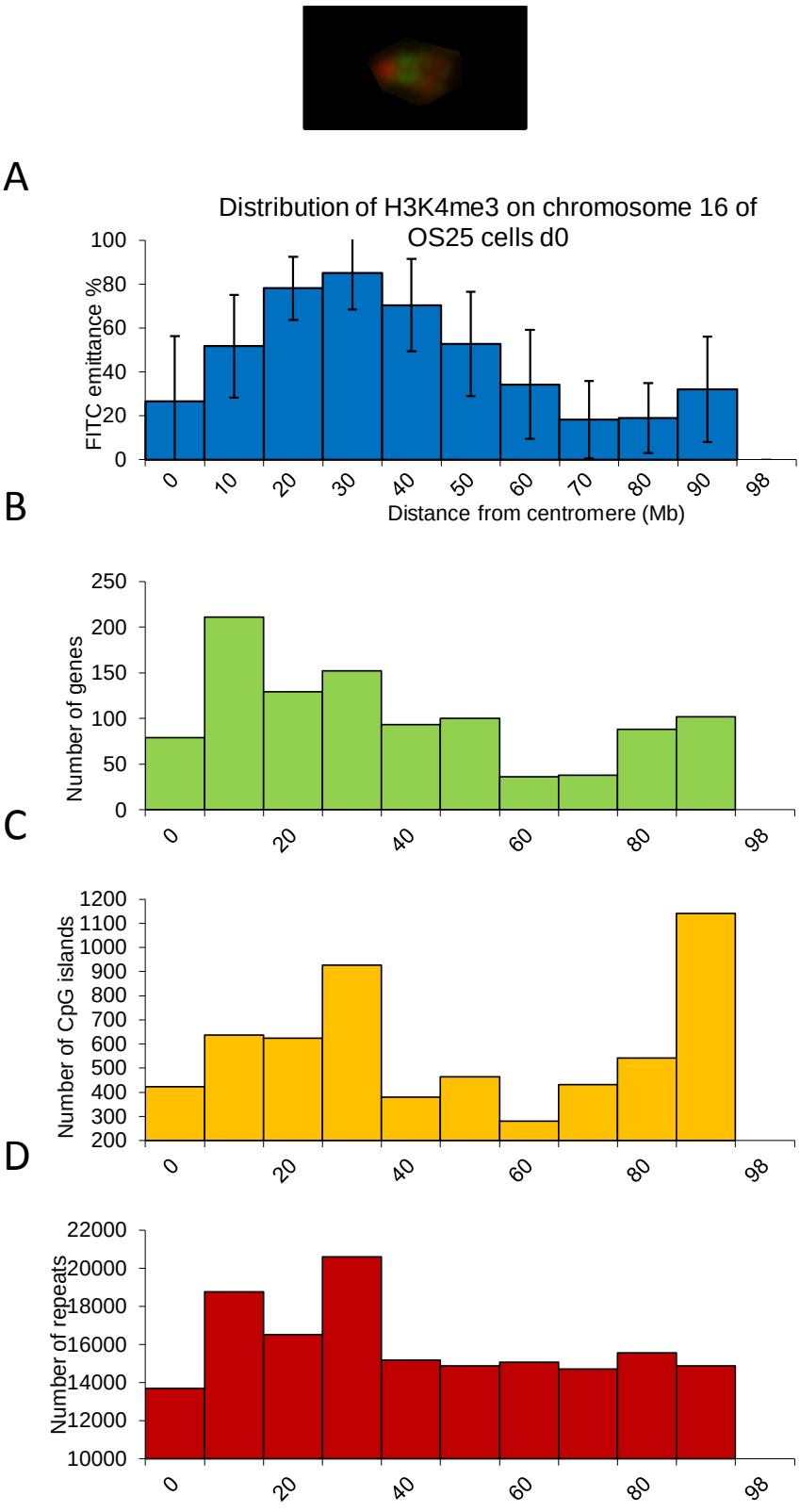
Appendix figure 13 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 13 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



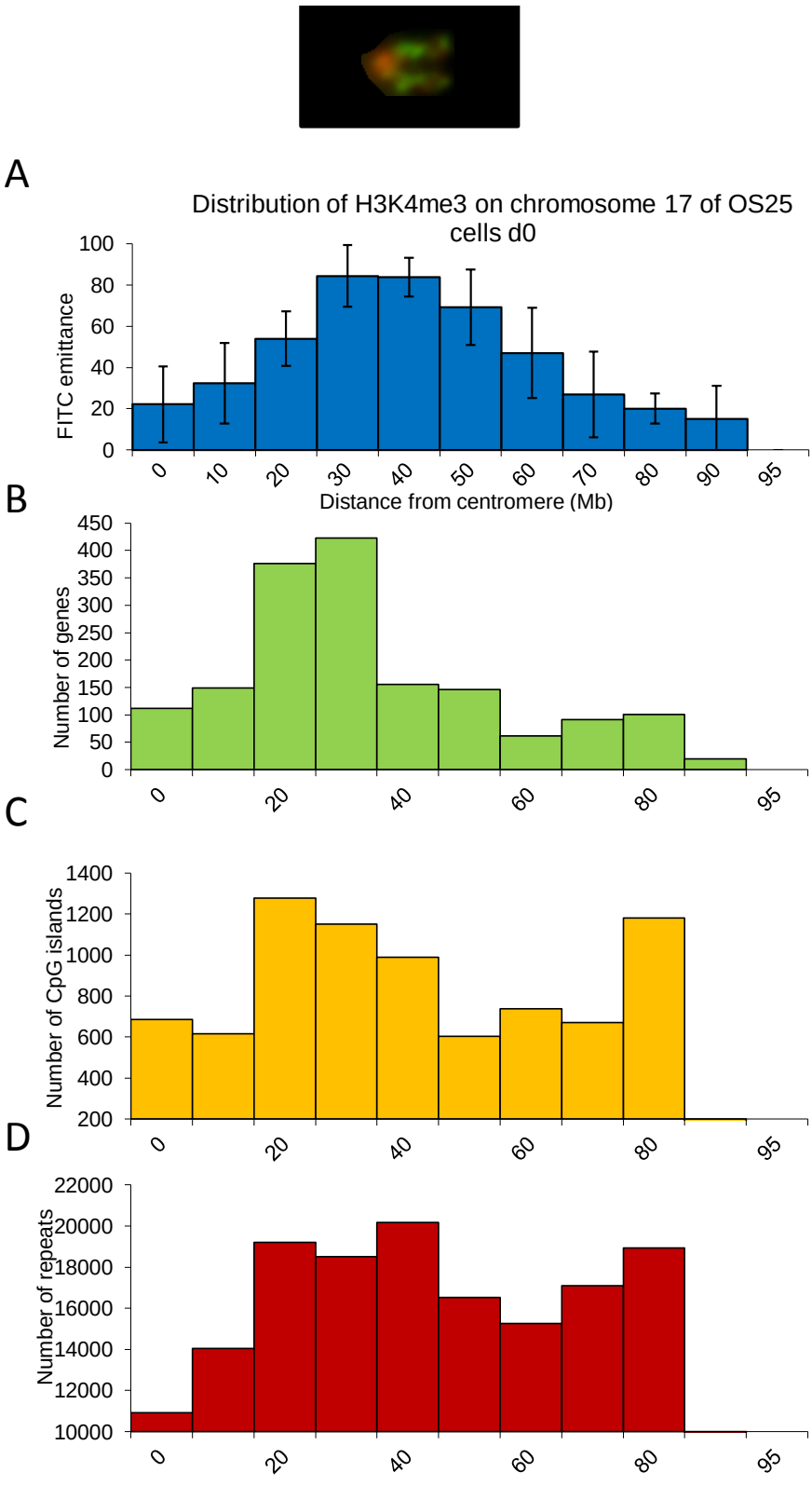
Appendix figure 14 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 14 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



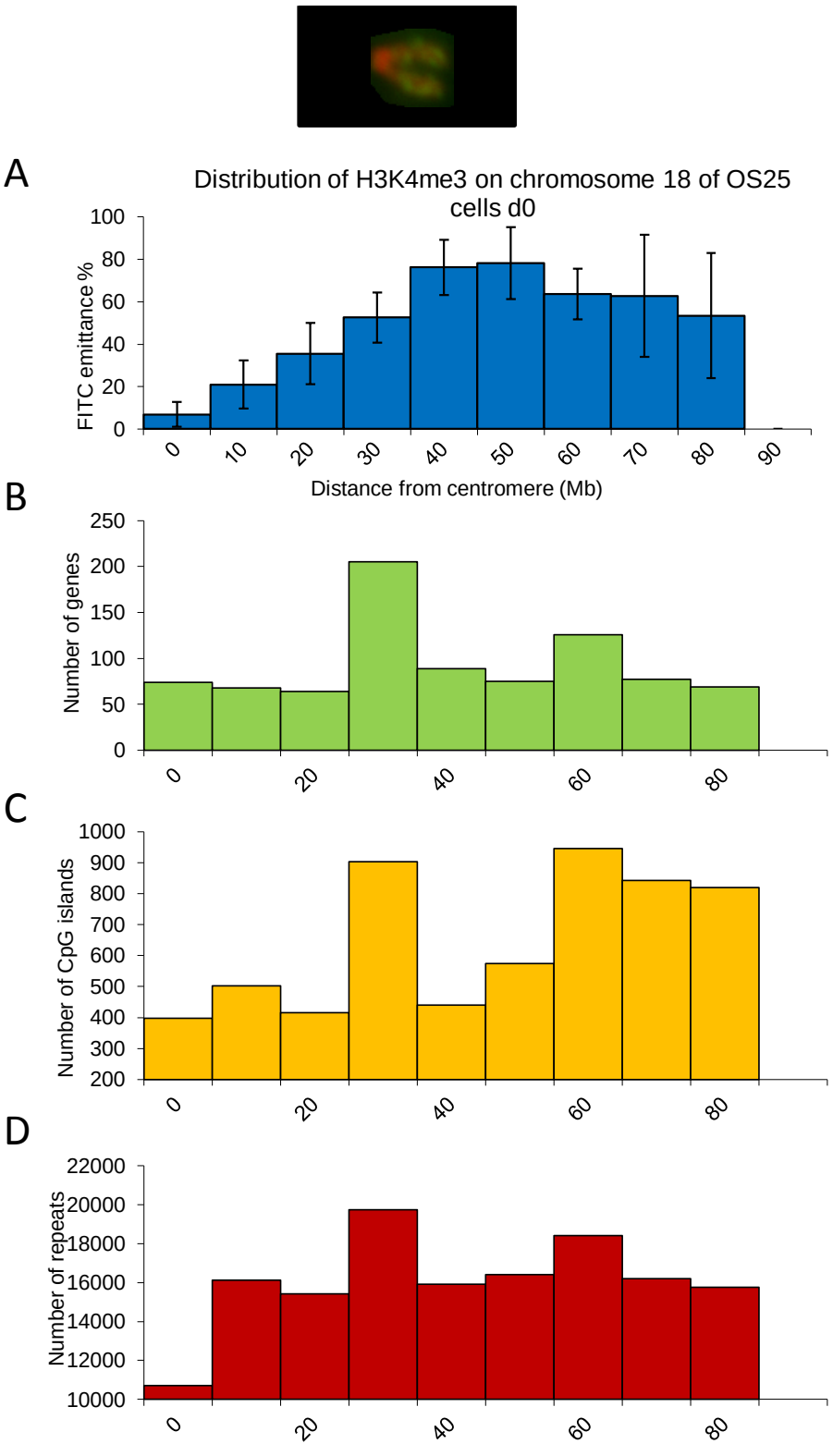
Appendix figure 15 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 15 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



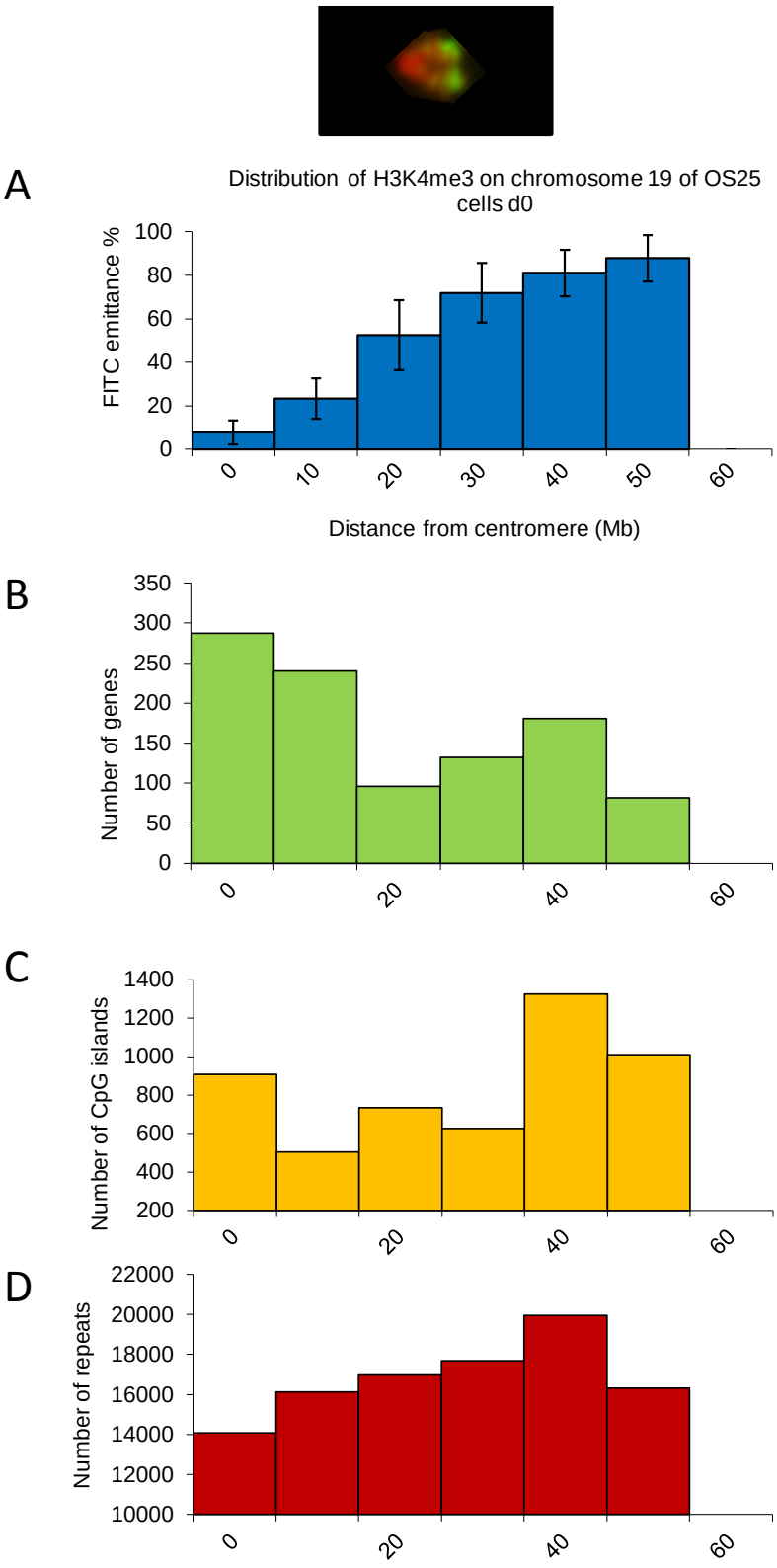
Appendix figure 16 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 16 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



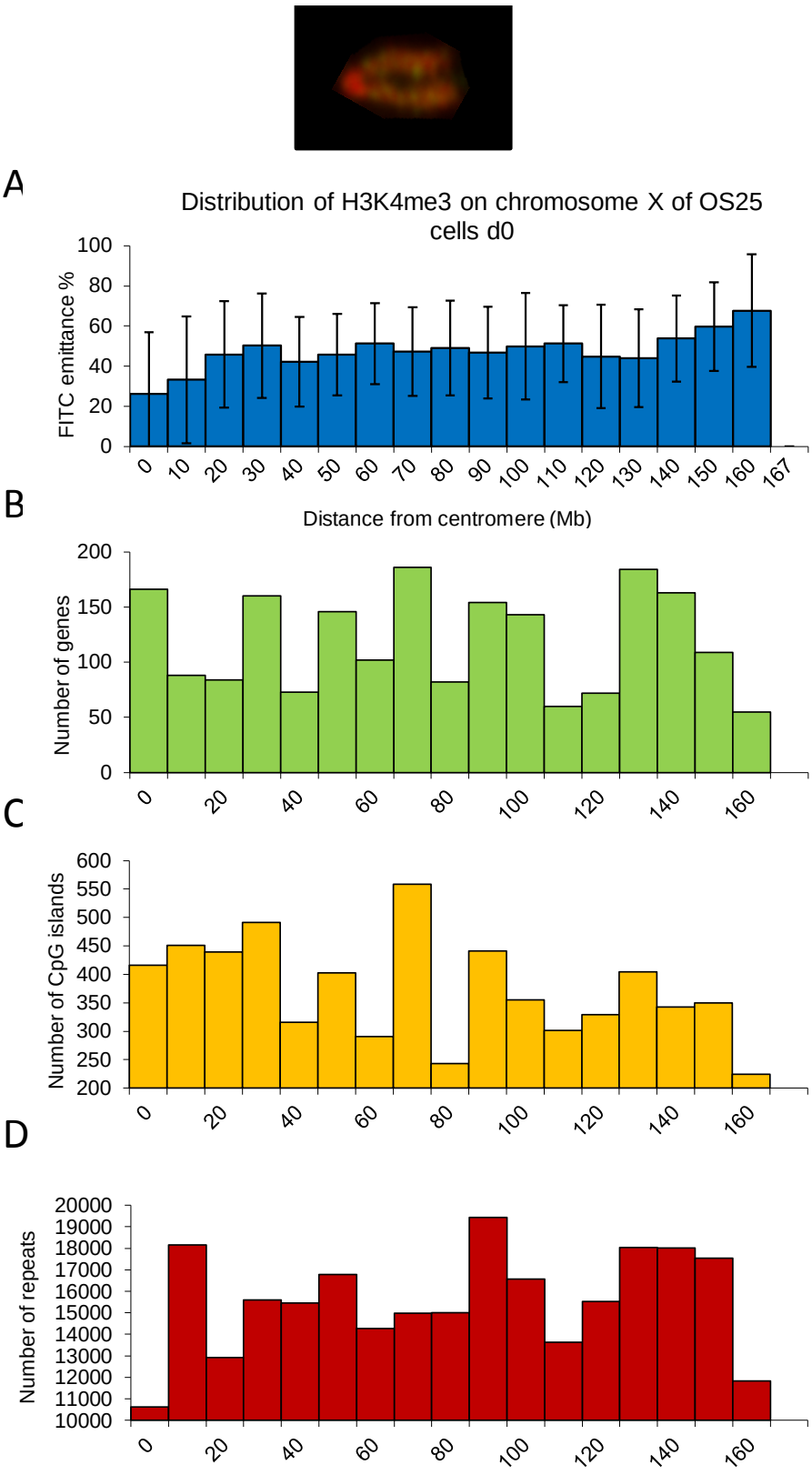
Appendix figure 17 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 17 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



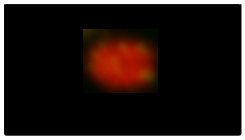
Appendix figure 18 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 18 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



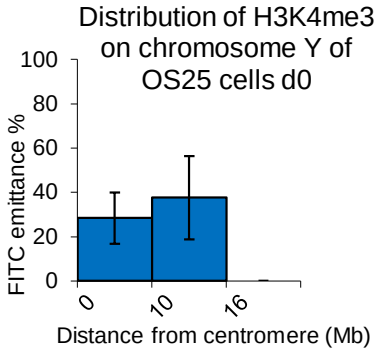
Appendix figure 19 Alignment of **A.** H3K4me3 FITC emittance data from chromosome 19 identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



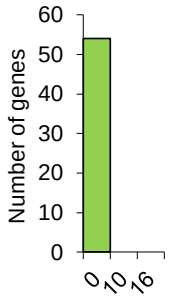
Appendix figure 20 Alignment of **A.** H3K4me3 FITC emittance data from chromosome X identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates



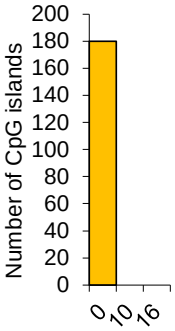
A



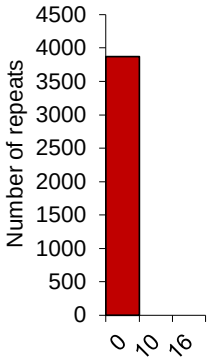
B



C



D



Appendix figure 21 Alignment of **A.** H3K4me3 FITC emittance data from chromosome Y identified by FISH (shown above) with genomic features. Shown here are the NCBI data for **B.** genes, **C.** CpG islands and **D.** repeats in 10 megabase windows. Error bars denote standard deviation between replicates