THE UNIVERSITY *of* EDINBURGH

H4K16 Acetylation during Embryonic Stem Cell Differentiation

Gillian C. A. Taylor

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2012

INSTITUTE OF GENETICS
& MOLECULAR MEDICINE

In Memory of my Dad, Peter Taylor

Declaration

I declare that this thesis has been composed by me, and all the work is my own unless otherwise stated.

Gillian Taylor

November 2012

Acknowledgements

There are many people without whom this thesis would probably have been replaced by a note reading “Gillian has done some science for four years, it was ChIP and some stuff and things please give her a PhD, love Gillian ☺”. Sorry if I forgot anyone!!

Firstly I’d like to thank my supervisor Wendy Bickmore, for letting my do a PhD in her lab and spend what felt to me like a vast amount of money. She has always had unwavering enthusiasm for my project, even when I didn’t! Secondly, thanks to my first day to day supervisor, Ragnhild Eskeland, who instigated the project that became my PhD, and who understands the motivational power of the high five. Finally in the supervisory department, I thank Pradeepa M. Marulasiddappa who inherited me as a student after Ragnhild’s departure (mostly due to proximity) and who does not understand the motivational power of the high five, but who always, always believes that the experiment will work.

Thanks also go to everyone in the Bickmore lab, past and present. You have been a source of inspiration, advice, and vast quantities of cake for the last four years. In particular, I would like to thank: Lady Shelagh Boyle - World Expert on FISH (and afternoon tea), Liz Kerr (I need a hero!), Rob “quick question” Illingworth, Pierre Therizols, Betül Hekimoglu-Balkan and Dr. Chief Iain Williamson, all of whom helped me out with protocols and problems over the years.

I would also like to thank the residents of the sadly demolished office E3.07 – a place any question could be answered in about thirty seconds flat. Several people already mentioned, plus Andrew Wood (a citation device in and of himself), Michael Moffat (a.k.a. the chocolate fairy), John Thompson (bioinformatics help and Mass Effect chat) and Donncha Dunican. You collectively improved my project and cut down on my googling time. Thanks also to everyone in Technical Services at the HGU, for making my life much easier!

I am also very grateful for the support of the students in my year (and others) who have put up with my whinging over the half decade (!!!!) we’ve spent together. The Chromatin Crew (willing or no); Sehrish Rafique, Sam Corless, and Adam Buckle; plus Kerrie Marie, Katy Astell, Xianghua Li, Grainne Gernon, Ewan McNeil, Emily Pritchard, Alisdair Jubb, Jess Woodward, Jenny Huffman, and (sensible non-PhD-doer) Heidi Mjoseng.

Finally, in the non-science category, the people who reminded me that there were more important things in life than failed western blots (such as new episodes of Game of Thrones, America’s Next Top Model, and The Great British Bake Off). My flatmates, particularly: Steph, Steff, Olly, Hannah, Siobhan, and Victoria. Sorry I disappeared for a year – thanks for all the cups of tea! I’m grateful to the Stockbridge branch of Starbucks, for giving me the motivation to get out of bed on many a cold winter’s morning, and Glenogle Swim Centre for delaying my inevitable cake-induced heart attack. Gemma and Helen – look! I have a PhD! Gemma you have a house! Helen, you have a child! Wait…does this mean we’re grownups now? I’m not sure I’m ok with that!! Kelly – grad school sucks, don’t do it!! To my family; thanks for all your support for the past four (or should I say 27…) years. Mum, I never surrounded my cells in imaginary purple light, but thanks for the thought, and for all the lasagne. Stevan, you smell (this fact shall be recorded in the Edinburgh university library for all eternity), but thanks for the amusement, and BC, TMP and M.

Abstract

Eukaryote DNA is organised into the more compact nucleosome by wrapping 147bp of DNA around a histone octamer core. The N-terminal tails of the histones protrude through the DNA and can be modified by a variety of enzymes. Acetylation of Histone 4 Lysine 16 (H4K16ac) is an important modification associated with an increase in transcription, and in flies is an important component of the doseage compensation system. It is also unique amongst histone modifications in that it has been directly associated with chromatin decompaction. H4K16ac has been linked to development through its Histone Acetyltransferase, MOF. Deletion of MOF in mice leads to mass chromatin defects, and embryonic lethality prior to the blastocyst stage.

I set out to understand the role of H4K16ac in differentiating Embryonic Stem cells (ES cells) and chromatin compaction *in vivo*. I generated a ChIP-seq profile for H4K16ac in undifferentiated ES cells, and after 3 days of retinoic acid (RA) differentiation. This revealed an association of H4K16ac with the promoters of transcribed genes in pluripotent ES cells, followed by loss H4K16ac on ES cell specific genes and gain of the modification on differentiation specific genes. There were some silent genes in ES cells, however, which were acetylated on their promoters. Through this study I also found that H4K16ac and MOF mark active enhancers in ES cells, along with H3K4me1 and H3K27Ac and p300. H4K16ac did not mark a known regulatory region in limb cells, and it is possible that it marks active enhancers only of ES cells.

Furthermore, I looked at the compaction state large regions (>100kb) which lost H4K16ac upon differentiation by FISH, to determine if loss of H4K16ac could predict compaction. The regions selected showed no change in compaction state between UD and D3 cells, meaning that loss of H4K16ac does not directly lead to chromatin compaction *in vivo*. However loss of H4K16ac may be necessary for any subsequent compaction, or the change in compaction may take place at nucleosomal level.

Finally, I attempted both to overexpress and reduce the level of MOF in ES cells. I was unable to manipulate the level of MOF in this cell type in either direction; expression of endogenous MOF was silenced after very little time, and stable MOF shRNA cell lines showed no reduction in levels of MOF. Therefore, potentially, dosage of MOF/H4K16ac in this cell type is critical.

This study may help to understand the significance of H4K16ac in ES cell differentiation and chromatin compaction.

Table of Contents

Chapter 1 Introduction

Chapter 2 Materials and Methods

Chapter 3 H4K16 Acetylation is found on active promoters and enhancers in Embryonic Stem Cells

Chapter 4 Profile of H4K16ac changes upon ES cell differentiation

Chapter 5 Loss of H4K16ac does not predict changes in chromatin compaction

Chapter 6 Preliminary investigation into the MOF complex in ES cells

Chapter 7 Discussion and future directions

Bibliograpy

Appendix I

List of Figures and Tables

1 Introduction

2 Materials and Methods

3 H4K16 Acetylation is found on active promoters and enhancers in Embryonic Stem Cells

4 Profile of H4K16ac changes upon ES cell differentiation

5 Loss of H4K16ac does not predict changes in chromatin compaction

6 Preliminary investigation into the MOF complex in ES cells

7 Discussion and Future Directions

Appendix I

Abbreviations

µl: Microlitre

µm: micrometres

14fp: 14 Forelimb Posterior mouse limb cells

293T: Human Embryonic Kidney 293T cells

2D: Two dimensional

A_{260}: Absorbance at 260nm

ac: Acetytion

BAM: Binary Sequence Alignment/Map

BED: Browser Extensible Data

bp: Base Pairs

BSA: Bovine serum albumin

cd: Chromodomain

CEAS: Cis-regulatory Element Annotation System

ChIP: Chromatin immunoprecipitation

ChIP-seq: ChIP sequencing

Cy3: Cyanine 3

d^2: interprobe distance squared

D3: ES cells differentiated by 3 days of treatment with retinoic acid

DAPI: 4',6-diamidino-2-phenylindole

DDR: DNA Damage Repair

DHS: DNase Hypersensitivity Site

DNA: Deoxyribonucleic acid

DNaseI: Deoxyribonuclease I

dNTP: Deoxyribonucleotide triphosphate

DSB: Double Strand Break

EDTA: Ethylenediaminetetraacetic acid

ENCODE: Encyclopedia of DNA Elements

ES cells: Embryonic Stem Cells

FACS: Fluorescence-activated cell sorting

FCS: Foetal Calf Serum

FISH: Fluorescence In-Situ Hybridization

FITC: Fluorescin isothiocyanate

GFP: Green Fluorescent Protein

GMEM: Glasgow Minimum Essential Medium

GO (terms): Gene Ontology

GOrilla: Gene Ontology enRIchment anaLysis and visuaLizAtion tool

GST: Glutathione-S-Transferase

GTE: Glucose, Tris, EDTA

H1, H2A, H2B, H3, H4: Histone 1, 2A, 2B, 3, 4

H4K16ac: H4K16 acetylation

HAT: Histone Acetyl Transferase

HDAC: Histone Deacetyl Transferase

HMT: Histone Methyl Transferase

HRP: Horseradish Peroxidase

IGB: Integrated Genome Browser

IgG: Immunoglobin

IP: Immunoprecipitation

iPS: Induced Pluripotent Stem Cells

IPTG: Isopropyl β-D-1-thiogalactopyranoside

IRES: Internal Ribosome Entry Site

K: Lysine

K^+: Potassium ion

kb: Kilobase

kDa: Kilodalton

LB: Luria-Bertani

LC-MS: Liquid Chromatography Mass Spectrometry

LIF: Leukemia inhibitory factor

MAA: Methanol Acetic Acid

MACS: Model-based Analysis for ChIP-Seq

Mbp: Mega base pairs

me/me1/me2/me3: Methylation, monomethylation, dimethylation, trimethylation

MEF: Mouse Embryonic Fibroblasts

mL: Millilitre

MNase: Micrococcal Nuclease

MOF: Males Absent On the First

MSL: Male Specific Lethal

NAD: Nicotinamide adenine dinucleotide

N-ChIP: Native ChIP

nm: Nanometres

NP-40: Nonidet P-40 (Sigma IGEPAL CA-630)

NPS: Nucleosome Positioning from Sequencing

NSL: Non Specific Lethal

NSL: Non-Specific Lethal

OD: Optical Density

PBS: Phosphate Buffered Saline

Pc: Polycomb

PCR: Polymerase Chain Reaction

PHD: Plant Homeo Domain

Pol II: RNA polymerase II

PTM: Post Translational Modification

PVDF: Polyvinylidene fluoride

PWWP: Pro-Trp-Trp-Pro Domain

Q: Glutamine

qRT-PCR: Quantitative Real Time PCR

R: Arginine

r^2: nuclear radius squared

RA: Retinoic Acid

RFP: Red Fluorescent Protein

RIN: RNA Integrity Number

RIPA: Radioimmunoprecipitation assay buffer

RNA: Ribonucleic acid

RPM: Reads Per Million

RT-PCR: Reverse Transcriptase PCR

SAM: Sequence Alignment/Map

SDS: Sodium dodecyl sulphate

SDS-PAGE: SDS Polyacrylamide gel electrophoresis

SICER: Spatial clustering approach for the Identification of ChIP Enriched Regions

SILAC: Stable Isotope Labelling by Amino Acids in Cell Culture

SMART: Simple Modular Architecture Research Tool

SSC: Saline Sodium Citrate

TEMED: Tetramethylethylenediamine

TES: Transcription End Site

TF: Transcription Factor

TSA: Trichostatin A

TSS: Transcription Start Site

UTR: Untranslated Region

WCB Edinburgh: Wellcome Trust Centre for Cell Biology

X-ChIP: Cross linked ChIP

zf: Zinc finger

Chapter 1: Introduction

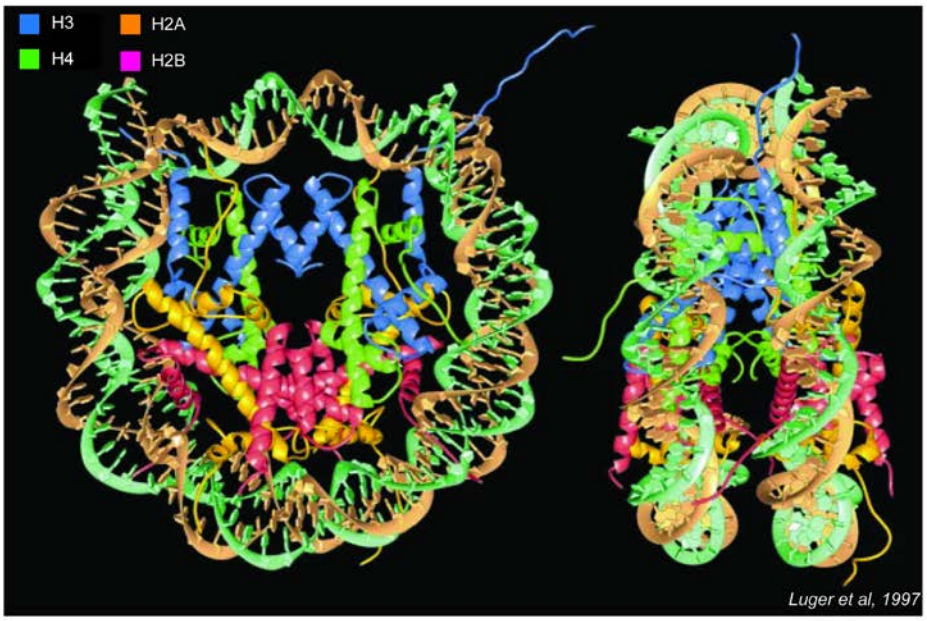
1.1 Chromatin Structure and Transcription

Chromatin structure is intrinsically linked to control of gene expression. For example; broadly speaking, open regions of the chromatin (euchromatin) are conducive to active transcription, whilst closed chromatin (heterochromatin) provides a more repressive environment and genes which reside here are transcriptionally silenced (Li et al., 2007). The differences between heterochromatin and euchromatin are not only revealed in regulation of transcription, but also DNA repair and replication. For example, experiments have indicated that chromatin remodellers may be of particular importance for replication of DNA through densely packed heterochromatin (Collins et al., 2002).

Although broadly, heterochromatin is silenced, transcription still occurs on some level in these regions (Grewal and Elgin, 2007), and by no means all are genes within euchromatin are transcriptionally active. DNA derived from more open/disordered fragments of the genome (separated by sucrose gradient) hybridize to more gene rich regions of the genome than do more closed fragments, though here the relationship to transcription is less clear (Gilbert et al., 2004). On a gene by gene level, changes in chromatin compaction state have been shown to correlate with transcription; at different stages of differentiation, globin genes are more accessible both to RNA polymerase II and DNAse I in chromatin isolated from cell types in which they are active (red blood cells) than cell types in which they are not expressed (liver cells) (Weintraub and Groudine, 1976). Additionally, binding of a potent transcriptional activator, VP16, leads to decompaction of heterochromatin (Belmont et al., 1999).

In eukaryotes, the nucleosome is the fundamental subunit of chromatin, and consists approximately 165bp of DNA wrapped around a histone octamer core (Kornberg, 1974; Kornberg and Thomas, 1974; Luger et al., 1997) (Fig1.1A). Inclusion of the linker histone (e.g. H1) forms the chromatosome. The DNA is wound in nearly two superhelical turns around the core; from free DNA, this results in a structure five to ten times more compact. This primary 10nm fibre structure may then be compacted further into a secondary 30nm structure via internucleosome interactions, and beyond, intro tertiary structures which are potentially formed through long distance interactions of secondary structures. There is considerable work to be done before the specifics of these higher order structures are fully understood (Woodcock and Dimitrov, 2001) (Fig1.1B).

A



B

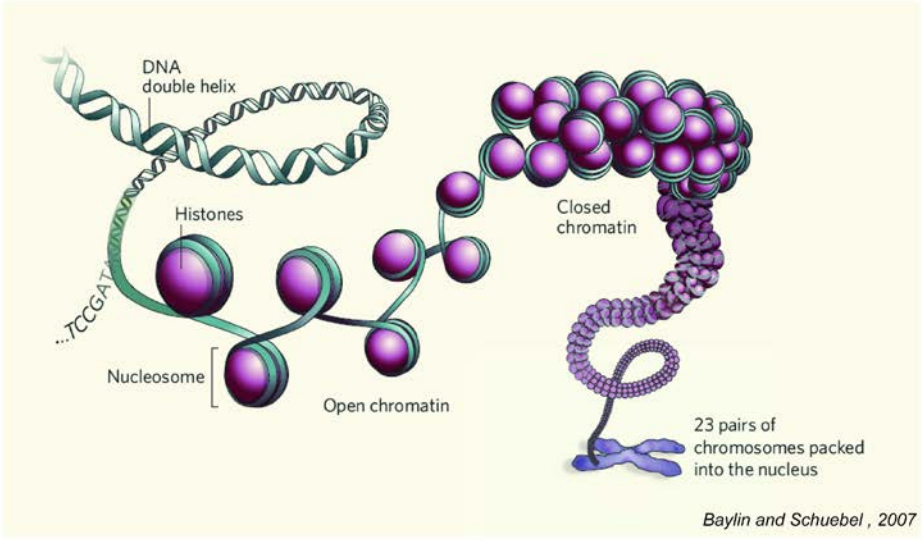


Figure 1.1 *Models of chromatin structure*

A: Crystal structure of the nucleosome core particle (Luger et al., 1997). Green and orange ribbon traces represent DNA, histones represented by colours as in key. Histone tails protrude from the core particle, beyond the DNA. B: A model showing potential levels of genome compaction from free DNA through to highly compacted mitotic chromosomes. Adapted from (Baylin and Schuebel, 2007).

The nucleosome structure, in and of itself, is not conducive to transcription. It provides a barrier to the transcriptional machinery; assembled nucleosomes *in vitro* do not support transcription initiation, and have a negative effect on elongation (Williamson and Felsenfeld, 1978; Knezetic and Luse, 1986). However, though the presence of nucleosomes over the promoter or transcription factor binding sites can obstruct the initiation of transcription *in vitro*, once initiation has taken place, the polymerase can continue through the nucleosome, displacing it (Lorch et al., 1987; O'Neill et al., 1992). Promoters *in vivo* are generally sensitive to nuclease digestion (Elgin, 1981; Sabo et al., 2004), and genes usually contain a nucleosome free region at the start (with two well positioned nucleosomes either side) (reviewed in (Jiang and Pugh, 2009)).

Chromatin packaging has a profound effect on transcription (and other DNA dependent processes). Therefore, systems exist both to encourage compaction of chromatin in areas of the genome which are required to be silenced, and to encourage decompaction in areas which are to be transcribed.

One set of targets for manipulation of the chromatin fibre are the core histone tails, which protrude through the DNA and provide a substrate for a variety of modifications, such as methylation, ubiquitination, phosphorylation, and most importantly for this work, acetylation. The histone core may also be modified, but these Post Translational Modifications (PTMs) have not been as widely studied as those on the N-terminal tails (Mersfelder and Parthun, 2006).

1.2 Histone Modifications

Soon after the discovery of histone tail modifications, it was postulated that they might have some effect on transcription. It was soon determined that acetylation of histones relieves the inhibition of RNA synthesis by nucleosomes (Allfrey et al., 1964). Broadly, post translational histone modifications fall into one of two categories; those which are associated with open, transcriptionally active chromatin, and those which are associated with more closed, transcriptionally silent chromatin (Fig 1.2).

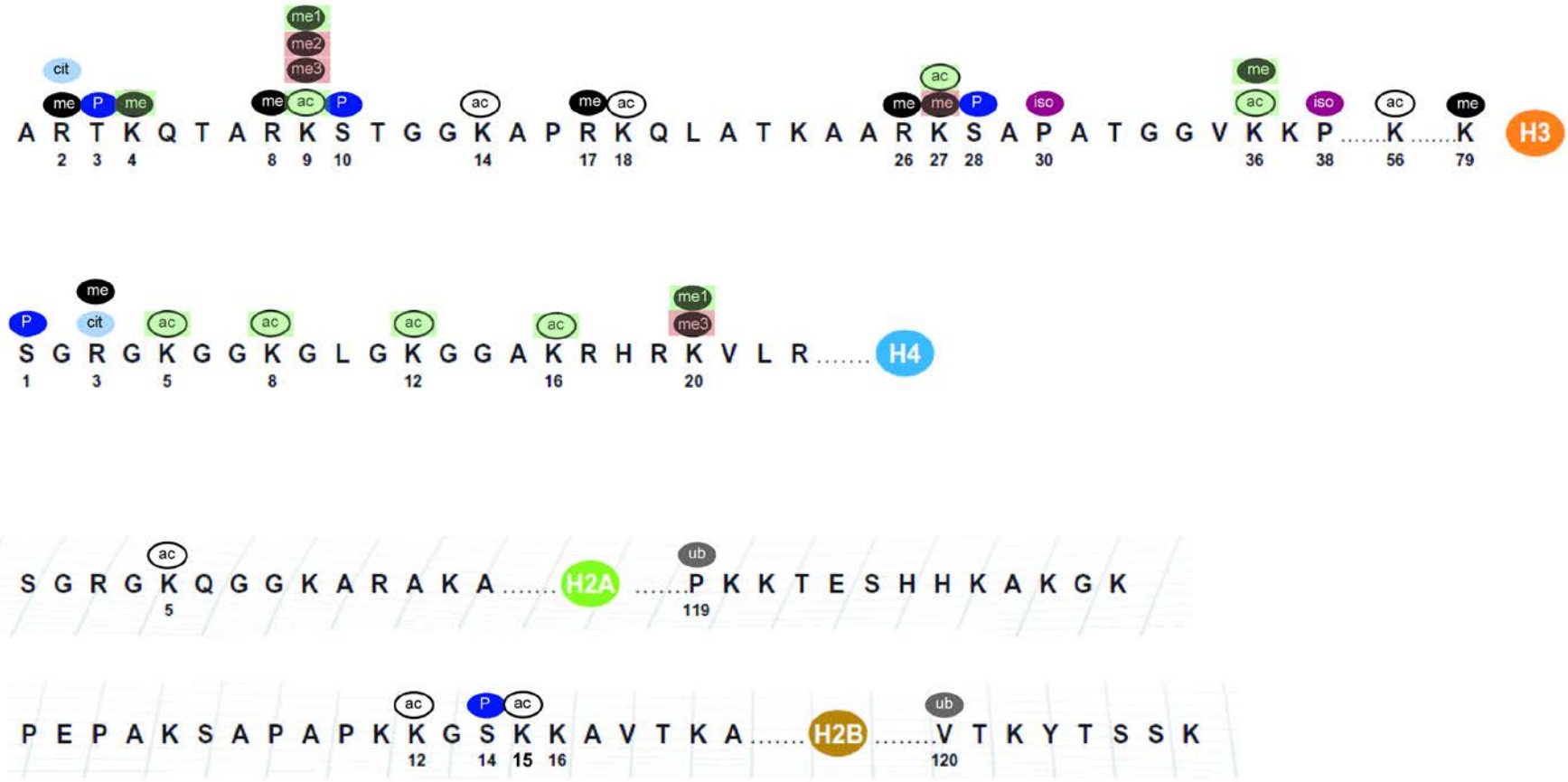


Figure 1.2 *Histone tail modifications.*

Known histone tail post translational modifications; for a selection of relevant modification, the association with active or inactive genes is indicated with a green or red box respectively. Modified from (Andy Bannister, 2012), using data from (Wang et al., 2008).

Hypoacetylation, and trimethylation of histone 3 lysine 27 (H3K27me3), H4K20me3, H3K9 methylation are all associated with silenced heterochromatin, whilst acetylation, H3K4me3, and H3K36me3 are examples of histone PTMs associated with active genes (Wang et al., 2008). The associations of the specific histone modifications with different types of chromatin are conserved across species. Indeed, the locations of specific histone modifications which are conserved between cell types in the same organism are conserved between species (in mammals) (Woo and Li, 2012).

There are two possible modes of action through which a particular histone modification can influence transcription. They may have an intrinsic effect, that is, the presence of the histone modification in and of itself produces an effect on nucleosome structure. Alternatively, a PTM may affect how a nucleosome interacts with its near (or potentially far) neighbours, since formation of a more compacted structure requires that a nucleosome come into closer contact with others nearby. The histone tails have been shown to have an effect on nucleosome oligomerisation *in vitro* (Gordon et al., 2005).

In vitro, H4K16ac is an example of a single histone PTM which has an intrinsic effect on chromatin structure (to be discussed later – section 1.3). More broadly, histone acetylation and phosphorylation both reduce the net positive electrostatic charge of the core histone. Acetylation has this effect by neutralising a positively charged lysine residue, and phosphorylation by addition of a negatively charged phosphate group. This theoretically reduces the attraction of the histone core to negatively charged DNA, causing chromatin decompaction. An increase in chromatin decompaction is a demonstrated effect of a global increase in histone acetylation induced by incubation with the Histone Deacetylase (HDAC) inhibitor TSA (Tóth et al., 2004; Eskeland et al., 2010). The effect of histone phosphorylation on chromatin condensation is less well studied, but phosphorylation of H1 is proposed to decondense chromatin at the replication fork in S phase (Alexandrow and Hamlin, 2005).

Histone modifications can provide a binding site for proteins such as transcription factors or recruiters of chromatin remodelling complexes. For example, the chromodomain of HP1 is known to recognise and bind to methylated H3K9 (Bannister et al., 2001). PWWP domains can recognise and bind to H3K36me3 (Dhayalan et al., 2010) and a variety of chromatin domains can bind acetyl lysine, such as bromodomains (reviewed in (Zeng and Zhou, 2002)), and PHD fingers (Lange et al., 2008). PHD fingers are also capable of binding to methylated lysine (Wysocka et al., 2006).

Histone PTMs can also block protein binding, as in the example of the INHAT (inhibitor of acetyl transferases) protein, which binds to the unmodified histone H3 tail, but whose binding can be abrogated by a variety of histone modifications, including histone acetylation or another overall charge reducing modification, phosphorylation (Schneider et al., 2004).

Downstream effects of binding proteins to histone PTMs can be further complicated by the fact that they are not always direct. For example, serum stimulation leads to increased transcription of the FOSL1 gene in human cell lines through the following pathway; initially, PIM1 kinase phosphorylates H3S10 on the gene's enhancer. This generates a binding platform for the 14-3-3 protein, which then recruits MOF (the specific histone acetyl transferase, or HAT, for H4K16ac), leading to H4K16ac, and then binding of the bromodomain protein BRD4, which subsequently recruits P-TEFb, a transcription elongation factor which facilitates release of the promoter-proximal paused polymerase and finally, transcription of the gene (Zippo et al., 2009). This system is also an example of a combinatorial pattern of histone modifications providing a binding platform for a single protein; *in vitro*, nucleosomes containing H3K9Ac (which is already acetylated prior to serum stimulation in the *in vivo* example), H3S10ph, and H4K16ac induced binding of P-TEFb/BRD4. There is also the example of the negative effect had by H3S10ph on binding of HP1 to H3K9me3 peptides. *In vivo*, loss of H3S10ph allows HP1 to remain bound to mitotic chromosomes, from where it is usually removed (Fischle et al., 2005). A genome wide study of a variety of histone acetylations and methylations showed that there are a group of 17 modifications which tend to coexist, though the functional significance or necessity to their coexistence remains to be elucidated (Wang et al., 2008).

All histone modifications discovered to date are reversible. This flexible control makes them ideal for situations where plasticity in gene expression is required, such as development. Histone methylation and acetylation are of particular relevance to this study and are discussed below in greater detail.

1.2.1 Histone Methylation

Compared with acetylation, methylations of histone tails have been thought of as rather stable marks. Both arginine and lysine residues can be methylated. Lysines can carry mono-di or tri-methylation, resulting in four different states for each lysine, and the potential for different readouts by, for example, binding of different effector molecules.

Methylation of histone tail lysines is carried out by SET domain containing Histone Lysine Methyltransferases (HKMTs) such as Ezh2, which is found within the PRC2 complex (Cao et al., 2002). Product specificity (i.e. whether the product is mono, di, or trimethylated lysine) of SET domain proteins can be defined by variations in the SET domain active site (Zhang et al., 2003). Substrate specificity can vary; in the example of Ezh2, its specificity is altered by interaction with a variety of Eed isoforms (Kuzmichev et al., 2004).

It was only recently that lysine methylation was shown to be reversible, with the discovery of the first histone demethylase, LSD1 (Shi et al., 2004). Another distinguishing feature of histone methylations is a variety of features between the different modified residues; for example though generally considered stable, lysine methylations have a different turnover rate, so some are more stable than others (Zee et al., 2010).

In contrast to histone acetylation (discussed below) which is generally associated with active transcription and open chromatin; methylation of histone tails can correlate with either activation or silencing, depending on which residue is modified (Fig 1.3). For example, along the H3 tail the residues are functionally diverse; H3K4me3 is broadly correlated with the promoters of active genes(Wang et al., 2008), and H3K36me3 with elongation of transcription (Krogan et al., 2003), whilst H3K27me3 is important for gene silencing via the polycomb repressive complexes (Cao et al., 2002). H3K9 methylation has been associated with formation of heterochromatin (Fischle et al., 2005). As expected from such a variety of different functions, the mechanism of action for histone methylation is not fully understood, but the diversity makes them powerful mediators of regulation.

Unlike lysine acetylation, lysine methylation does not appear to have a direct effect on chromatin structure, nor does it affect the charge of the lysine, which remains cationic (Fig 1.4B). Instead, the most studied regulatory effects of lysine methylation are mediated through proteins which contain a methyl-binding domain such as a chromodomain, PWWP domain or WD40 repeats. Again, although broadly the domains' binding is affected by methylated lysine, there is variety. Some are prevented from binding lysine by methylation of the residue; for example, the PHD domain of TRIM24, (a multifunctional protein, overexpression of which correlates with poor prognosis in breast cancer patients) binds to only unmethylated H3K4. This leads to co-activation of the oestrogen receptor alpha (ERα), and expression of the target genes which are generally associated with cellular proliferation (Tsai et al., 2010). Other binding proteins lead to effects on transcription by binding to methylated lysines. For example, the PHD domain containing proteins are known to bind to H3K4me3. BPTF contains two PHD domains, and is the biggest subunit of the ATP

dependent chromatin remodelling complex NURF. BPTF binding to chromatin is dependent on H3K4me3, and Xenopus mutants which lack H3K4me3 show the same developmental defects as BPTF mutants (reviewed in (Taverna et al., 2007).

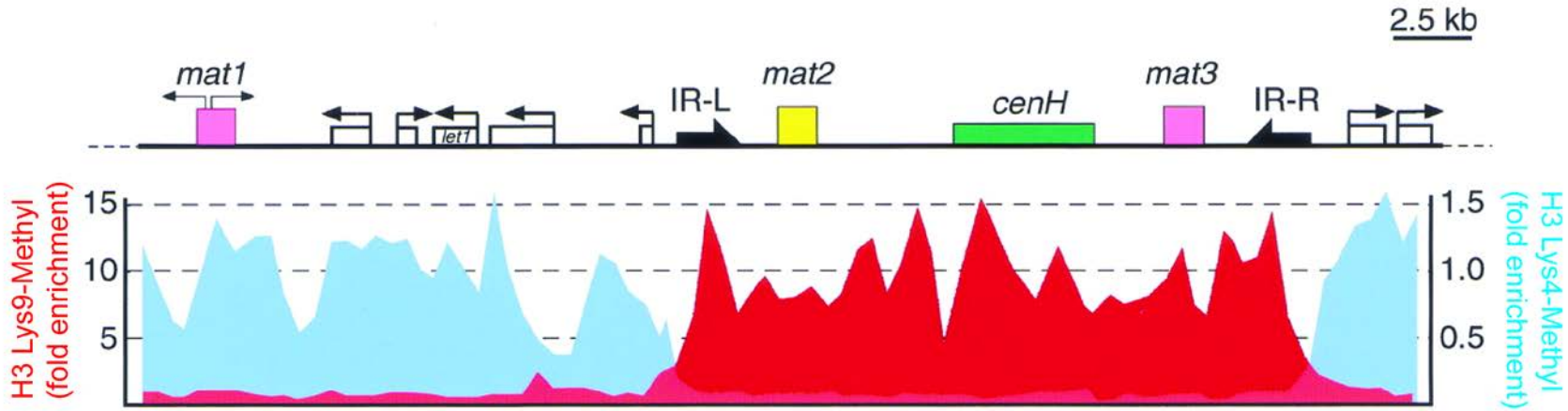


Figure 1.3 *Different H3 methylation marks can associate with euchromatin or heterochromatin*

Euchromatic and heterochromatin regions of the *S. pombe* mating locus contain different H3 methylation marks. Open reading frames are shown with arrows indicating direction of transcription. Region between IR-L and IR-R (inverted repeats) is heterochromeric. Graph represents ChIP experiments for H3K9me (red), and H3K4me (blue). From (Zhang and Reinberg, 2001).

1.2.2 Histone Acetylation

The various histone modifications in this category are frequently counted together under the umbrella term of histone acetylation; though not all have the same function, some of the lysine acetylations have a coordinated function and an additive effect.

Gene activity and histone acetylation were first positively correlated in the 1960s when it was discovered that nucleosomes with acetylated histones show less inhibition of RNA polymerase, and that 'activated' lymphocytes incorporated a much larger amount of acetate into their histones than did inactivated cells (Allfrey et al., 1964; Pogo et al., 1966). This correlated with a large increase in transcription in the activated cell type. Subsequent investigations showed directly that histone acetylation was present on transcriptionally active genes (Hebbes et al., 1988; Wang et al., 2009).

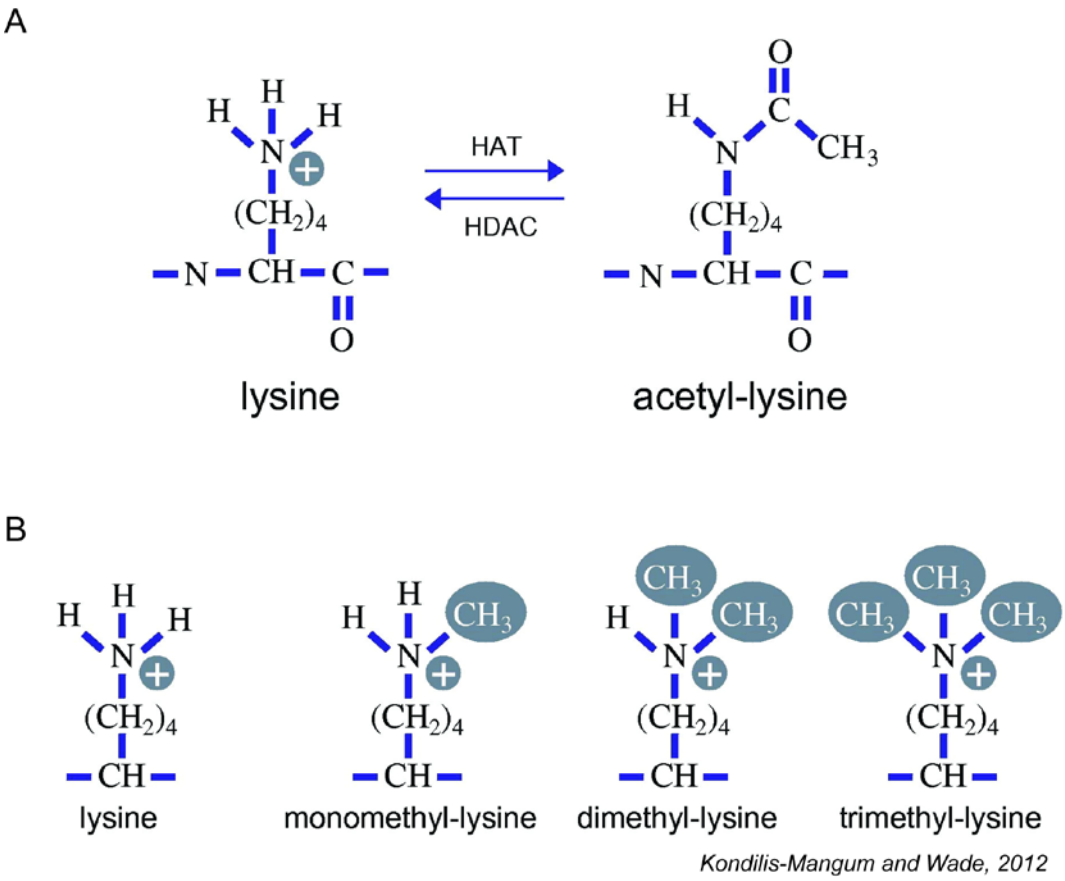


Figure 1.4 *Covalent modification of lysine*

A: ε-amino group of lysine acetylated by Histone Acetyl Transferase (HAT), resulting in loss of positive charge. Reverse reaction is catalysed by Histone DeACetlyases (HDAC).

B: ε-amino group of lysine modified by addition of 1-3 methyl groups, without resulting in loss of the positive charge. Image modified from (Kondilis-Mangum and Wade, 2012)

There are 12 potential sites for histone acetylation on the histone tails (five on H3, four on H4, two on H2B and just one on H2A – Figure 1.2) All are lysine acetylations, which is important given that lysine is a positively charged amino acid, and acetylation neutralises the charge. This means that the net positive charge of the histone tail is reduced, reducing its attraction to the DNA (Fig 1.4A). Since the histone tails lie outside the DNA, their binding to the DNA could obscure transcription factor binding sites and thus repress transcription; a process which would be reduced by acetylation. An early report suggested that on nucleosomes assembled on 5S rDNA *in vitro*, acetylation of the histones, or removal of the histone tails, allows binding of the transcription factor TFIIIA, where unacetylated histones do not (Lee et al., 1993). Acetylations on H2A and H2B are poorly studied, while those of H3 and H4 are widely studied. There is a striking difference between the phenotypes of histone tail deletions of H3 and H4 in yeast; both deletions are viable, whilst the double mutant is not. Deletion of the H3 tail leads to activation of a selection of genes (Mann and Grunstein, 1992) where deletion of the H4 tail results in a loss of repression on the silent mating loci (Johnson et al., 1990), but also a reduction in transcription of certain genes (Durrin et al, 1991). Deletion of other histone tails has no effect on the transcription of the subset of genes affected by H4 or H3 tail deletion.

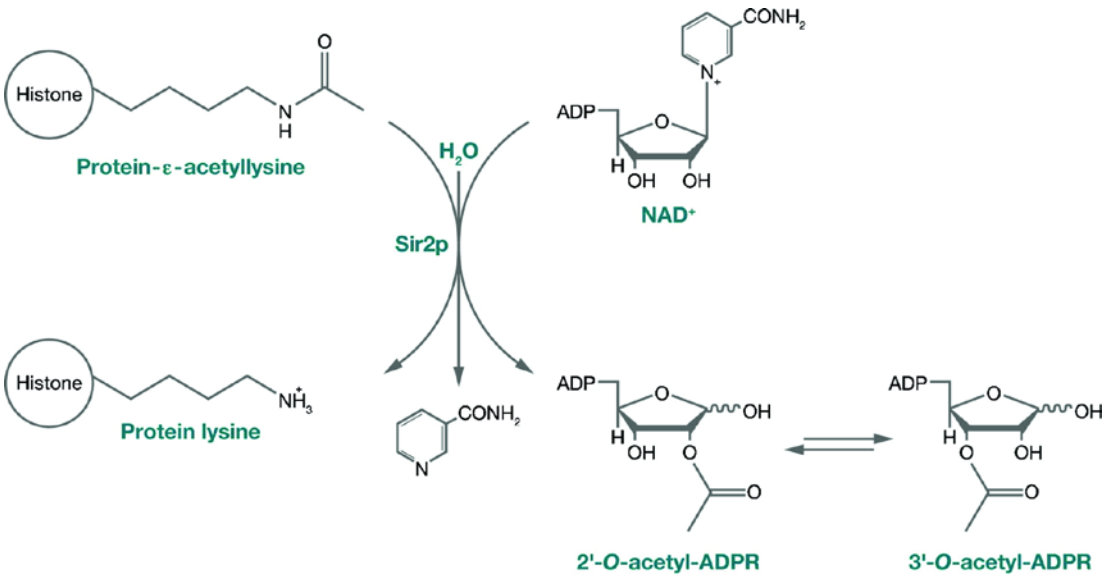
Histone acetylation on H4 is also more enriched on euchromatin, whereas heterochromatin is generally hypoacetylated - however, no correlation was found between total H4 acetylation and transcription (O'Neill and Turner, 1995). Subsequently, the four H4 lysine acetylations were examined individually. Interestingly, in spite of the fact that acetylation is generally associated with activation and euchromatin, in Drosophila, H4K12Ac is enriched at heterochromatin and at telomeric heterochromatin in yeast (Turner et al., 1992; Zhou et al., 2011), though not in humans. On newly synthesised histones (in a manner conserved across species), H4 is diacetylated, on lysines 5 and 12 (Sobel et al., 1995), but the bulk of nuclear histones are instead monoacetylated on lysine 16 (and monoacetylated H4 tails contain only H4K16ac). Histone acetylation proceeds in a preferred order towards the N-terminal through H4K12, K8 and K5 (Turner et al., 1989; Thorne et al., 1990; Zhang et al., 2002).

The acetylation status of histones at particular loci is dictated by a balance between HATs and HDACs. The conservation of these enzymes from yeast to human implies they are functionally extremely important. The importance of HATs in transcription was confirmed with the discovery that a number of well-established transcriptional co-activators (such as CBP and Gcn5) had HAT activity (Bannister and Kouzarides, 1996; Brownell et al., 1996).

By contrast, it was discovered that a transcriptional repressor carried HDAC activity (Taunton et al., 1996).

Based on their sequence and function, HATs fall into one of five families (GNAT, MYST, p300/CBP, $TAF_{II}250$, and nuclear receptor coactivators). MOF belongs to the MYST family of HATs, which frequently contain Zinc finger domains and a PHD or chromodomain (Roth et al., 2001).

Similarly, HDACs fall into one of four classes (I-IV). Class I, II, and IV HDACs remove the acetyl group by hydrolysis (reviewed in (Lombardi et al., 2011)), whilst Class III HDACs (also known as the Sirtuins) rely on NAD^+ (Fig 1.5), and can be specifically inhibited by Sirtinol (reviewed in (Sauve et al., 2006). Class III HDACs are known to deacetylate H4K16ac (reviewed in (Vaquero et al., 2007).



Sauve AA, et al. 2006.
Annu. Rev. Biochem. 75:435–65

Figure 1.5 *NAD$^+$ dependent histone deacetylation by Class III HDAC*

From (Sauve et al., 2006). First, the Sir enzyme binds to NAD$^+$, and catabolises this into activated ADP-ribose and nicotinimide (released). Then the Sir enzyme transfers the acetyl group from the histone lysine to the ADP-ribose, generating *O*-acetyl-ADP ribose as a by-product.

The association of histone acetylation with active regions of the genome and connection with transcribed genes is extremely clear, but what is less clear is whether some, all, or any of the histone modifications are necessary for transcriptional activation and chromatin decompaction at different gene loci.

A comprehensive genome wide study in human (CD4+T cells) looked at the profile of eighteen histone acetylations (including some histone core modifications, as well as those of the N-terminal tails) (Wang et al., 2008). All the histone acetylations studied showed a positive correlation with expression – genes which showed presence of the acetylation had higher expression than those which did not. Although acetylations were increased on active genes, the profile of the modification across those active genes was very different. For example, H3K9ac (along with H2AK9ac, H2BK5ac, H3K18ac, H3K27ac, H3K36ac and H4K91ac) shows a sharp peak over the TSS region, and is absent from the body of the gene. By comparison, H3K23ac (and H2BK12ac, H2BK20ac, H2BK120ac, H3K4ac, H4K5ac, H4K8ac and H4K16ac) has a lower peak over the TSS, and tails off gradually through the body of the gene. The profiles of histone acetylations are not identical. For example H4K12ac also has a low peak over the TSS, which reduces slightly at around 10% through the gene length, but remains even through the rest of the gene, tailing off to background level only after the TES (Fig1.6).

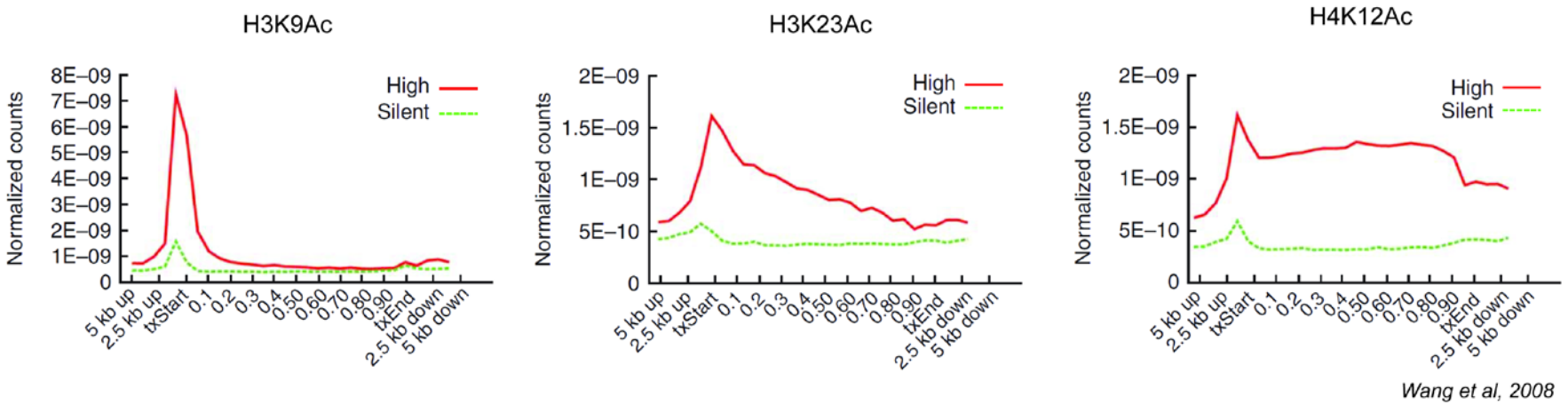


Figure 1.6 *Histone acetylation profiles around the TSS.*

Normalised tag counts at active or silent genes across averaged gene bodies +/- 5kb. txStart, transcription start, txEnd, transcription end. Modified from (Wang et al., 2008).

Although histone acetylation has a general association with active genes, the different binding profiles suggest that some may have individual effects, though this has not yet been widely studied. Of the H4 lysine residues in yeast, it has been shown that mutation of K5, K8 and K12 to Arginine has an additive, negative effect on transcription. Individually, these mutations have a similar effect on the same genes. Mutation of K16, however has a specific effect, separate from the other residues on the histone tail (Dion et al., 2005). The specific effects of H4K16ac will now be discussed in detail.

1.3 H4K16 Acetylation

1.3.1 Direct Effects of H4K16ac

Acetylation of H4K16 is one of the only histone marks which has a direct, causative relationship with chromatin decompaction and upregulation of transcription. Though, as mentioned, many histone modifications have a correlation with euchromatin or heterochromatin and gene expression or silencing, H4K16ac is one of the few which has been shown to have a direct impact on either.

Acetylation of the H4 tail has an impact on higher order chromatin folding. Acetylation of lysines results in an overall reduction in positive charge of the histones (Fig 1.4A), and it is this non-specific electrostatic mechanism which is responsible for chromatin condensation between nucleosome arrays (i.e. aggregation of single nucleosomes or self-association of nucleosome arrays) (Allahverdi et al., 2010; Liu et al., 2011).

However, *in vitro*, acetylated nucleosome arrays are more disordered than unacetylated arrays, and it was found that this effect is specific to residues 14-19 of the H4 tail (Tse et al., 1998; Dorigo et al., 2003). This specific region of the H4 tail has been shown to make contact with the H2A/H2B dimer on the following nucleosome (Fig 1.7), which implicates the region in intramolecular nucleosome folding. This region is functionally important in the formation of heterochromatin in yeast, since it binds to the silencing SIR3 protein (Hecht et al., 1995).

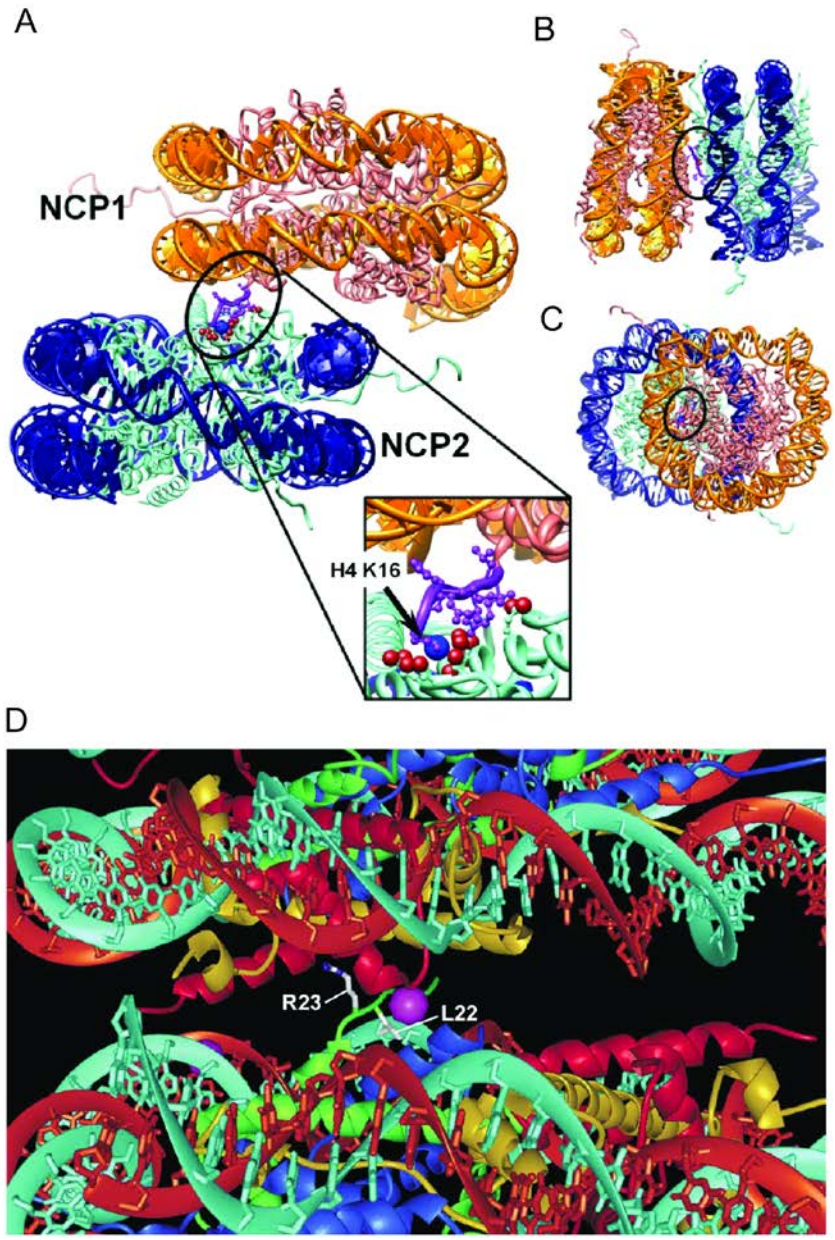


Figure 1.7 *H4 tail modifications interact with H2A/H2B of following nucleosome*

A: Nucleosome Core Particle (NCP) stacking derived from NCP crystals, using the frequently reported structure in which H4K16-R23 (magenta – blue sphere represents NZ atom of H4K16, red spheres = oxygen atoms of carboxylate groups which form the electronegative patch) locates between two NCPs and interacts with the acidic patch of NCP2 H2A. B and C: As A, for different projections. From (Liu et al., 2011). D: H4R23 and H4L22 (and a Mn^{2+} ion in magenta) interact with the acidic region of H2A-H2B dimer of second NCP. H4 = green, H2A = red. From (Davey et al., 2002).

H4K16 is found within this specific region, and its acetylation has been shown to specifically disrupt intramolecular nucleosome folding to a similar extent to deletion of the entire H4 N-terminal tail (Shogren-Knaak et al., 2006). Not only can H4K16ac disrupt the formation of higher order structures *in vitro* but its removal is also necessary for their formation, even on extremely long nucleosome arrays. The longer nucleosome arrays were generated in order to more accurately represent the situation *in vivo* and also contain the chicken erythrocyte isoform of linker histone H1, known as H5. Decompaction of arrays requires removal of the linker histone, and the effect of acetylation of H4K16 and removal of the linker histone are additive (Robinson et al., 2008), adding further evidence that H4K16ac has a vital role in decompaction. Decoupling the specific effect of H4K16ac on intermolecular nucleosome folding from the non-specific electrostatic mechanism; the charge neutralising K16Q mutation does not cause decompaction of nucleosome arrays (Robinson et al., 2008). This adds further evidence that the interaction of the H4 tail with H2A/H2B is the important mechanism in intra-array folding. Additionally, the basic region of H2B can also be bound by the K+ ion – in this situation, 12-mer nucleosome arrays are unable to compact. It was proposed that binding of the K+ ion prevents binding of the H4 lysine residues, which would otherwise be able to bind to the H2B region and allow chromatin compaction (Allahverdi et al., 2010).

Although H4K16ac is associated with decompacted chromatin *in vitro*, the role of H4K16ac in modulating chromatin compaction *in vivo* is less well-studied.

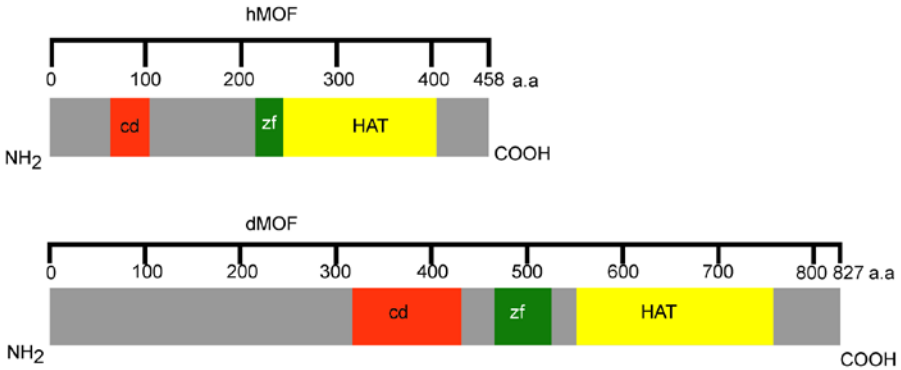
H4K16ac has also been shown to have a direct effect on transcription. In *in vitro* transcription assays (Akhtar and Becker, 2000), targeting MOF to a nucleosome array for the Drosophila *hsp1* gene lead to H4K16ac and derepression.

1.4 Roles of H4K16ac *in vivo*

In vivo, H4K16 is specifically acetylated by MOF (also known as MYST1 or KAT8) (Morales et al., 2004), and in mammals, MOF is required for H4K16ac, but not acetylation of other H4 lysine residues (Taipale et al., 2005). In mouse, MOF is expressed uniformly in adult tissues, but is overexpressed in testis (Thomas et al., 2007, 2008). Like most MYST family HATs, it contains a chromodomain as well as a histone acetyl transferase domain (Fig 1.8A). These domains share homology in mice and flies (Fig 1.8B).

There are many cellular processes in which H4K16ac and MOF are integral. A selection will be discussed below (and Figure 1.9).

A


hMOF
0
100
200
300
400
458 a.a
cd
zf
HAT
NH2
COOH
dMOF
0
100
200
300
400
500
600
700
800
827 a.a
cd
zf
HAT
NH2
COOH

B

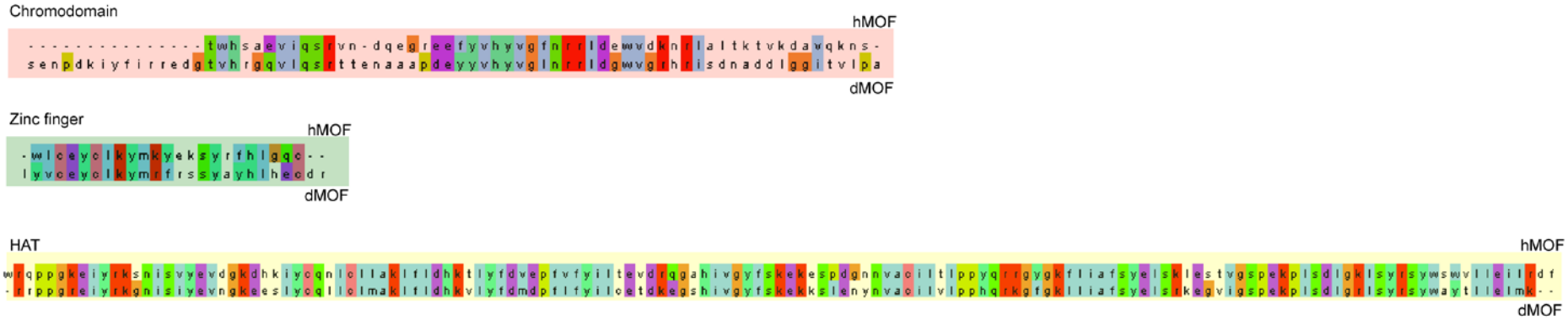

Chromodomain
hMOF
dMOF
Zinc finger
hMOF
dMOF
HAT
hMOF
dMOF

Figure 1.8 *Conserved domains of mouse and Drosophila MOF (Previous page)*

A: Locations of conserved domains detected using UniProtKB (The UniProt Consortium, 2011) and InterPro (for zinc finger domain) (Hunter et al., 2012). Domains are represented in orange (cd = chromodomain), green (zf = zinc finger domain) and yellow (HAT = histone acetyl transferase domain). hMOF adapted from (Gupta et al., 2005a), dMOF adapted from (Akhtar, 2001).

B: Sequence alignment for conserved domains, generated through ClustalOmega (Sievers et al., 2011). Domains highlighted with colours as in A.

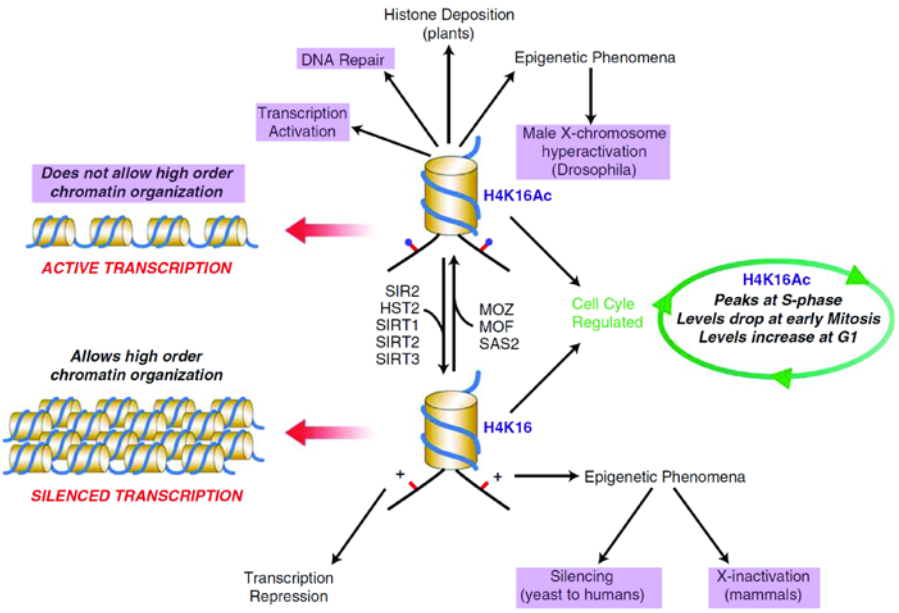


Figure 1.9 *Functions of H4K16ac.*

Nucleosomes represented by gold cylinder with DNA represented by blue line. H4K16 acetylation is depicted in the top half of the figure by a blue circle over K16 (red line). Functions discussed below are highlighted in purple. Modified from (Vaquero et al., 2007).

1.4.1 Dosage Compensation in Flies and Mammals

X chromosome dosage compensation is an example of a situation where histone acetylation is important, in both flies and mammals. In mammalian systems, where females carry two copies of the X chromosome, and males only one, the females silence one entire chromosome to ensure that the dose of genes on the X-chromosome is equivalent to that in males. This silencing is achieved by formation of facultative heterochromatin, and involves multiple layers of regulation. In flies, the situation is reversed, and males must upregulate the genes on the X-chromosome twofold to ensure equal dosage (reviewed in (Baker et al., 1994)) .

Modification of chromatin structure and its ability to allow transcription is a hallmark of dosage compensation across species. In Drosophila, the generation of more relaxed chromatin on the male X chromosome is seen on the polytene chromosomes after knockdown of the ISWI chromatin remodelling complex– by comparison to the wild type, the chromosome is 'bloated' and expanded. Preventing acetylation of H4K16ac rescues this phenotype (Straub and Becker, 2007).

MOF was first discovered as a regulator of X-chromosome dosage compensation in flies, which is reflected in its name; MOF, or Males absent on the first. In the *mof* mutant male flies, the X chromosome loses H4K16ac (Turner et al., 1992; Hilfiker et al., 1997), and the male flies die.

In mammals, the mechanism of dosage compensation is regulated differently, but the inactive X is hypoacetylated on K5, 8, 12 and 16 (Jeppesen and Turner, 1993), and hypoacetylation has been proposed to be important for maintainance of the inactive state (Keohane et al., 1996; O'Neill et al., 1999).

1.4.2 The role of H4K16ac in genome wide transcription

The role of MOF in Drosophila was initially thought to be limited to dosage compensation, due to the *mof* mutant's male specific lethality (Hilfiker et al., 1997). However, it was later shown that females homozygous for a different *mof* mutation (lacking the entire HAT and chromodomain) are sterile, and have a significantly shortened lifespan compared with wild type flies (Conrad et al., 2012a). In addition, it was noted that the distribution of MOF across the genome is much wider than that of the MSL proteins (Gelbart et al., 2009) and

that MOF binds to the majority of active autosomal promoters (Raja et al., 2010). MOF is also necessary for genome wide H4K16ac in both male and female flies (Conrad et al., 2012a), and H4K16ac correlates with transcriptional activation on female chromosomes and male autosomes (Prestel et al., 2010).

In mammals, genome wide studies of H4K16ac have shown it to be associated with active promoters across the genome (Wang et al., 2008; Wong et al., 2011), and it also has a functional connection with transcription; in HeLa cells, knockdown of MOF leads to a reduction in both H4K16ac at the promoter of *HOXA9*, and transcription of the gene (Dou et al., 2005).

1.4.3 Functional Separation of MOF roles

In flies, MOF and H4K16ac have a role in both dosage compensation and genome wide transcription. Separation of these two roles is achieved through differing interaction partners of MOF. The interaction partners are important for the histone substrate specificity of MOF, which is unusually narrow for a HAT; other HATs generally acetylate a broad group of histone residues (summarised in (Rea et al., 2007)). In isolation, MOF is less specific when acetylating free histones, and has no HAT activity on nucleosomal histones (Morales et al., 2004). By contrast, when partnered with MSL1 and 3, MOF acetylates only H4K16.

The reliance of MOF on its interaction partners for enzymatic activity is key to organisms' ability to segregate its functional roles.

For hyperacetylation targeted to the male X chromosome in flies (Bone et al., 1994), the MSL complex and a series of non-coding RNAs, expressed only in male flies are vital. Together with MOF, these are known as the Dosage Compensation Complex (DCC). The first step is production of the MSL2 protein, which is silenced in females. MSL2 is also an E3 ligase, ubiquitinating H2B (Wu et al., 2011). The rest of the DCC is assembled using MSL2 as a scaffold, and consists of MSL1, MSL3, MOF, and MLE, which are also expressed in female flies, and the male specific non coding RNAs (roX), which are not. The DCC specifically binds to the male X chromosome in a manner dependent on the roX non coding RNA transcripts. MOF can bind to the roX2 RNA via its chromodomain (Akhtar et al., 2000). Since roX RNAs are very quickly degraded it is thought that the DCC is fully assembled at their site of transcription on the X chromosome, and then spreads to the rest of

the chromosome, allowing MOF to acetylate H4K16ac. H4K16ac is required for spreading of the complex across the chromosome, indicating that a certain extent of chromatin decompaction mediated by acetylation may be required for the DCCs progress along the chromosome (reviewed in (Conrad and Akhtar, 2012)).

Though MOF can bind to one of the roX RNAs, this function can be dispensed with for its integration into the DCC (Morales et al., 2004). Instead, MOF has a direct interaction with MSL1 through its zinc finger domain. A point mutation in the zinc finger can prevent MOF association with the male X chromosome entirely (Morales et al., 2004).

The chromodomain of MOF is dispensable for its integration into the DCC and efficient targeting of the DCC to the male X chromosome, but its highly conserved state indicates that it still holds some functional importance. Interestingly, in Drosophila, there is a large region between the chromodomain and the N-terminal of the protein. This region is not conserved in species with different dosage compensation systems. Though point mutations in the chromodomain did not affect DCC localisation (Morales et al., 2004), the same point mutation led to total male lethality and defective spreading of MSL1 and 3 (Conrad et al., 2012a).

H4K16ac is central to the eventual downstream effect of the DCC, though the exact mechanism of upregulation of transcription has yet to be elucidated. A clue comes from ChIP-sequencing experiments for MOF, H4K16ac and a subunit of RNApolII performed in flies. These experiments also show how H4K16ac profile changes according to function.

On the autosomes of male flies (and all chromosomes of female flies), H4K16ac is found only around the gene promoters. However, on the male X chromosome it is found through the length of the genes (Fig 1.10). Additionally, on the male X chromosome, MOF is enriched towards the 3' ends of the gene (Kind et al., 2008; Larschan et al., 2011). This correlates with increased enrichment of RNApolII on the genes of the male X (Conrad et al., 2012b). These genes also have increased transcriptional elongation, which has been posited as a potential mechanism for H4K16ac action (i.e. the increased accessibility of the chromatin after acetylation allows the transcriptional machinery an easier path through the gene) (Larschan et al., 2011). This is supported by the fact that in isolated nuclei the male X chromosome is highly accessible to the DNA methylase M.Sssl (Bell et al., 2010).

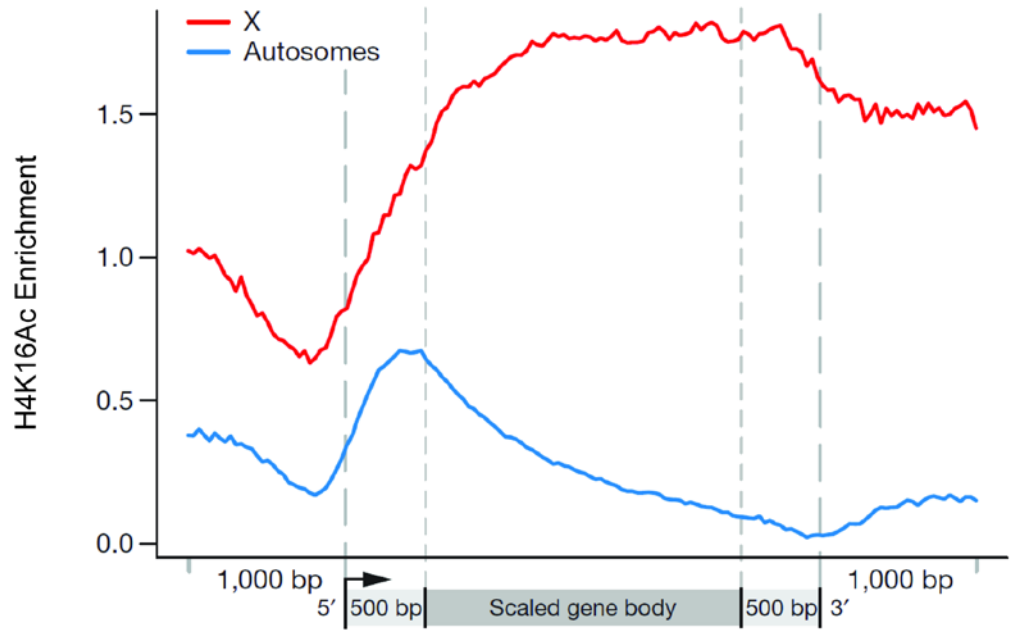


Figure 1.10 *Differing profiles of H4K16ac on Drosophila male X or autosomes*

From (Larschan et al., 2011). H4K16ac is localised across the body of the genes on the active genes of the male X chromosome (red), but is restricted to the 5' end of active genes on the autosomes.

In Drosophila, targeting of MOF and H4K16ac to the promoters of male and female autosomal genes relies on interaction with the 'Non-specific lethal' (NSL) complex (Feller et al., 2011) – so called because its disruption is lethal to both male and female flies. In flies, the NSL complex is capable of transcriptionally activating a promoter if it is artificially tethered there; importantly, depletion of MOF seriously reduced the level of reporter gene transcribed (Raja et al., 2010). The biological role of NSL targeting to autosomal genes seems to be regulation of the housekeeping genes (Lam et al., 2012), though not all NSL bound housekeeping genes are active (Feller et al., 2011). The activation potential of MOF, measured by transcriptional output of its target genes, is much higher when MOF is in the context of the NSL complex than the MSL complex. As part of the MSL complex MOF only effects a twofold increase of expression, as required for accurate doseage compensation (Prestel et al., 2010).

Both the MSL and NSL complexes are evolutionarily conserved, indicating that there may also be a separation of function for MOF in mammals. In humans, like in flies, the NSL proteins do not join the MSL complex, but instead form a distinct complex with a size of approximately 300-400kDa, which includes MOF (Mendjan et al., 2006; Cai et al., 2009). In mammals the NSL complex is also known as MSL1v1, and KANSL; for simplicity here it will be referred to as the NSL complex in both species. The human NSL complex is enzymatically active *in vitro*, not only on H4K16, but also on H4K5 and H4K8 (Cai et al., 2009). However, *in vivo* in Drosophila, relaxation of MOF substrate specificity was not observed (Lam et al., 2012). MOF interacts with the NSL complex via an interaction between its HAT domain and the C-terminal domain of NSL1 (Raja et al., 2010). The functional purpose behind segregation of MOF into two complexes in mammals is still unknown, but it has been suggested that as part of the NSL complex, the principle substrate of MOF is the non-histone protein p53 (Li et al., 2009). Alternatively, MOF has also been shown to potentially interact with the H3K4 methyltransferase MLL. This interaction is mediated through WDR5, which is shared between the NSL-MOF complex and the MLL complex, but not the MSL-MOF complex (Dou et al., 2005). The interaction with MLL has been disputed, however (Cai et al., 2009).

It is interesting to note that the 17q21.31 microdeletion syndrome (which causes intellectual disability, epilepsy, heart and urogenital defects, along with a facial morphology phenotype) was recently shown to be caused by a mutation in the NSL complex protein KANSL1 (or NSL1). Cell lines established from patients showed transcriptional misregulation of a number of genes, which are enriched for similar GO terms as NSL1 bound genes in

Drosophila. This implies that the NSL-MOF complex may target MOF binding to a specific subset of genes (Koolen et al., 2012; Zollino et al., 2012). NSL/MSL known functions are summarised in Figure 1.11.

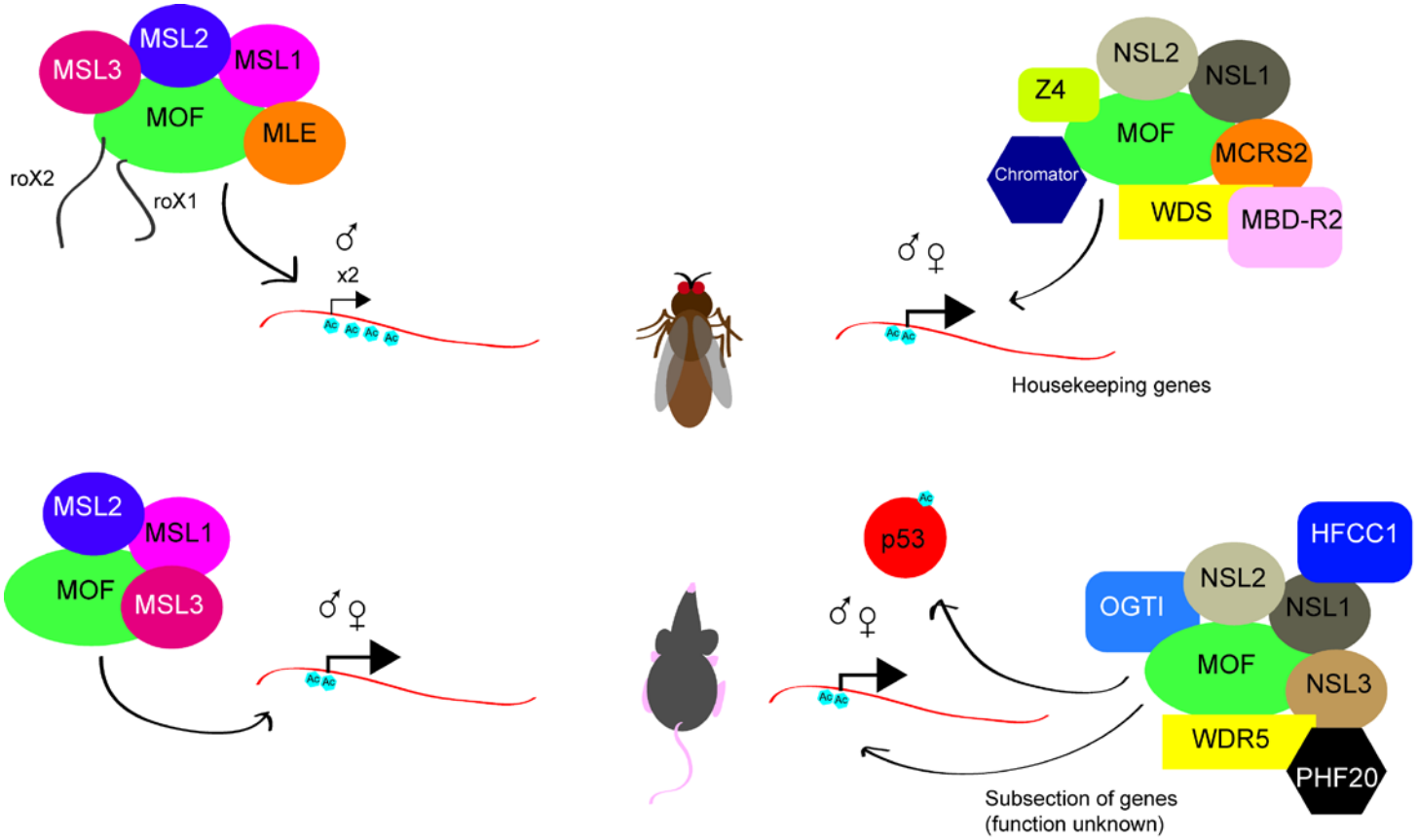


Figure 1.11 *Segregation of MOF function by integration into separate complexes in flies and mammals*

Model of MOF functions in MSL and NSL complexes in flies (top) and mammals (bottom), mammalian data summarised from pull downs in HeLa cells (Cai et al., 2009). Drosophila complexes from (Raja et al., 2010) and (Mendjan and Akhtar, 2007).

1.4.4 Histone Acetylation and H4K16ac in Mammalian Development

ES cell differentiation is an example of a developmental situation in which histone acetylation may be important. Leukaemia Inhibitory Factor (LIF), is added to the culture media of mouse ES cells, in order to maintain them in the pluripotent state. If the LIF is withdrawn, the cells begin to differentiate non-specifically (i.e. down a variety of lineages) (Williams et al., 1988). There is a global increase in bulk acetylation of H3 and H4 upon differentiation, and treatment with a HDAC inhibitor can reversibly induce differentiation without removal of LIF (McCool et al., 2007). These experiments did not distinguish between the four acetylated lysines on the H4 tail, but H4K16ac has also been shown to be specifically involved in development.

Mice deleted for MOF die prior to the blastocyst stage, with extreme morphological abnormalities. At E4.5, wild type cells show distinct development of the Inner Cell Mass (ICM), an ordered group of cells distinct from the Blastocoelic Cavity (BC). However, in MOF-/- embryos, the ICM does not develop properly and by E5.5 has degenerated into an entirely disordered structure (Fig 1.12A). *In vitro* culture of the embryos revealed that the cells die by apoptosis, but prior to activation of caspase-3, they show chromosome abnormalities characterised by a loss of focused bisbenzimide staining (the staining is present throughout the nucleus in strongly affected cells). $MOF^{-/-}$ cells also show a loss of H4K16ac (Thomas et al., 2008). This data showed that H4K16ac is important for development in mammals.

In addition, MOF was recently shown to be of vital importance in the maintenance of pluripotency of ES cells. Interestingly, prior to this discovery, it was shown that HDAC inhibitors were amongst a variety of molecules which enhanced formation of induced pluripotent stem cells (iPS cells) *in vitro* (Feng et al., 2009), in spite of the fact that TSA (a HDAC inhibitor) can induce pluripotent cells to differentiate (McCool et al., 2007). Thomas et al (2008) were unable to produce ES cells from $MOF^{-/-}$ mice, and therefore a recent study (Li et al., 2012) instead used an inducible Cre-floxing approach, though even this method resulted in eventual cell death. In wild type cells, MOF protein levels are reduced upon retinoic acid differentiation, supporting the idea that it has a specific role in ES cells. Further, by working with $MOF^{flox/flox}$ cells (prior to cell death) Li et al showed that chromatin structure is greatly perturbed by loss of MOF in ES cells, with a large increase in apparently condensed chromatin compared to wild type (Fig 1.12B). $MOF^{flox/flox}$ ES cells

show a reduction in markers for pluripotency, including reduced Alkaline Phosphatase (AP) staining (O'Connor et al., 2008), and reduced expression of key pluripotency associated genes, such as *Oct4* and *Nanog*. The pluripotency associated genes *Klf4* and *c-Myc* showed an increase in expression, surprisingly, though MOF binding and H4K16 acetylation were reduced in $MOF^{flox/flox}$ cells on these genes. MOF has a broad role in transcriptional regulation in ES cells. Upon removal of MOF from ES cells, approximately 3000 genes show a 2-fold change in expression (around half are upregulated and half downregulated). These genes fall into many categories, including cell cycle regulation and DNA repair, though downregulated targets are enriched for pluripotency maintainance, and upregulated targets for differentiation specific genes. Interestingly, the misregulated genes correlated highly with targets of *Nanog*, and overexpression of *Nanog* was capable of rescuing the $MOF^{flox/flox}$ phenotype (Li et al., 2012).

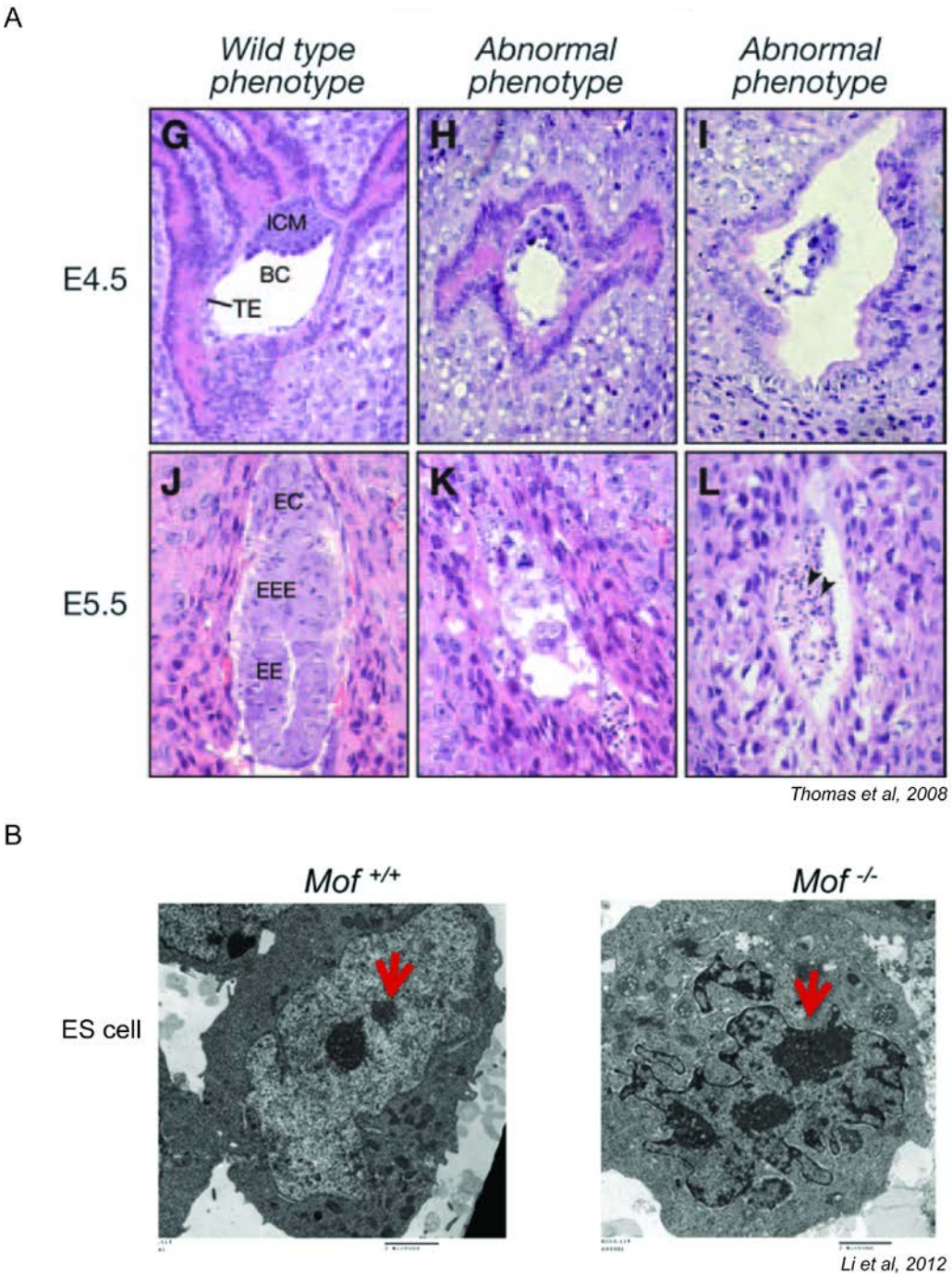


Figure 1.12 *Effects of Mof knockout on development and ES cell nuclear morphology*

A) Sectioned uteri from timed matings at E4.5. BC, blastocoelic cavity; EC, ectoplacental cone; EE, embryonic ectoderm; EEE, extraembryonic ectoderm; ICM, inner cell mass; TE, trophectoderm; ZP, zona pellucida. From (Thomas et al., 2008). B) Electron Microscopy (EM) images of wild type (right) and Mof knockout ES cell nuclei. Dense staining of heterochromatin is highlighted with red arrow. From (Li et al., 2012).

1.4.5 Histone Acetylation and DNA Damage Repair

Since histone acetylation, and H4K16ac in particular, are associated with changes in chromatin structure, it is to be expected that other processes for which the organisation of the DNA would have some impact would also be affected by acetylation. An example of this is DNA damage repair. Exactly how the DNA damage machinery accesses damaged DNA when it is located in compacted chromatin in mammals is not yet known, but chromatin remodelling complexes and histone modifications are involved (reviewed in (Wong et al., 2006)).

In yeast, there is a significant amount of work linking histone acetylation and double strand break repair. Indeed, acetylation of H4 is considered to be essential. Mutation of lysine residues to glutamine results in hypersensitivity to double strand breaks induced by camptothecin (CPT). Replacing only one of the lysines rescues the phenotype (lysines 5-12 show equal rescue of CPT sensitivity, but lysine 16 is slightly less effective), insertion of an extra lysine (which is acetylated) also rescues the phenotype (Bird et al., 2002). Deletion of Esa1 (a MOF homolog in yeast which is responsible for acetylation of all the lysines on the H4 tail) also leads to sensitivity to DSBs (Bird et al., 2002).

Additionally, after a double strand break, the histone acetylation is removed; HDACs are recruited approximately three hours after induction of the DSB, and levels of all H4 acetyl-lysines is reduced (Tamburini and Tyler, 2005).

The function of histone acetylation in DNA damage repair is linked to the histone H2AX, which is phosphorylated on serine 139 upon breakage of the DNA (Rogakou et al., 1998). Upon induction of a double stranded break (DSB) in yeast, H2AX phosphorylation may recruit the HAT complex NuA4 through an interaction between phosphorylated H2AX, and a subunit of NuA4 (Arp4). The HAT complex is found at the site of double stranded breaks, and acetylates the H4 tail. The mechanism by which H4 acetylation aids in the cells ability to cope with DSBs is not known, but replacing the lysines in the H4 tail, or mutating the HAT complex responsible leads to hypersensitivity to DNA damage by induction of DSBs (Attikum and Gasser, 2005). This points to a general role of histone acetylation in DNA damage repair (suggested to be due to ‘opening’ of the chromatin at the site requiring repair), but it has also been shown that H4K16ac and MOF have a more specific role in this process.

In humans, MOF has been identified as an interaction partner of the protein ATM (ataxia-telangiectasia-mutated) which is a critical part of the cellular response to DNA damage. Cells with reduced level of MOF activity (either through knockdown or expression of a mutant with no HAT activity) show an increase in the number of spontaneous chromosome breaks, but also decreased survival after irradiation (Gupta et al., 2005b). Reduction in levels of MOF mediated H4K16ac did not lead to reduction in expression of any proteins involved in DDR. Instead, it has a direct effect on H2AX focus formation, which is delayed upon reduction of H4K16ac (Sharma et al., 2010), though, again, the exact mechanism has yet to be elucidated.

1.5 Histone Acetylation, H4K16ac and Disease

1.5.1 Cancer

Consistent with its role in regulation of transcription and DNA damage repair, histone acetylation has also been implicated in situations when regulation of these processes goes wrong. Histone acetylation and cancer is one example of this association. HDAC inhibitors that are FDA approved for use in anti-cancer therapy include Vorinostat and Romidepsin. HDAC inhibitors are proposed to act on cancer via their interference with histone acetylation, but they could have many off target effects (Mack, 2010). The DNA binding activity of p53, for example, is activated by acetylation; binding of DNA is a feature of p53 frequently absent in oncogenic mutants (Gu and Roeder, 1997).

Acetylation of p53 is a process which is affected by inhibition of HDACs, which are known to deacetylate p53 in mammals (Brooks and Gu, 2011). Histones do, however, represent the most abundant substrate for HDACs, so at least some of the effect of HDAC inhibitors may be due to changes in histone acetylation.

The p53 protein is widely regarded as one of the most important in the cancer field. Its inactivation leads to tumours due to its ability to trigger either arrest of the cell cycling process or apoptosis leading to cell death, by stimulating transcription of the genes involved in one pathway or the other (Sykes et al., 2006).

Interestingly, MOF is amongst the HATs which are known to acetylate p53. MOF and another HAT, Tip60, acetylate p53 on lysine 120. Deacetylation of p53 K120ac prevents p53 from activating transcription of its pro-apoptotic target genes, but does not affect its

ability to upregulate the transcription of pro-cell cycle arrest targets, such as *p21* (Sykes et al., 2006).

Evidence for a direct involvement in H4K16ac in cancer, aside from its effect on double strand break repair, discussed above, is limited, but human MOF has been shown to be downregulated in a subset of breast cancer cases (Pfister et al., 2008).

1.5.2 Ageing

Like organismal development, the process of ageing is linked changes in chromatin, and histone acetylation has been implicated particularly in determination of lifespan, through the action of the Sir proteins, a group of HDACs. Ageing and epigenetics is reviewed in (Fraga and Esteller, 2007).

In yeast for example, 'ageing' is a function of the number of daughter cells one parent cell can produce – the two cells can be distinguished in budding yeast, since the daughter cell is smaller. The mother cell will eventually cease to divide and die, after a number of divisions, (between approximately 10 and 40) (Mortimer and Johnston, 1959).

Screens in yeast for proteins which affected lifespan yielded the SIR proteins (reviewed in (Kennedy et al., 1997) – HDACs important for gene silencing in yeast. Sir mutants can increase lifespan as well as reduce it – for example, in the SIR4-42 mutant, two Sir proteins relocate from the telomeres to the nucleolus and lifespan is extended (Kennedy et al., 1997). Deacetylation of H4K16ac has subsequently been shown to be important for telomeric silencing (Kimura et al., 2002), and H4K16ac is also important in regulation of cellular lifespan in yeast.

The profile of histone modification post translational modifications changes upon aging in yeast, as it does upon cellular differentiation in mammals. For example, H3K56ac is increased in younger cells which have produced few daughter cells, whereas in older yeast, H4K16ac is increased as the level of the HDAC Sir2 is decreased. In these older cells, as would be expected from the role of histone deacetylation on silencing of telomeres, the increased H4K16ac is associated with a reduction of silencing on the telomeres. Lifespan of yeast can be extended somewhat by either deletion of the HAT responsible for H4K16ac, or overexpression of Sir2 (Dang et al., 2009).

In mammals, the effect of H4K16ac on ageing is less well studied, and whilst the life elongating effect of Sir2 expression is replicable in both *C. elegans* and Drosophila, similar pathways have not yet been found in humans (reviewed in (Longo and Kennedy, 2006)). Telomeres in mammals are, like in flies and yeast, heterochromatic, and are hypoacetylated on H4 (Benetti et al., 2007).

A role for H4K16ac in mammalian ageing has, however, been found, although the mechanism is rather different, and is associated with Lamin A and the role of H4K16ac in DNA damage. Lamin A is a nuclear membrane protein associated with premature ageing disorders when truncated (Eriksson et al., 2003). In mice which cannot produce mature Lamin A and which show a premature ageing phenotype, there is a reduction in global H4K16ac. There is an interaction between mature Lamin A and MOF, which is not replicated with prelamin A (the immature version of the protein). Lack of this interaction leads to a reduction in MOF presence at the nuclear envelope, and a poor response to DNA damage which can be rescued by treatment with HDAC inhibitors or MOF overexpression (Krishnan et al., 2011).

1.6 Long Range Control

1.6.1 Enhancers in development

Although some genes are controlled with only the core promoter, for many genes, regulation of expression is achieved through the action of cis-regulatory elements. The first enhancers were discovered in 1981; sequences from the SV40 virus that are capable of upregulating expression of a transgene up to 200 fold, even when positioned ~3000bp downstream of the transcription start site (Banerji et al., 1981). Other cis-acting regulatory regions have been shown to have a silencing effect, for example sequences 5' of murine *Igβ* repress its transcription (Malone et al., 1997). Cis-regulatory regions have been located as much as 1 or 1.5Mb away from the transcription initiation site (Lettice et al., 2002; Benko et al., 2009).

Cis-regulatory regions are also potentially interesting since 88% of trait associated SNPs fall within non-coding regions (Hindorff et al., 2009), and mutations of cis-regulatory regions have been shown to explain disease phenotypes.

For example, an insufficient doseage of Pax6 has been associated with aniridia (where patients lack an iris) for some time(Axton et al., 1997). However, patient studies revealed

that there were some cases of aniridia which showed no mutation in the coding region of PAX6, even though the phenotype was the same as those which did (Fantes et al., 1995). This implied that another region was exerting control over the eventual dosage of Pax6.

The functional region of PAX6 includes a regulatory region >150kb downstream of the PAX6 promoter, known as the Downstream Regulatory Region (DRR). Selected conserved regions of the DRR are capable of driving tissue specific expression of a reporter gene in mice (Kleinjan et al., 2001).

The DRR is located within the final intron of the ELP4 (a.k.a PAXNEB) gene (Fig 1.13), which is a universally expressed HAT which functions to enhance transcriptional elongation (Kleinjan et al., 2001). However, it was subsequently shown that mutations in ELP4 do not lead to aniridia, but the phenotypes of chromosome breakages in the DRR are due to disruption of PAX6 regulation(Kleinjan et al., 2002).

Another example is that of the Sonic Hedgehog (Shh) regulatory region. Sonic hedgehog is a morphogen with a vast array of functions in a variety of systems including neuronal development (reviewed in (Jessell, 2000)), muscle determination (Borycki et al., 1999), and limb development (reviewed in (Anderson et al., 2012)).

In a healthy mouse limb bud, *Shh* is expressed in the posterior (the Zone of Polarising Activity, or ZPA) limb mesenchyme to form a gradient of Shh to the anterior side. When grafted to the anterior of the limb, the ZPA induces formation of a mirror image digit duplication. This is due to the effect of the Shh gradient. Ectopic expression of *Shh* in the anterior limb buds also results in mirror image digit duplications (Riddle et al., 1993).

Mutations in intron 5 of the *Lmbr1* gene (which is some 1Mb away from the Shh transcription start site) lead to a phenotype of extra digits - preaxial polydactyly (PPD), due to the region's regulatory effect on *Shh* (Lettice et al., 2002). Genetic analysis in mice has shown that mutations in intron 5 of *Lmbr1* lead to ectopic expression of *Shh* in the anterior of the limb bud (Lettice et al., 2003), and deletion of the conserved regulatory sequence abrogates expression of *Shh* (Sagai et al., 2005). The region was titled the ZRS, or ZPA Regulatory Sequence.

Point mutations in the ZRS can cause PPD and other limb defects in mice (as in the Sassquatch mutant and humans (Lettice et al., 2003; Furniss et al., 2008), cats (Lettice et al., 2008) and chickens (Maas et al., 2011) (Fig 1.14).

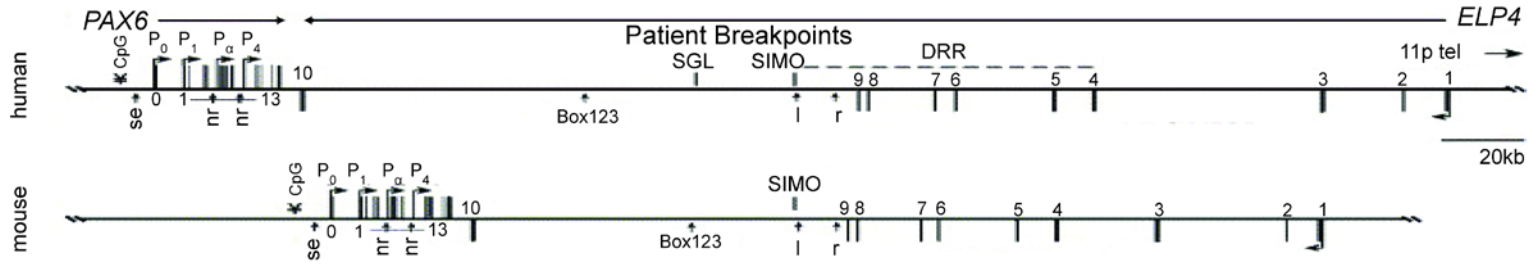


Figure 1.13 *Human and mouse PAX6 locus and regulatory regions*

SGL and SIMO indicate aniridia associated chromosomal breakpoints. Eye specific regulatory elements (which can drive specific expression in reporter assays) are indicated for surface ectoderm (se), neural retina (nr), lens (l), and a retina (r) enhancer which drives expression in neural retina and retinal pigment epithelium. Box123 indicates a regulatory region for neural retina. DRR denotes the *PAX6* distal regulatory region. Exons are shown in by black boxes, and direction of transcription is indicated with arrows. From (Kim and Lauderdale, 2006).

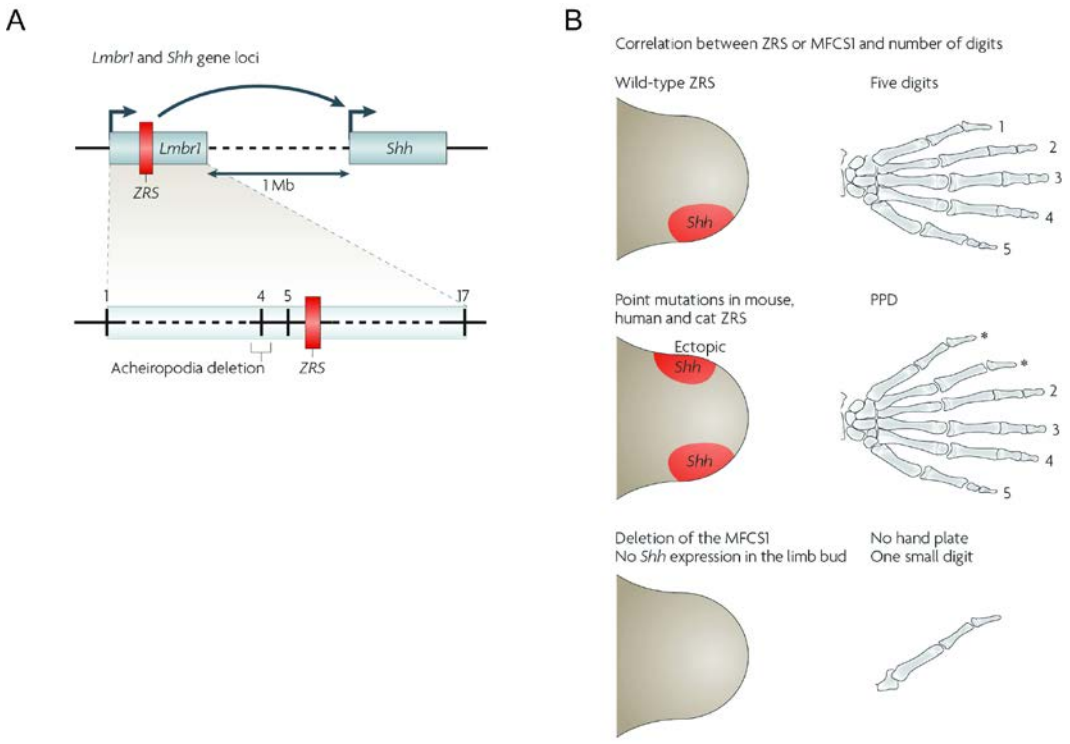


Figure 1.14 *ZRS region and functional consequences of mutations*

A: Schematic representing location of the *Lmbr1* locus with 1.7kb limb specific *Shh* regulatory region (red box). *Shh* loss of function in mice phenotypically resembles human acheiropodia; mutations for this disease map close to the ZRS.

B: Phenotypes of ZRS mutations. Mutations that lead to PPD have their effect through ectopic expression of *Shh*. Other mutations, such as deletion of an 800bp region within the ZRS (known as the Mouse Fish Conserved Sequence 1, or MFCS1), and human acheiropodia deletions lead to loss of *Shh*, resulting in a complete lack of hand plate. From (Zeller et al., 2009).

The *Hox* loci are an interesting example in which long range regulatory regions control changes in function by generating a variety of expression profiles in different embryonic structures. During development of vertebrate organisms, the *Hox* genes are activated according to 'temporal colinarity', i.e. temporally according to their relative positions along the genome. This pattern of expression leads to different combinations of *Hox* genes instructing different cell fates for groups of cells along the axis (reviewed in (Kmita and Duboule, 2003)).

The core anterior posterior body patterning functions of the *Hox* loci seem to be regulated from within the clusters themselves; expression of the *Hoxd* cluster from a PAC clone is capable of re-capitulating its colinar expression pattern as it is found in wild type embryos (Spitz et al., 2001).

However, in vertebrate evolution, organisms have taken advantage of the duplications of *Hox* genes which have occurred, by putting the genes to use in other contexts at different locations and times during development. For example, expression of the *Hoxd* genes (particularly *Hoxd13*) is required for correct limb formation in mice (Davis and Capecchi, 1996). *Hoxd* expression occurs in two waves, the first for limb growth and patterning, and the second at embryonic day 10.5 for formation of the digits (Spitz et al., 2003). Expression of the *Hoxd* genes in the limbs is not reproduced by its introduction by PAC clone (Spitz et al., 2001), implying that they are under control of a regulatory region which is not found within the *Hox* clusters proper.

An example of such a regulatory region is the GCR, or Global Control Region which was discovered to be located upstream of the *Hoxd* locus. The GCR can drive expression of a transgene in the developing mouse limb in a pattern which recapitulates that of the *Hoxd* genes. The GCR is the region mutated in the mouse *Ulnaless* mutant, which has reductions in the size of the proximal limb due to deregulation of the posterior *Hoxd* genes (Peichel et al., 1997; Spitz et al., 2003).

Interestingly, expression of *Hoxd13* in cells derived from the posterior of the distal limb is accompanied by decompaction of the chromatin, compared to cells derived from the anterior of the distal limb, where *Hoxd13* is not expressed. Additionally, there is colocalisation between the GCR and the 5' end of the *Hoxd* locus which occurs only in the posterior derived cells. This indicates that chromatin organisation may be extremely important for the action of this enhancer on transcription of the *Hoxd* genes (Williamson et al., 2012).

There are several enhancer regions which act on the *Hoxd* genes (Fig 1.15). For example, *Hoxd10* and *11*, in wild type embryos, are acted on by the same enhancers, and are found in the same expression pattern both in the distal and proximal parts of the developing limb. Splitting the genes by inverting a section of the loci from between *Hoxd10* and *Hox11* to between *Itga6* and *Itga6r* removes each of the genes from the influence of one of these enhancers, and therefore splits their expression pattern so that *Hoxd11* is expressed only in the distal portion of the limb, whilst *Hoxd10* is found in the proximal portion (Spitz et al., 2005). The GCR is not the only enhancer region responsible for the effects of *Hoxd* genes in digits as subsequent studies determined that the 'Prox' enhancer (located between *Lnp* and *Evx2*) also has activity in digit formation. Additionally, a further five regulatory regions located in the gene desert upstream of *Lnp* and the GRC, and were shown to have a positive effect on *Hox* gene transcription. Deletion of these elements leads to malformation of the digits in mice (Montavon et al., 2011).

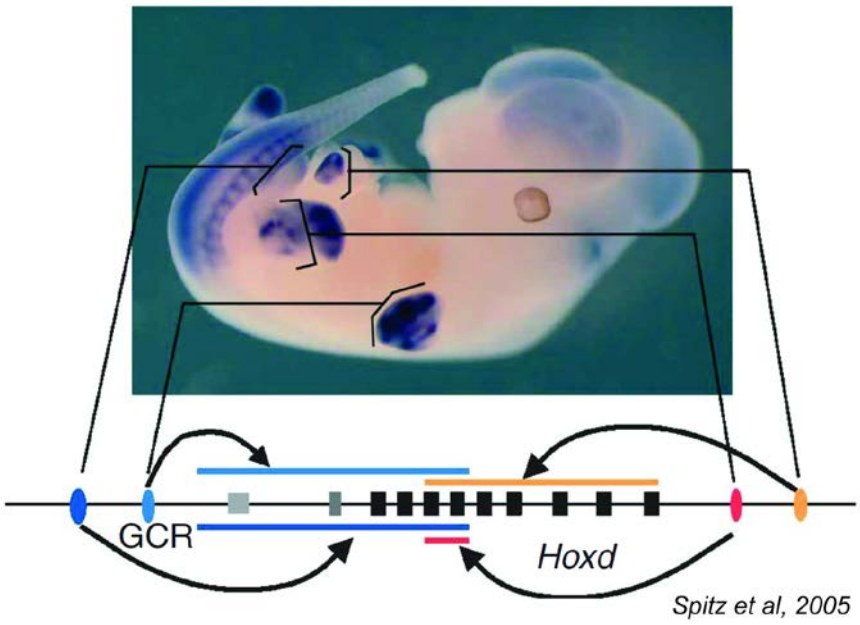


Figure 1.15 *Several enhancers act on the Hoxd genes leading to different expression patterns*

β-gal staining for Hoxd10 in mouse embryo. Cis-acting sequences are represented by ovals which control expression of genes highlighted in the same colour. Domain of expression induced by these genes is indicated. GCR = global control region. From (Spitz et al., 2005).

1.6.2 Chromatin State and Long Range Control

Chromatin state is very important for enhancer action. Classically, DNaseI Hypersensitive Sites (DHSs) have been used to analyse the location of regulatory elements in the genome (reviewed in (Gross and Garrard, 1988)). DHSs are likely to be nucleosome free regions, and are sensitive to nuclease cleavage (Elgin, 1981; McGhee et al., 1981) possibly due to binding of a transcription factor or factors (Felsenfeld et al., 1996). DNAse I footprinting has shown that 75% of DHSs contain a bound protein (Neph et al., 2012).

A genome wide study noted that approximately 80% of DHSs fell within a regulatory region or regions (Crawford et al., 2006), and 95% of DHSs are located outside coding regions, making them important for identification of regulatory elements genome wide (Thurman et al., 2012).

Reflecting the importance of cis-regulatory regions in control of cell fate and development, many DHSs are cell type or tissue specific, meaning that the chromatin accessibility of enhancers differs depending on whether they are active or not (Thurman et al., 2012).

Histone modifications are another method of predicting the location of regulatory regions which has recently come to the fore. An early genome wide ChIP experiment noted the association of H3 acetylation with regulatory regions (Roh et al., 2005), and that this was capable of predicting the presence regulatory elements, even those which are not conserved by sequence (Roh et al., 2007). The H3 acetyltransferase p300 is also known to bind at enhancers, as well as promoters (Heintzman et al., 2007), and other HATs have been shown to mark enhancers which are not p300 bound (Krebs et al., 2011). Recently H3K9ac and H3K14ac were shown to mark regulatory regions in ES cells (Karmodiya et al., 2012).

ChIP-chip experiments highlighted for the first time the potential for histone methylations as a predictor for chromatin function. Looking at 30Mb of the human genome in HeLa cells, H3K4 methylation state can distinguish promoters from enhancers, which are marked by H3K4 trimethylation, and H3K4 monomethylation respectively. The methylation state was shown to be capable of predicting the location of novel enhancers (Heintzman et al., 2007). This was followed by a genome wide profile of 20 histone methylations (of lysine and arginine) in $CD4^{+}$T cells, which showed that amongst enhancer elements identified by DHSs, H3K4me1, me2 and me3 were all elevated, with H3K4me2/3 showing more defined peaks, and H3K4me1 marking the regions more broadly (Barski et al., 2007). Further genome wide

studies provided evidence that H3K4me1 is a signature of enhancers (Birney et al., 2007; Koch et al., 2007), and reviewed in (Bulger and Groudine, 2011)), whilst H3K4me3 is a signature of promoters (Fig 1.16A). H3K4me1 is now commonly used as a marker of both active and inactive enhancers, but since many histone modifications remain to be tested, a single histone modification 'profile' for enhancers has not yet been defined.

As noted by Roh and colleagues, H3 acetylation can be useful in predicting enhancer sites. A more specific genome wide profile of 18 individual acetylation marks showed that H3K18ac was associated with the most enhancers (defined by DNAseI hypersensitivity), associating with about 20% of the enhancer set used in the paper. H3K27ac was the next highest of the H3 acetylations, associating with around 15% (Wang et al., 2008).

The enhancer regions responsible for the expression profile of *Hoxd* are present in all cell types, but are clearly active in very specific locations and times during development to restrict ectopic expression which can cause deformities. How the enhancers are controlled so specifically is not fully understood, but it has been established that the histone modification profiles of enhancers are very variable, depending on the cell type. Whilst H3K4me3, H3K4me1, and H3K27ac remained relatively stable at the promoters of the five different cell types, the profiles on the enhancers were significantly different. For example, a HeLa specific enhancer close to a gene generally meant the gene would have higher expression in HeLa cells than other cell types (Heintzman et al., 2009).

A broader study which looked at 19 different cell types in mouse (including both tissue sections and cell lines) also showed that these histone modifications on enhancers were very cell type specific (Shen et al., 2012). This difference in enhancer histone modification profiles has also been shown in differentiating cell types, implying its importance for *in vivo* developmental control. Although the H3K4me1 profile does differ between cell types (Heintzman et al., 2009), it was shown that some enhancers marked by H3K4me1 were not active. It was therefore proposed that H3K4me1 may mark enhancers which have the potential to become activated later, making it of particular interest in developmental systems. Notably, H3K4me1 is lost on the enhancers of pluripotency genes upon differentiation. This is dependent on LSD1, without which differentiation does not proceed normally (Whyte et al., 2012). Genes close to distal enhancers which are marked by H3K27ac and H3K4me1 have higher expression level on average than genes close to enhancers marked only by H3K4me1, and GO term annotations showed that cell type specific genes have H3K27ac and H3K4me1 marks on the closest enhancer in the correct cell type (i.e. neural specific genes have enhancers marked by both histone modifications in neural cells, but not ES cells)

(Creyghton et al., 2010). 'Poised' enhancers in human embryonic stem cells were also determined to show not only a lack of H3K27ac, but instead, trimethylation of that residue (Rada-Iglesias et al., 2011). Upon differentiation along a neuronal pathway, H3K27me3 was lost, and H3K27ac was gained. Later, three classes of enhancer were defined in mouse ES cells; poised enhancers which contained H3K27me3 and H3K4me1, 'intermediate' enhancers, which contained H3K4me1 but were not covalently modified on H3K27, and active enhancers, with both H3K4me1 and H3K27ac (Zentner et al., 2011) (Fig 1.16B). Poised enhancers could also be distinguished from the intermediate set by presence of H3K9me3. Though earlier studies showed that H3K4me3 was present on regulatory regions (Barski et al., 2007), Zentner and colleagues noted a general depletion of H3K4me3 on enhancers.

Histone modifications are certainly associated with enhancers in a tissue specific manner. It remains to be determined exactly which histone modifications are associated, and potentially more 'classes' of enhancers will be defined as further ChIP-seq datasets become available. Acetylation in particular may be generally associated with enhancers, though H3K27ac has been studied in more detail in this context than other acetyl PTMs. Significantly more work is needed to define the function of histone modifications on enhancers, but it has been suggested that they may contribute to a receptive chromatin environment. That is, the histone modifications keep enhancers poised to maintain their "activation potential" for binding of transcription factors and gene induction upon differentiation/stimulation (Ong and Corces, 2011).

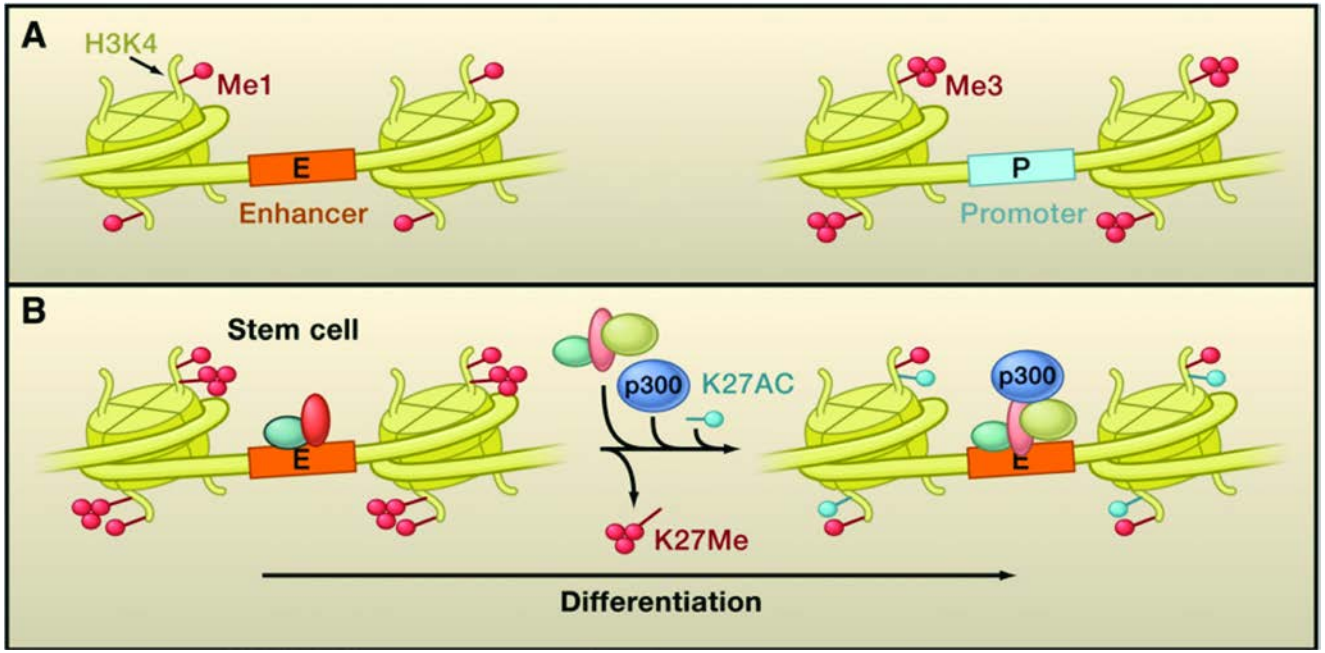


Figure 1.16 *Histone modifications at enhancers and promoters*

A: Red circles indicate methyl groups on H3K4. Monomethylation is found around enhancers, whilst trimethylation is found around the TSS.

B: An example of the subset of enhancers marked by both H3K4me1 and H3K27me3. These enhancers are referred to as “poised”. Upon differentiation, H3K27 gains acetylation in place of trimethylation, and the enhancer becomes active. From (Bulger and Groudine, 2011).

1.7 ChIP and ChIP-sequencing technology

1.7.1 Chromatin Immunoprepitation

Proteins and DNA interact in ways which have functional consequences for processes like transcription, DNA repair and replication and so on. Studying these interactions has been significantly helped by development of the Chromatin Immunoprecipitation (ChIP) technique. ChIP can be used to analyse the localities of transcription factors and other proteins that interact with DNA, but also to isolate the DNA which is wound around nucleosomes with particular PTMs. The principle of ChIP is to use antibodies against the protein or histone modification of interest to isolate both the protein and its bound DNA sequences. The DNA can then be purified and analysed for enrichment of sequences of interest, using conventional or quantitative PCR, microarrays, or most recently, deep sequencing. There are two methods for ChIP, both of which have their advantages and disadvantages (Turner, 2001). More commonly, proteins are bound to the DNA by UV or formaldehyde cross linking prior to shearing of the chromatin by sonication (X-ChIP) (Gilmour and Lis, 1985; Solomon et al., 1988). The protein-DNA complexes are then isolated using antibodies bound to protein beads. This technique works well and is necessary for analysis of transcription factors and other DNA binding proteins which often have transient interactions with their target sequences. However, X-ChIP can be inefficient, especially when working with histone modifications, as formaldehyde cross linking frequently involves lysine groups, which can cause problems with antibody specificity to lysine containing epitopes. This is of particular relevance to this study, concerned with lysine acetylation, but also many other histone modifications of interest. In the second method of ChIP, native chromatin is used (N-ChIP), omitting the cross linking step (Hebbes et al., 1988). For transcription factors this is impractical, as their transient interactions would not survive for the length of the protocol, but for histone modifications, which by their nature are more stably associated with DNA, it is a more useable method (O'Neill and Turner, 2003).

Another advantage of using native chromatin is the fact that there is no epitope loss due to formaldehyde cross linking, which may be the reason that N-ChIP is more efficient in pulling down DNA than X-ChIP. Downstream, this means that analysis can take place on unamplified DNA. Amplification is often a necessary process for X-ChIP, but can introduce

biases, or an increase in background from amplification of non-specific DNA. N-ChIP has some disadvantages, for example nuclease digestion is not without bias, but these can be overcome by using input digested chromatin as a control; this DNA is not enriched for any modification, but is subject to the same digestion bias.

After isolation by ChIP, the DNA is analysed, typically an initial test would involve qPCR on control regions to ensure specificity. This would then be followed by analysis of a broader region of interest using mircoarrays, or an unbiased genome wide analysis using deep sequencing.

1.7.2 Analysing DNA isolated by ChIP

Analysing ChIP DNA using microarrays (ChIP on chip) has advantages over simple qPCR analysis, principally the fact that a much broader region of interest (even up the entire genome) can be analysed (reviewed in (Buck and Lieb, 2004)). By hybridizing the DNA to input, positive biases from nuclease digestion are minimised, and the assay is swift and relatively inexpensive over loci that can be covered on one array, allowing multiple experiments and biological replicates, which may be impossible presently using genome wide sequencing.

Although ChIP on chip can be useful for studying particular loci of interest, the requirement for large amounts of DNA ($\geq$ 100ng per hybridization), and the expense and impracticality of using arrays for genome wide analysis means that in recent years, deep sequencing has come to the fore as a method to map the genome wide location of DNA binding proteins and histone modifications (Park, 2009). A variety of technologies for deep sequencing have become available, including 454 sequencing (Roche) and sequencing by synthesis (Illumina), which is used in this study.

After DNA isolation by ChIP, there are three major steps involved in Illumina's sequencing by synthesis technology (http://www.illumina.com/technology/sequencing_technology.ilmn). The first step in the analysis is the preparation of the libraries. The sheared DNA is size selected, repaired, and adenylated. Adapter sequences are then attached to both ends of the fragment. These adapter sequences match those which are covalently bonded to the flow cell used for sequencing. The second step is cluster generation, which means generating many copies of each piece of DNA to be sequenced. The adaptor sequence bound DNA hybridises to the

‘lawn’ of matching oligos on the flow cell and is copied. The copy is therefore covalently bonded to the flow cell. Each of the potentially hundreds of millions of sequences is copied by extension and bridge amplification. The reverse strands are then cleaved and washed away, and end blockers and sequencing primers are added, resulting in a large number of clusters, ready to be sequenced.

All the clusters are sequenced simultaneously, base by base. Fluorescent (one colour for each base), reversibly terminated nucleotides are added, and are free to compete for binding to the DNA templates during the synthesis reaction. After each round of synthesis, a laser excites the clusters, and an image is taken of the colour pattern on the flow cell, which corresponds to the nucleotide which successfully bonded to the template DNA. Then the blocking group is removed, and the next round of synthesis takes place so that the next base in the template sequences can be detected.

After synthesis is complete, the images are analysed and converted into sequences, which can then be mapped to the genome assembly of the organism from which they came. The data are then used in downstream analyses, such as peak calling, or visualisation of the modification/DNA binding protein enrichment over a particular locus of interest.

1.8 PhD Aims

Research on H4K16ac has largely focused on the association of the mark with transcriptional activation and euchromatin *in vivo*, whilst studies into its effect on chromatin compaction have been restricted to *in vitro* work. As H4K16ac clearly has an important role in mammalian development, it is possible that it may be linked to changes in chromatin compaction associated with differentiation.

Additionally, studies of the interaction partners of MOF have revealed that in mammals, as well as flies, it is part of two complexes which are functionally different. However, these studies have all taken place using overexpression of tagged proteins in human cancer cell lines, rather than endogenous protein.

The aim of this thesis is to:

- Use differentiation of ES cells along the neuronal pathway as a model, and to determine how global H4K16ac profile changes upon differentiation in conjunction with expression.
- To look for large scale changes in H4K16ac upon differentiation and determine whether these can be used to predict changes in chromatin compaction state.
- To use antibodies against endogenous MOF in ES and somatic cells to determine whether MOF interaction partners differ between different cell types.

Chapter 2: Materials and Methods

2.1 Stock solutions, reagents, and buffers

Water used in solutions is deionised, and solutions were stored at room temperature unless otherwise stated.

Acetate Buffer: 3M Potassium Acetate, 11.5% glacial acetic acid (v/v).

Antibiotic Stock Solutions:

Chloramphenicol, 25µg/µl (w/v in ethanol), stored at -20ºC.
Ampicillin, 100mg/mL (w/v in water), aliquoted and stored at -20ºC.
Kanomycin 35 mg/mL (w/v in water), aliquoted and stored at -20ºC
Puromycin 10mg/mL (w/v in 20mM HEPES buffer), stored at -20ºC (purchased prepared and sterile filtered from Invitrogen, cat#A11138)
Zeocin/G418 50mg/mL (w/v in water), stored at 4ºC (purchased prepared and sterile filtered from Sigma, cat# G8168)

ChIP Blocking Solution: 1 x PBS, 0.5% BSA (w/v), premade and stored at -20ºC.

ChIP Elution Solution: 0.1M $NaHCO_3$, 1% SDS, made fresh before each use. For X-ChIP, elution solution was passed through a 0.2 micron filter prior to use.

N-ChIP Wash Buffers: Wash 1, 150mM NaCl, 10mM TrisHCl pH8, 2mM EDTA, 0.05% Triton X-100 (v/v); Wash 2, 10mM TrisHCl pH 8, 1mM EDTA, 150mM LiCl, 0.1% sodium deoxycholate (w/v). Washes were made in the absence of detergents at 10 x concentration and stored at room temperature. Protease Inhibitor Cocktail (Calbiochem, cat #539131), 5mM sodium butyrate, 0.2mM PMSF, 1mM DTT, and 5µM Sirtinol were added to both 1 x washes immediately prior to use.

Colloidal Coomassie Stain: Colloidal Blue Staining Kit was purchased from Life Technologies (cat #LC6025) and used as specified by manufacturer instructions. Destain was performed with water.

Coomassie Destain: 30% methanol (v/v), 10% glacial acetic acid (v/v).

Coomassie Stain: 0.25% Coomassie brilliant blue R-250 (w/v) dissolved in 45% methanol (v/v), with the addition of 10% glacial acetic acid.

DNA Loading Buffer: 50% glycerol, 5mM EDTA pH8, 0.3% Orange G (v/v).

Farnham Lysis Buffer: 5mM PIPES pH8, 85mM KCl, 0.5% NP-40 (v/v), Protease Inhibitor Cocktail was added immediately prior to use. Buffer was made fresh each use, and passed through a 0.2 micron filter.

FISH Hybridisation Mix: 50% deionised formamide (v/v), 10% dextran sulphate (v/v), 1% Tween 20 (v/v), prepared in 2x SSC.

Fixation Solution (2D FISH): 75% methanol (v/v), 25% glacial acetic acid (v/v), prepared fresh before each use.

Glutathione Elution Buffer: 50mM Tris-HCl pH8, 10mM reduced glutathione.

GTE Buffer: 50mM glucose, 25mM Tris-HCl pH8, 10mM EDTA, stored at 4ºC. Approximately 10mg lysozyme/mL was added prior to use.

Hypotonic Buffer (2D FISH): 33mM KCl, 17mM tri-sodium citrate.

Luria agar*: Per litre: 10g tryptone, 5g yeast extract**,** 10g NaCl, 15g agar.

Luria-Bertani (LB) Broth*: Per litre: 10g tryptone, 5g yeast extract, 10g NaCl, 1g glucose.

Lysis Buffer (alkaline lysis mini-prep): 0.2M NaOH, 1% SDS (v/v), made fresh prior to each use.

Micrococcal Nuclease (MNase): Definition of one unit of MNase differs between suppliers. Boehringer Mannheim (Roche) Units are used throughout this work. MNase was purchased from Sigma (500 Sigma Units/vial, cat #N3755-500UN), and resuspended in MNase Nuclease Buffer to 50 Roche units/µl.

MNase Buffer: 50% glycerol, 10mM Tris pH7.6, 50mM NaCl.

MNase Stop Buffer: 215mM NaCl, 10mM TrisHCl pH8, 20mM EDTA, 5.5% sucrose (v/v) 0.2mM PMSF, 0.2% Triton X-100.

Nick Translation Salts: Made as a 10x solution, which was aliquoted and stored at -20ºC. 0.5M Tris pH7.5, 0.1M $MgSO_4$, 1mM DTT, 0.5mg/mL BSA fraction V (Sigma).

NP-40 Lysis Buffer: 50mM TrisHCl pH8, 150mM KCl, 1.5mM $MgCl_2$ 0.1M DTT, 0.2mM EDTA. Aliquoted at stored at -20ºC. Protease inhibitor cocktail (Calbiochem) was added immediately prior to use.

Nuclear Buffer A (NBA): 85mM NaCl, 5.5% sucrose, 10mM TrisHCl pH7.5, 0.2mM EDTA. 0.2mM PMSF, 5mM sodium butyrate, 5µM Sirtinol and 1mM DTT added fresh prior to use.

Nuclear Buffer B (NBB): 85mM NaCl, 5.5% sucrose, 10mM TrisHCl pH7.5, 0.2mM EDTA, 0.2mM PMSF, 5mM sodium butyrate, 5µM Sirtinol, 1mM DTT and 0.1% NP-40 added fresh prior to use.

Nuclear Buffer R (NBR): 85mM NaCl, 5.5% sucrose, 10mM TrisHCl pH7.5, 3mM $MgCl_2$, 1.5mM $CaCl_2$, 0.2mM PMSF, 5mM sodium butyrate, 5µM Sirtinol and 1mM DTT added fresh prior to use.

Phosphate Buffered Saline (PBS)*: 160mM NaCl,3mM KCl, 8mM Na_2HPO_4, 1mM KH_2PO_4 from tablets (Oxoid, ThermoScientific, cat# BR0014). PBS was made by technical services at MRC HGU, except for PBS used in embryonic stem cell culture.

Proteinase K: 20mg/mL (w/v) in 50mM Tris (pH8), 1.5mM calcium acetate, 50% glycerol (v/v). After preparation, proteinase K was aliquoted and stored at -20ºC.

RIPA Buffer: 1% NP-40 (v/v), 0.5% sodium deoxycholate (v/v), 0.1% SDS in 1x PBS. Protease Inhibitor Cocktail added immediately prior to use. Buffer made fresh each use, and passed through a 0.2 micron filter. For use in SILAC IP Benzonase (20U/mL, Novagen) and DNAseI (200U/mL, Invitrogen) were added prior to use.

Saline Sodium Citrate Buffer (SSC)*: Per litre for 20x solution: NaCl 175.3g, $Na_3C_6H_5O_7$ 88.2g.

SDS Loading Buffer: 50mM TrisHCl pH 6.8, 2% SDS (w/v), 10% glycerol (v/v), 1% β-mercaptoethanol, 12.5mM EDTA, 0.02% bromophenol blue (w/v). Made as 5x stock, aliquoted and stored long term at -20ºC, and short term at 4ºC.

Sodium Dodecyl Sulphate*: Prepared as 20% stock (w/v).

Sonication Buffer (SB): 2mM NaCl, 5M Urea.

Swelling Buffer (SILAC IP): 10mM Hepes pH9.0, 1.5mM $MgCl_2$, 0.5mM DTT. Buffer prepared fresh for each experiment, with addition of protease inhibitor cocktail immediately prior to use.

* Prepared by Technical Services at MRC HGU

Tris-Acetate EDTA (TAE) Buffer*: 40mM Tris base, 20mM acetic acid, 1mM EDTA. Prepared as 50x stock.

TE Buffer*: 10mM Tris-HCl pH 7.6, 1mM EDTA.

Tris Buffered Saline (TBS): 50mM Tris-HCl pH 7.4, 150mM NaCl. Made as 10x stock.

Tris-Glycine SDS-Page Running Buffer: 25mM Tris base, 250mM glycine pH 8.3, 0.1% SDS (w/v). Prepared as 10x stock.

Trypsin: Trypsin-EDTA solution at 10x concentration (0.5% Trypsin, 0.2% EDTA) was purchased from Sigma (cat #59418C), aliquoted and stored long term at -20ºC, and short term at 1x concentration at 4ºC.

Western Transfer Buffer: 25mM Tris base, 200mM glycine, 20% methanol (v/v), 0.02% SDS (v/v). Made fresh prior to use.

X-ChIP Lithium Chloride Wash Buffer: 100mM Tris pH 7.5, 250mM LiCl, 1% NP-40 (v/v), 1% sodium deoxycholate (v/v). Made fresh each use, and passed through a 0.2 micron filter.

X-ChIP PBS/BSA: 5mg/mL BSA (fraction V), prepared in 1x PBS fresh for each use, and passed through a 0.2 micron filter.

Table 2.1 Primers used.

	Primer Set Name	Forward Sequence	Reverse Sequence
ChIP qPCR	ActinPromoter	CCTCGATGCTGACCCTCATCC	GACACTGCCCCATTCAATGTCTC
	Olig2Promoter	GCCTGACGCTACAGTGACAA	GGCTAATTCCGCTCAATGAA
	Hoxb1Promoter	TTAGCCCATTGGCCTGGGAGAGAT	TGAAGCTTGAGCTTGAGCCCATGGCCCG
	Hoxb13Promoter	ATGAGCCTCTCTCCCCCAGG	AATCGCTCCCAGCTCGAACGG
	Hoxd1Promoter	GAGTAACTTGACCTTCTCAGAG	ATTGCGGGAGAAAGGCAGGGAAG
	Pou5f1Promoter	GCTGGCGGAAAGACACTAAG	CAGAGCATGGTGTAGGAGCA
Reverse Transcription PCR	MofExpression	ACAACTTCAGGAACTGTAGG	TCGCCTGTTAAAGCCC
	Hoxb2Expression	AGGGAGCCAAAAGCAAGCC	CGCCGATTCTGGCAATCC
	Olig2Expression	AGAGCCAGGTTCTCCTCCGC	ACTAGACACCAGGCTGGCGTC
	Spry4Expression	GCAGCTCCTCAAAGACC	TGGAGCCGGCTGTGGG
	Lrrc2Expression	AGCGCCCTGGAGAAGATA	AGGCAGCTCCTTCCAC
	Cyp26b	ATTGGCGACATCCACCG	GCACGGCCATTCGGAAGGT
	Hoxd1	CCACAGCACTTTCGAGTGGA	ACTCTTTCTCTAGCTCTGTCAG
	Pou5f1	CGAGAACAATGAGAACCTTC	CCTTCTCTAGCCCAAGCTGAT
	GAPDH	TGGTGAAGGTCGGTGTGAACG	TGAGTGGAGTCATACTGG
	Hoxb1Expression	ACCTCCTCTCTGAGGACAAGG	AGTCCCAGCTCGGACACCTTC
	Hoxb13Expression	CATTCTGGAAAGCAGCGTTTGCAG	GATAACTTGTTGGCTGCATACTCC
Cloning	MOFFlag	AGTTAATTAATGGACTACAAGGACGACGATGACAAGGCGGCACAGGGAGCTACAGC	AAGCTAGCTCACTTCTTGGAAAGCTTGACTTG
	MOFChromo	CACCGGATCCTGGCATTCAGCAG	GTTGCGGCCGCTCACTTCTGCAC

	MOFBam1Not1Sites	GTAGCGGCCGCTCACTTCTTGGAAAGCTT	ACGCGGATCCATGGACTACAAGGACGACGATG
Sequencing	MofM	AGAACCTGTGTCTACTGGC	AGTAGACACAGGTTCTGAC
	SP6	ATTTAGGTGACACTATAG	n/a
	T7	TAATACGACTCACTATAGGG	n/a

Table 2.2: Fosmid Probes

	Region	name	Ensembl name	start	end
Alpha globin	I19r	WI1-2837A17	G135P60495H4	32056569	32100774
	Hbq1	WI1-2903N21	G135P603718B2	32196269	32241313
Sox2	Sox2Up	WI1-1766B8	G135P601568H9	34503694	34543065
	Sox2Down	WI1-1185J11	G135P603742D7	34686049	34722693
Nanog	NanogUp	WI1-2049I9	G135P604583D8	122574082	122610708
	NanogDown	WI1-196C18	G135P62993G4	122916571	122959014

Names are ensembl (r45) (http://jun2007.archive.ensembl.org/Mus_musculus/index.html). Alpha globin fosmids were previously used in (Eskeland et al. 2010).

Table 2.3: Plasmids used

Plasmid	Source	Antibiotic Resistance	Use	Reference
pTLC	gift from Josh Brickman lab	Ampicillin	FLAG-MOF overexpression	Eskeland et al, 2010
pGAGASIZXN (pCAGxn)	pCAGASIZ plasmid with deletion of Alkaline Phosphotase (N. Gilbert)	Ampicillin	FLAG-MOF overexpression	
pGEX6P1	GE Healthcare, cat#28-9546-48	Ampicillin	MOF Chromodomain bacterial expression	
pEGFP-N1	Clontech, cat#6085-1, gift from Ian Adams lab	Kanomycin	Positive control for GFP in FACS sorts	
CAG-Cre	Addgene plasmid 13775	Ampicillin	Floxing of pTLC plasmid	

2.2 Methods

Centrifugation steps were performed at room temperature unless otherwise stated.

2.2.1 Cell Culture

All mammalian cells were cultured in 5% CO_2 at 37ºC unless otherwise stated.

ES cells:

Low passage cells (<40) were used throughout.

Embryonic Stem cell (ES cell) strains used were OS25 (Billon et al. 2002) and E14 (Hooper et al. 1987). OS25 ES cells are derived from E14 cells and contain the *hygromycin thymidine kinase* gene targeted to the *Pou5f1* locus. Undifferentiated cells were selected for by inclusion of Hygromycin in media.

OS25s were cultured in Glasgow Minimal Essential Medium (GMEM) (Gibco) supplemented with 10% (v/v) foetal calf serum (FCS), 0.1mM β-mercaptoethanol, penicillin (10,000 units/mL) and streptomycin (650µg/mL), 0.3mg/mL L-Glutamine, 1mM sodium pyruvate, and non-essential amino acids (Sigma). For undifferentiated cells, 1000 units/mL human recombinant LIF and 0.1mg/mL hygromycin (Roche) was added to the media.

E14s were cultured in GMEM as above, but supplemented with 15% (v/v) FCS, and grown without hygromycin.

ES cells were cultured on either 0.1% gelatin coated tissue culture flasks (Corning) or without gelatin on hydrophilic culture flasks (Cell+ culture flask, Sarstedt). Media was replaced every 24 hours, and cells were passaged every 72 hours and seeded at a density of $0.1x10^6$ cells/mL.

For differentiation using retinoic acid (RA), undifferentiated ES cells were passaged and seeded at a density of $0.2x10^6$ cells/mL into GMEM supplemented with 10% (v/v) FCS, 0.1mM β-mercaptoethanol, penicillin (10,000 units/mL) and streptomycin (650µg/mL), 0.3mg/mL L-Glutamine, 1mM sodium pyruvate, and non-essential amino acids (Sigma). After 24 hours, RA was added to a concentration of $5x10^{-6}$M. 48 hours after seeding, cells were checked for morphological changes (an elongated shape, and a lack of rounded colonies), and fresh media containing RA was added. Cells were used in experiments 72 hours after seeding (D3 differentiated).

Other:

Human HEK 293T (293T) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% FCS, penicillin (10,000 units/mL) and streptomycin (650µg/mL). Media was replaced every 2-3 days, and cells were passaged shortly prior to 100% confluence.

Mouse limb mesenchymal 14fp cells were kindly supplied by Iain Williamson and Bob Hill (Williamson et al., 2012). 14fp cells were cultured at 33ºC (permissive temperature for temperature sensitive T-antigen) in DMEM supplemented with 10% FCS, penicillin/streptomycin, and 20ng/mL γ-Interferon (Peprotech). Media was changed every 3 days and cells were passaged prior to 100% confluence.

Transfection:

For transient transfection, ES cells and 293T cells were transfected using Lipofectamine LTX plus reagent (Invitrogen, cat#15338) according to manufacturer instructions.

For stable transfection of pTLC-MOF plasmid into E14 ES cells, approximately 15µg linearized plasmid was cleaned by double phenol:chloroform extraction, and resuspended in tissue culture quality PBS (Sigma). Approximately $0.8x10^7$ cells were used per electroporation, at a concentration of $1x10^7$ cells/mL in a 0.4cm electroporation cuvette (BioRad) on a Gene Pulser II (BioRad). Electroporation conditions are as follows; High Capacitance (>50µF), infinite resistance, 250V, 500uF, time constant between 5-8 ms.

Cells were seeded into two gelatinised 10cm tissue culture dishes (Nunc, ThermoScientific), puromycin selection (1.2µg/mL) began after 24 hours and continued with daily media replacement until colonies had formed in positive samples and all cells in non-transfected controls were dead (approximately 9 days). Colonies were picked and grown under selective conditions, then analysed for RFP fluorescence by Fluorescence Activated Cell Sorting (FACS).

pTLC-MOF plasmid was floxed by introduction of pCAG-Cre to cells by transfection using Lipofectamine LTX plus reagent (Invitrogen). After 5 hours incubation with lipofectamine and plasmid, the cells were split harshly to encourage colony growth. After 5 days, the plate was trypsinised and cells analysed for GFP fluorescence by FACS.

2.2.2 Bacterial Culture

Genomic clones were supplied by BacPac Resources Centre at the Children's Hospital Oakland Research Institute (http://bacpac.chori.org/). For FISH experiments, fosmid clones were used from the WIBR2 library derived from a human female (Broad Institute) in a pEpiFOS-5 vector in DH10B T1 resistant competent cells, and supplied as stab cultures in agar. Clones used are listed in Table 2.1.

Bacteria (from stab cultures or frozen glycerol stocks) were streaked out onto LB agar (section 2.1) to form single colonies and grown overnight at 37ºC; antibiotics were added to cool, melted agar prior to plates being poured. Fosmids were grown on agar supplemented with 25µg/mL chloramphenicol. Plasmids containing ampicillin resistance were selected for with ampicillin at 100µg/mL. Plasmids containing kanamycin resistance were selected for with kanamycin at 50µg/mL.

To prepare DNA, a single colony was picked from an LB agar plate and used to inoculate 3-5mL LB broth (antibiotics used as required; for fosmids, 12.5µg/mL chloramphenicol was used in LB broth, for plasmids, concentrations are as above). Cultures were then incubated with shaking at approximately 300rpm (InnOva 4230 incubator, New Brunswick Scientific) at 37ºC overnight with a 5:1 air to liquid ratio.

To prepare larger amounts of DNA, 1mL of inoculated LB broth (cultured overnight as above) was used to inoculate 200mL LB broth supplemented with appropriate antibiotics, and incubated for not more than 16 hours at 300rpm at 37ºC with 5:1 air to liquid ratio.

2.2.3 Glycerol Stocks

Glycerol stocks of bacteria were prepared by adding glycerol to a concentration of 40% v/v to 1mL of an overnight culture, and stored frozen at -70ºC.

2.2.4 Protein overexpression in bacteria

The chromodomain of mouse MOF (as defined by NCBI protein) was amplified by PCR with inclusion of BamHI and NotI restriction sites, and cloned into a pGEX6P1 plasmid for expression with glutathione-S-transferase (GST) in *E. coli*. Insertion was confirmed by size separation of digested DNA on an agarose gel, and the insert was sequenced on a 3130/3170 genetic analyser (Applied Biosystems) by technical services at MRC HGU. Plasmid was transformed into bL21-RIPA competent bacteria (transfection proceeded as below, with the exception that SOC media was heated to 42ºC before addition to heat shocked bacteria). A colony of transformed bacteria was grown in 10mL LB broth containing 100µg/mL ampicillin and 50µg/mL chloramphenicol for 3 hours at 37ºC until an Optical Density (OD_{600}) of 0.6, as measured on an Ultraspec 3000pro spectrophotometer. Protein production was induced with 0.5mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 hours at 37ºC and cells were lysed with NP-40 lysis buffer, proteins separated on an SDS-PAGE gel, and analysed by Colloidal Coomassie and western blot with α-GST primary antibody (Abcam, cat#ab6613, used at a dilution of 1:5000) to confirm expression of specific protein. 1mL of uninduced culture was then used to inoculate 500mL LB broth containing ampicillin and chloramphenicol until an OD_{600} of 0.6, then induced with 0.2mM IPTG for 3 hours at 37ºC.

2.2.5 Preparation and handling of DNA

Extraction of DNA from overnight bacterial culture:

Fosmid DNA was extracted using alkaline-lysis miniprep. Approximately 3mL of cultures were pelleted at 16,000g for 30s, then resuspended in 200µl GTE buffer for 5 minutes before addition of 400µl ice cold lysis buffer and incubation on ice for 5 minutes. 300µl of acetate buffer was added, and the preparation was incubated on ice for a further 5 minutes. The flocculent precipitate was centrifuged at 16,000g for 5 minutes at 4ºC, and the clear supernatant was removed to a fresh eppendorf. Phenol:chloroform extraction was then performed by adding an equal volume phenol-chloroform to the sample, centrifuging (at 16,000g for 2 minutes at 4ºC), and removing the top layer. The DNA was precipitated with an equal volume of isopropanol, and incubated for >1 hour at -20ºC, then pelleted by centrifugation at 16,000g for 15 minutes at 4ºC. The DNA was then washed in 70% ethanol, re-pelleted as previously for 5 minutes, and finally resuspended in approximately 30µl of TE, and stored at -20ºC.

Plasmid DNA was extracted using a Qiagen mini- or maxiprep kit, according to manufacturer instructions.

Restriction Digestion and ligation of DNA:

Plasmid DNA was digested using appropriate restriction enzymes (all purchased from New England Biolabs and used according to manufacturer instructions). Typically 1-2 units of enzyme were used to digest ≤ 1ug DNA which was diluted in the required buffer. Digestion was carried out for 1 hour at 37ºC.

Ligation of insert DNA into vector DNA was performed using T4 ligase (NEB) as manufacturer instructions. A 3:1 molar ratio of insert to vector was used, and the reaction was incubated overnight at 4ºC.

Transformation of competent bacteria:

Approximately 10ng plasmid DNA was added to 50µl Library Efficiency DH5α Competent Cells (Invitrogen, cat#18262-012) and incubated on ice for 45 minutes to 1 hour. The cells were then subjected to heat shock at 42ºC for 45 seconds, and incubated on ice for 2 minutes. 900µl SOC medium (Invitrogen) was added and the cells were incubated for 1 hour at 37ºC shaking at 300rpm. The cells were then plated on LB-agar containing appropriate antibiotics.

Quantification of DNA using spectrophotometry:

Throughout experiments, DNA concentration was measured using spectrophotometry with a Nanodrop ND-1000 (Nanodrop Technologies) or a Nanodrop 8000 (Thermoscientific) according to manufacturer instructions. Absorbance of sample at a wavelength of 260nm (A_{260}) was measured. An A_{260} measurement of 1 equals a double stranded DNA concentration of 50µg/mL.

Resolution of DNA on agarose gels:

Routinely, DNA was resolved on gels made to appropriate agarose percentage ("Hi-Pure" Low Eeo agarose, Biogene UK, w/v) in TAE with 0.5µg/mL ethidium bromide. 5x DNA loading buffer was added to the sample to result in a final 1x concentration. Commercial DNA size marker ladders (NEB, Invitrogen) were used for reference, and gels were visualised under UV light.

For analysis of MNase digest ladder for ChIP sequencing, input chromatin was resolved on a 1.3% agarose gel with no ethidium bromide, stained with SYBR Gold (Invitrogen) according to manufacturer instructions, and scanned using a 532nm laser with a 575nm Long Pass filter (LPG filter on Fuji FLA-5100 phosphoimager).

Extraction of DNA bands from agarose gels:

Gel slices containing correct size DNA fragment were cut out of the agarose gel using a clean razor blade, and DNA was extracted using a QIAquick Gel Extraction kit (Qiagen, cat# 28704) according to manufacturer instructions.

2.2.6 Preparation of RNA and cDNA for Expression Analysis:

RNA Extraction:

Total RNA was isolated from cells using Trizol reagent (Invitrogen) according to manufacturer instructions. Briefly, 1mL of Trizol per 5-10x10^6 cells was added directly to the culture flask, and incubated for 5 minutes at room temperature. Per 1mL Trizon, 200µL chloroform was added, the tubes mixed vigorously, incubated at room temperature for 2 minutes and centrifuged at 12,000g for 15 minutes at 4ºC. The aqueous phase was removed and precipitated with isopropanol at room temperature for 10 minutes, then centrifuged at 12,000g for 10 minutes at 4ºC. The pellet was washed in 75% ethanol, centrifuged as previously for 5 minutes, and dissolved in 25-50µl H_2O. Concentration was estimated using Nanodrop (A_{260} of 1 equals 40µg/mL concentration of RNA).

RNA was then DNAse treated with 2 units of DNAseI (Ambion), for 30 minutes at 37ºC, and analysed for quality on an Agilent 2100 Bioanalyser with an RNA 6000 Nano chip according to manufacturer instructions (analysis was performed by technical services at MRC HGU). Only RNA with a RIN number of >8 was used in subsequent experiments.

cDNA Preparation:

cDNA was prepared using a First Strand cDNA Synthesis kit (AMV) (Roche, cat#11483188001) according to manufacturer instructions, and including a negative control reaction without reverse transcriptase. Samples which failed the negative control were assumed to contain genomic DNA and discarded.

2.2.7 Expression Analysis:

RT-PCR:

cDNA generated using First Strand kit (above) was amplified by PCR using primers for gene of interest (Table 2.1) with Reddymix PCR Mastermix (Thermoscientific, cat# AB0575) according to manufacturer instruction. Reddymix contains Thermoprime Plus DNA Polymerase, buffering salts, and dNTPs, in addition to a precipitant and dye for subsequent electrophoresis.

PCR reaction was incubated at 95ºC for 15 minutes, followed by 30 cycles of amplification (95ºC for 20 seconds, 56ºC for 30 seconds, 72ºC for 30 seconds), and products were checked by electrophoresis (see section 2.2.4).

Gene Expression Microarrays:

Illumina:

For Illumina expression arrays, RNA was extracted from 3 biological replicates of undifferentiated ES cells, and 400ng of RNA were amplified using the Illumina TotalPrep RNA Amplification Kit according to manufacturer instructions (Ambion, cat#AMIL1791), and checked for quality using Agilent 2100 Bioanalyser as described above. Amplified, biotinylated cRNA was labelled and hybridized to Illumina MouseRef6 Gene Expression beadchip arrays and run on an Illumina Beadstation by the Wellcome Trust Clinical Research Facility.
Probe level expression data was analysed by Graeme Grimes (MRC HGU) using R Bioconductor packages Limma and beadarray. Genes were sorted according to probe detection level, and the top 500 were designated as highly active genes, whilst the lowest 500 were designated as genes with little expression, and used in downstream analysis.

Agilent:

1ug of total RNA from 3 biological replicates of undifferentiated and differentiated (3 days in Retinoic Acid) was amplified using one round of amplification with an Amino Allyl MessageAmp II aRNA Amplification kit (Ambion, cat#AM1753) according to manufacturer instructions, and checked for quality on the Agilent 2100 Bioanalyser. The RNA was dye coupled to Cy3 using the same kit, and for 3 technical replicates per biological replicate, 1.34µg cRNA was hybridized to Agilent whole mouse genome 4x44k oligomicroarrays (Agilent cat# G4122F) according to manufacturer instructions (hybridisation proceeded for

17 hours). Arrays were scanned on a Nimblegen MS200 scanner at 5µm resolution and spot intensities were extracted using Array-Pro Analyser (by Jurriaan Hölzenspies, University of Copenhagen). Quantile normalisation and calculation of $\log_2$ fold change and adjusted p values were performed in Limma by Graeme Grimes (MRC HGU). Genes differentially expressed with adjusted p value of <0.005 were used for H4K16ac tag density analysis.

A selection of genes highly upregulated in UD or D3 samples was checked by RT PCR (see above) using primers detailed in Table 2.1. To ensure global gene expression changes were consistent with expected results from a retinoic acid differentiation, the 7758 differentially expressed genes (5190 upregulated in undifferentiated cells and 2568 upregulated in D3 cells) with an adjusted p-value of <0.005 were searched for enriched gene ontology (GO) terms against a background of total probes on array, using Gene Onotology enrichment analysis and visualisation tool GOrilla (Eden et al. 2009). To avoid false positives, a P value of 10^{-7} based on Bonferroni correction was used as the cut off for GO terms to be considered.

2.2.8 Chromatin Immunoprecipitation (ChIP)

Native ChIP:

Approximately $2x10^8$ ES cells were treated for 2 minutes prior to trypsinisation with 5µM sirtinol and 5mM sodium butyrate. After trypsinisation, the cells were washed twice in ice cold PBS containing 5µM sirtinol and 5mM sodium butyrate, and pelleted at 200g for 4 minutes at 4ºC. The cell pellet was resuspended in 5mL NBA buffer, and 5mL NBB buffer were added. The cells were incubated on ice for 3 minutes, then centrifuged for at 400g for 4 minutes at 4ºC to pellet nuclei. Nuclei were resuspended in 10mL NBR buffer and centrifuged again at 400g for 4 minutes at 4ºC.

Chromatin concentration was measured by taking an aliquot of chromatin diluted 1:20 in NBR buffer and treating with DNAseI for 5 minutes. 400µl of Sonication Buffer were added, and absorption at 260nm was measured on an Ultraspec 3000pro spectrophotometer. 500µg of chromatin were used in each IP, and digested with 40-50 Boehringer units of MNase (Sigma) to generate a chromatin ladder enriched in tri-, tetra-, and pentanucleosomes for qPCR analysis, or 70-100 units to generate a chromatin ladder enriched in mono-, di-, and trinucleosomes for deep sequencing (which requires smaller fragments of DNA). MNase digestion proceeded at room temperature for exactly 10 minutes, and was stopped by addition of an equal volume of stop buffer, and incubated on ice at 4ºC overnight.

Between 50-150µg of released chromatin were incubated with 10ug anti-H4K16ac (Millipore, cat#07-329) or Rabbit IgG (Santa Cruz, cat#sc-2027) prebound to Protein A Sepharose beads (Amersham cat#17-5280-01) with 100µg BSA for 3 hours at 4ºC. Beads were washed twice with wash buffer 1, once with wash buffer 2 for 10 minutes at 4ºC, and once in TE. Bound complexes were eluted from beads using elution solution. 10% input and immunoprecipitated chromatin were adjusted to pH8 using 2M TrisHCl pH6.8, and treated with 40µg Proteinase K (Genaxxon) at 55ºC for 1 hour. DNA was purified using a Qiagen PCR purification kit, according to manufacturer instructions, and quantified on Nanodrop.

Cross Linked ChIP

In order to compare H4K16ac profile in 14fp limb cells with existing data for H3K27ac and H3K4me1, it was necessary to generate a cross linked ChIP dataset for H4K16ac in this cell line.

Approximately $1.5x10^8$ cells were cross-linked with 1% formaldehyde (v/v) (Sigma, cat#F87750) for 10 minutes at room temperature. Cross-linking was halted by addition of 0.125M glycine for 5 minutes. Cells were then washed with ice cold PBS, and removed from the tissue culture flask with a cell scraper, then pelleted at 400g for 5 minutes at 4ºC. Samples were subsequently kept on ice. Cells were resuspended in 5mL Farnham Lysis Buffer, incubated for 5 minutes on ice, then pelleted at 420g for 5 minutes at 4ºC and resuspended in 5mL ice cold RIPA buffer for sonication. Sonication was carried out with the resuspended nuclei in 1mL aliquots, using a Bioruptor Next Gen (Diagenode) for 60 minutes with 30 second on off cycle at 4ºC. Samples were spun down at 20,000g for 15 minutes at 4ºC and supernatant collected. Sonication efficiency was checked with a 20µl sample of chromatin treated with 10µg RNAse A at 37ºC for 15 minutes, followed by treatment with 10µg Proteinase K at 65ºC for one hour, and visualisation on an agarose gel. Samples which were over-sheared (significant portion of chromatin <100bp) or under-sheared (>1000bp) were discarded. Chromatin concentration was checked on a spectrophotometer.

For IP, magnetic Protein A Dynabeads (Invitrogen, cat#100-02D) were briefly pre-blocked using 5mg/mL BSA in PBS, then coupled to 5µg H4K16ac or Rabbit IgG antibodies (as above) for 1 hour at 4ºC, then washed in BSA/PBS as above, and resuspended in BSA/PBS. 200µg of chromatin were incubated with antibody-coupled beads overnight at 4ºC. Beads were collected using a magnet and washed five times with 3 minute incubations at 4ºC using LiCl wash buffer, and once in TE. Bound material was eluted from beads using 200µl elution solution at 65ºC for 1 hour, with intermittent vortexing. Supernatent was then collected and treated with RNAse A for 30 minutes at 37ºC before addition of 40µg Proteinase K and incubation at 65ºC for 5 hours for reversal of formaldehyde cross-links.

The DNA was then purified using a Qiagen PCR purification kit, according to manufacturer instructions, and quantified on Nanodrop.

2.2.9 ChIP Analysis

qPCR

For relative quantification of ChIP DNA by real-time qPCR, DNA isolated from 10% of total MNase digested native chromatin input was used to generate a standard curve for immunoprecipitated H4K16ac and IgG samples. For each of the three biological replicates, PCR was performed in triplicate using Quantitect SYBR Green detection kit on a Biorad CFX96 Real Time System C1000 Thermal Cycler using primers described in Table 2.1. Thermal cycler programme: 15 minutes hotstart at 95°C, and 45 times 95°C for 20s, 56°C for 30s, 72°C for 30s and plate read. Analysis was performed in Biorad CFX Manager version 2.0.

ChIP on chip

For H4K16ac, in UD and D3 ES cells, array analysis was previously performed for 3 biological replicates with dye swaps (Gillian Taylor, MSc thesis, 2009) on custom arrays printed at Liverpool Microarray Facility (Eskeland et al. 2010), and so only one biological replicate was performed on the custom Nimblegen arrays used in this work. For 14fp cells, two biological replicates with dye swap were performed.

400ng (for OS25s) or 500ng (for 14fp) ChIP/Input DNA were labelled with Cy3/Cy5 by random priming according to NimbleGen ChIP-chip protocol (Roche). Briefly, the DNA was denatured with the Cy3/Cy5 random nonamers at 98ºC for 10 minutes, then incubated on ice for 2 minutes. The sample was then incubated with dNTPs and Klenow fragment at 37ºC for 4.5 hours (for OS25s) or 3 hours (for 14fp). An equal amount (15µg) of ChIP/Input DNA was combined for each experiment, and hybridized to custom 3x720K NimbleGen mouse tiling arrays (Roche) (genomic regions covered by arrays based on the mm9 (NCBI37) mouse genome build are listed in Appendix 1). Hybridisation proceeded for 20 hours at 42ºC, and arrays were scanned on a NimbleGen MS200 Microarray Scanner (Roche) using 100% laser power at 2µm resolution. Signal intensities were quantified from raw TIFF images with MS200 Data Collection software.

Data processing was performed by Graeme Grimes (MRC HGU) in R. Mean signal was taken for duplicate probes on the array, and scale normalisation was applied between biological replicates, with loess normalisation to correct for dye bias.

Log_2 ChIP/input was calculated by Log_2 of the product of division of ChIP signal by input signal. Analysis was performed in Bioconductor in R using Epigenome (PROT43 protocol). The R Ringo sliding quantile function was used to generate the mean Log2 ChIP/input score over 1000bp sliding windows across the array.

ChIP sequencing

Library preparation and Next Generation Sequencing of ChIP and Input DNA were performed by GenePool (University of Edinburgh) on an Ilumina Genome Analyser IIx. Library preparation and sequencing proceeded according to the standard Illumina protocol, with paired end sequencing of 50bp read length from either end of the fragment. Input DNA was sequenced as a control for nucleosome occupancy, MNase digestion bias, and sequencing bias. Briefly, DNA was end repaired, and polyadenylated. Illumina paired end adapters were ligated to the ChIP DNA, and the ligated DNA was purified and amplified, and library fragments of >200bp were gel purified, then captured on the Illumina flow cell for cluster generation and sequencing.

Processed image data was converted to FASTQ format (Solexa v1.3) by GenePool, and then used for further processing in-house.

ChIP sequencing analysis

I thank James Prendergast, Rob Illingworth, John Thomson and Duncan Sproul for their generous assistance with analysing ChIP-seq datasets.

In addition to the dataset generated in this study, the following publicly available datasets were downloaded in SRA lite format from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/)*:* H3K4me1, H3K27ac , H3K4me3, p300 (GSE24165)(Creyghton et al. 2010), MOF (GSE37268)(Li et al. 2012) and RNApolII 8WG16 (GSE34520)(Brookes et al. 2012). All datasets were aligned to the mm9 genome assembly (July 2007, NCBI37) in Bowtie (Langmead et al., 2009), allowing a maximum of two mismatches per read. Only sequences which mapped to one location were kept, and output was in SAM format.

Format Conversion:

SRA-lite formal files were converted into FASTQ format using the SRA-toolkit fastq-dump tool. FASTQ files were used as input for Bowtie alignment for which the output was a SAM file. SAM files were conversed to BAM files using Samtools view function (Li et al., 2009), and BAM files were converted to BED files (tag location in “chromosome, start, stop” format) using BEDtools (Quinlan and Hall, 2010).

Wiggle files for visualisation of tag density across the genome were generated in WCB Edinburgh Galaxy server in 200bp windows with 20bp step, and were normalised to reads per million using a custom Perl script.

Quality Control:

Solexa quality scores were visualised (along with other quality control information such as overrepresented reads and base composition of reads) using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Transcription Start Site/End Site Tag Density Graphs:

TSS/TES tag density graphs were generated using custom perl scripts written by Duncan Sproul. Briefly, TSS/TES locations were downloaded from Ensembl61 Biomart (NCBI 37, http://feb2011.archive.ensembl.org/index.html), and processed to generate a list of unique, autosomal TSS/TESs. For each ChIP-seq tag the distance from this midpoint to its nearest TSS/TES was calculated. Tags equidistant from two TSS/TESs were discarded. A histogram was then generated using a list of genes of interest, and a region of interest around the TSS/TES. Each tag was analysed to determine whether it mapped to a TSS of interest, and whether it fell within an interval of interest. Tag density was then normalised to number of reads per million aligned reads (RPM) in the ChIP-seq dataset used.

Average Gene Profile Graphs:

Profile of histone modifications over an 'average' transcript were generated using custom perl scripts and the coverageBED function of BEDtools. Transcript locations were downloaded from Ensembl61 Biomart (NCBI37), and the location (chromosome, start, and stop) of a division of each transcript in 10% intervals was determined, along with the region 2kb upstream and 2kb downstream in 200bp intervals. The tag density of each interval was determined using coverageBED, and plotted.

Promoter Window Correlations:

Promoter windows were defined as 2kb upstream of the TSS, and the normalised (RPM) number of tags for each ChIP-seq dataset within each TSS promoter window was calculated using custom perl scripts and the coverageBED function of BEDtools. Spearman correlation of the distributions between each dataset was calculated and plotted as a heatmap in R version 2.14.0.

Visualisation of Tag Density:

Wiggle tracks were uploaded to the Integrated Genome Browser (IGB) version 6.7.1 (Nicol et al., 2009), found at http://bioviz.org/igb/releases.html, along with RefSeq gene set for mm9 (July 2007) genome assembly. Higher peaks indicate an increased depth of sequencing.

Normalisation:

For the bulk of analysis, ChIP-seq datasets were normalised according to total tag number by division by number of million aligned reads in the total dataset.

For exploratory quantile normalisation, which was not selected as the most appropriate method of normalisation, custom R scripts written by Rob Illingworth, and the Limma package in Bioconductor were used. Minimum read depth for ChIP and Input wiggle tracks were set to 0.1, and quantile normalisation (using Limma NormaliseBetweenArrays function with method=Quantile argument) was performed on the two datasets. Wiggle tracks were then generated for the normalised ChIP and input samples, or for Log2 ChIP/Input and visualised in the Integrated Genome Browser.

Peak Calling:

Three methods were tested for peak calling; Nucleosome Positioning from Sequencing (NPS), Model-based Analysis for ChIP-seq (MACS), and Spatial clustering approach for the Identification of ChIP-Enriched Regions (SICER).

NPS version 1.3.2 (Zhang et al., 2008b) identifies positioned nucleosomes from histone modification ChIP-seq data. Both undifferentiated and differentiated datasets were used as input for one run then tag density was calculated within each positioned nucleosome for each dataset separately using custom perl scripts. Nucleosomes which had a significantly different tag number from either UD or D3 datasets (Fisher's Test) were taken for further examination.

MACS version 1.3.7.1 (Zhang et al., 2008a) uses the fact that ChIP-ed fragments are sequenced from one end or the other (or in the case of paired end sequencing, both ends) of a larger fragment, and that therefore tags form a bimodal distribution around the "binding site" (or enriched region, for histone modifications). To pinpoint the exact peak location, the algorithm takes the distance between the two bimodal peaks (d), and shifts tags a distance of d/2, to centre them onto the actual peak location. It then slides a window of d2 distance

across the genome to locations of potential peaks, and overlapping enriched regions are merged. The local background around these peaks is calculated from sequencing input chromain, and false positives are removed.

SICER version 1.1 (Zang et al., 2009) was developed in response to a lack of algorithms designed to cope with the frequently diffuse profile of histone modifications as compared to transcription factor binding sites for which most peak calling algorithms had been developed. The algorithm partitions the genome into non-overlapping windows (200bp), and the number of reads in each window is calculated, and each window is given a score based on the probability of finding *x* number of reads in that window, given the input background. Eligible windows are then clustered according to their local environment (i.e. windows located close to other eligible windows are clustered together to form 'islands'), depending on a 'gap size' (600bp for H4K16ac, but user-variable depending on the observed profile of the modification; for example, H3K4me3 forms sharp peaks rather than domains, so a smaller gap size would result in more relevant data for this modification). Eligible windows which are located at a distance less than the gap size from another eligible window are clustered together.

Saturation Curves:

A random selection of a proportion of reads (from 10% to 100% in 10% intervals) was selected from aligned sequencing data in Unix, then used as input for SICER peak calling. The number of peaks for each fraction of data was then plotted.

CpG Promoter Definition:

A list of CpG promoters was determined by using a defined list of CpG islands (Illingworth et al., 2010). CpG promoter windows were defined over a 2kb region of TSSes which overlapped a known CpG island. Non-CpG promoters were defined from TSSes which did not overlap one of the known CpG islands. Tag density within the promoter regions was determined using the coverageBED function of BEDtools and plotted as a boxplot in R.

Genomic Distribution of Histone Modification/HAT peaks:

The genomic distribution of H4K16ac, H3K27ac, H3K4me3, MOF and p300 were calculated using Cis-Regulatory Element Annotation System (CEAS) (http://liulab.dfci.harvard.edu/CEAS/), using default parameters, with the exception that the promoter region was set to extend to 10,000kb upstream of the TSS.

Enhancer Definition:

For enhancer heatmaps, candidate active enhancers were defined as regions with overlapping peaks (at least 50bp) of H3K4me1 and H3K27ac (Creyghton et al. 2010), and inactive candidate enhancers were defined as H3K4me1 peaks which did not overlap any peak of H3K27ac. The number of reads in twenty 500bp windows +/-5kb around the enhancer midpoint was determined using coverageBED, normalized to reads per million, and clustered based on sum of windows for H4K16ac using CreateTreeView Files in WCB Edinburgh Galaxy server, then visualised using Java TreeView (http://jtreeview.sourceforge.net/).

Pluripotency transcription factor bound candidate enhancers were taken from a publicly available dataset (Whyte et al. 2012), and overlapped with H3K4me1 peaks defined in that study. Unbound candidate enhancers were defined as H3K4me1 peaks which did not overlap with TF bound enhancers. The number of reads in twenty 500bp windows +/-5kb around the enhancer midpoint was determined using coverageBED, and normalized to average RPM per enhancer.

Subtraction and overlap of peaks was performed in WCB Edinburgh Galaxy server with custom tools.

Bioconductor Tools:

Bioconductor version 2.9 was used in R version 2.14.1 for several applications. ChIPpeakAnno (Zhu et al., 2010) was used to generate the Venn diagram of overlapping histone modification peaks. The closest TSS to H4K16ac/H3K27ac+/- candidate enhancers was defined using the AnnotatePeakInBatch function of ChIPpeakAnno to annotate peak locations with their nearest TSS, and selecting those less than 5000bp from the enhancer using custom R commands.

2.2.10 Fluorescence In-Situ Hybridization (2D FISH)

Fosmid Labelling

Fosmids were labelled with biotinylated dTTP or digoxygenin dUTP by nick translation. 500ng to 1µg of fosmid DNA prepared as described in section 2.2.4 was incubated with 2µl Nick Translation Salts, 0.5U DNAseI (Invitrogen, cat#18047), 5U DNA polymerase I (Invitrogen, cat#18010), 2.5µ each of dATP, dCTP, dGTP (0.5mM), and either 2.5µl bio-16-dUTP (Roche) or 1.5µl digoxigenin-11-dUTP and 1µl dTTP for biotin or dig labelling respectively. The incubation took place for 90 minutes at 16°C, and the reaction was stopped by addition of 2µl 20% SDS and 3µl 0.5M EDTA. Reaction was made to 90µl volume by addition of TE, then processed through QuickSpin sephadex columns (Roche, cat#11273965001) according to manufacturer instructions.

Labelling was detected using Streptavidin alkaline phosphatase or anti-DIG alkaline phosphatase on nitrocellulose filters which were prepared by soaking in 20x SSC and dried. A standard curve of biotin and DIG labelled λ DNA was spotted onto the filtres, along with 4 dilutions of fosmid DNA from 1/500 to 1/10,000. DNA was crosslinked to the filter by UV irradiation (1500mJ) in a UVC500 Crosslinker (Hoefer), then the filter was washed for 5 minutes in 0.1M Tris pH7.5 and 0.15M NaCl, followed by addition of 3% BSA (w/v) and incubation at 60°C for 30 minutes. The filter was then transferred to 0.1M Tris pH7.5 and 0.15M NaCl and 1% streptavidin alkaline phosphatase and 1% anti-DIG alkaline phosphatase for 15 minutes at room temperature, then washed in 0.1M Tris pH7.5 and 0.15M NaCl twice for 15 minutes, and once in 0.1M Tris pH 9.5. The biotin/DIG were detected using a Vector BCIP/NBT kit according to manufacturer instructions, and fosmid probe concentrations were calculated by comparison to known standards.

2D FISH

Approximately $1x10^7$ cells were washed with PBS and trypsinized, then swollen in hypotonic buffer for 15 minutes, and centrifuged at 1200g for 4 minutes, after which the hypotonic buffer was removed, and the cells resuspended in fixation solution (3:1 Methanol:Acetic Acid). Cells were then centrifuged at 1200g for 4 minutes, and fresh fixation solution was added. The cells were then stored at -20°C.

Glass slides were stored in dilute HCl in ethanol, and were dried prior to use. Cells were resuspended in 1-3mL fresh fixation solution and dropped from a height of approximately 20cm and dispersed by blowing on the slide. Quality of slides was monitored by phase

contrast microscophy, and slides were aged by incubation at room temperature for 48 hours, or baking at 60°C for 1 hour.

Slides were then treated with RNAseA at 100mg/mL for 1 hour at 37°C then washed in 2xSSC and dehydrated through 70, 90 and 100% ethanol for 2 minutes each, and air dried. Slides were then incubated at 70°C for 5 minutes, and denatured in 70% formamide (in 2xSSC, pH7.5) at 70°C for 1 minute exactly. Slides were then quickly transferred to ice cold 70% ethanol and dehydrated through ethanol series as previously, then air dried.

200ng of labelled probe was used per slide, with 8ug CotI DNA (Invitrogen, cat#18440-016) and 10ug sonicated salmon sperm DNA (Sigma, cat#31149) and denatured in hybridization mix at 70°C for 5 minutes, the preannealed at 37°C for 15 minutes. The probe was then hybridized to the denatured slides in a humid chamber at 37°C overnight.

The slides were washed for 4x3 minutes in 2xSSC at 45°C, then 0.1xSSC at 60°C, and were transferred to 4xSSC. Slides were incubated for 5 minutes in 4xSSC, 1% Marvel (w/v). Digoxigenin labelled probes were detected using Fluorescin anti-DIG FAB fragments (Roche, cat#11207741910) followed by FITC-conjugated anti-sheep (Vector, cat#FI-6000). Biotin labelled probes were detected in a double reaction, first using Texas Red conjugated avidin (Vector, cat#A-2016) then by biotinylated antiavidin (Vector cat#BA-0300), and then again by Texas Red conjugated avidin. Slides were washed between each antibody incubation for 3 x 2 minutes in 4xSSC containing 0.1% Tween20 (v/v) at 37°C. After final wash, slides were counterstained in 0.5 µg/mL DAPI and mounted.

Image Analysis

Scripts for 2D Image analysis were written by Paul Perry (MRC HGU), who also provided assistance for all microscopy work.

Slides were imaged on a Hamamatsu Orca AC CCD camera (Hamamatsu Photonics (UK) Ltd) and a Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives, a 100W Hg source (Carl Zeiss), and Chroma #83000 triple band pass filter set (Chroma Technology Corp) with excitation filters installed in a motorised filter wheel (Prior Scientific Instruments). Image capture and image analysis were performed with in-house scripts written by Paul Perry for IPLab Spectrum (Scanalytics Corp).

Only nuclei which were evenly shaped and not in contact with another nucleus were selected for analysis. Probe spots which were bright and easily distinguished from background were

used, and only nuclei with two pairs of probe spots were used (selecting for diploid cells). Nuclei were imaged using a 100x objective in bin2. The squared interprobe distances (d^2) were measured and normalised to nuclear radius squared (r^2) or left unnormalised. Statistical significance of mean interprobe distance were examined with non-parametric Mann-Whitney U test in R version 2.14.0 with the null hypothesis that the two datasets contain the same distribution. Each dataset consisted of at least 50 nuclei (100 loci) for each cell type and probe combination. Control probe sets and experimental probe sets were hybridised to cells in parallel, within the same experiment.

2.2.11 Protein handling and preparation

Preparation of Whole Cell Extract

Cells were trypsinised and washed in ice cold PBS and pelleted at 400g. Cell pellet volume (in µl) was estimated by eye. The cell pellet was resuspended in 5x its volume of NP-40 lysis buffer and incubated on ice for 30 minutes with intermittent vortexing. Lysed cells were centrifuged at 15,000g for 10 minutes at 4°C, and supernatant collected and stored in the short term at -20°C.

SDS PAGE Resolution of Proteins

Cell extracts were resolved by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Denaturing polyacrylamide gels were made with 10% (v/v) Acrylamide/Bis (29:1, Biorad cat#161-0156), 0.33M Tris-HCl pH8.8, 0.1% SDS (v/v), 0.1% APS (ammonium persulphate, v/v), and 0.05% Tetramethylethylenediamine (TEMED, v/v) in water. The gel was set with a cover of isopropanol, which was removed when the gel was set, and stacking gel made with 5% Acrylamide/Bis, 0.125M Tris pH6.8, 0.1% APS and 0.1% TEMED.

Gels were run in a Novex minicell X-Cell Sure Lock (Invitrogen) electrophoresis tank in SDS-PAGE running buffer at 130V for approximately 90 minutes. A prestained protein size ladder of standard molecular weights was loaded alongside all gels (PageRuler Plus, 10-250kDa, Thermoscientific, cat#26619).

Coomassie Staining for loading control and visualisation

The stacking gel was removed from SDS-PAGE gel, and the SDS-PAGE gel was washed in water, then fixed in Coomassie destain solution for 10 minutes at room temperature. The gel was then transferred to 30mL 10% acetic acid, and 10% coomassie stain was added (v/v).

Gel was incubated for 1 hour at room temperature, then transferred to destain solution for >2 hours.

Colloidal coomassie staining was carried out with a Colloidal Blue Staining Kit purchased from Life Technologies (cat #LC6025) and used as specified by manufacturer instructions.

Western Blotting

Stacking gel was removed from SDS-PAGE gel, and samples were transferred to polyvinylidene difluoride (PVDF) membrane. PVDF membrane was soaked in methanol and then transferred to western transfer buffer for 2 minutes before being used in transfer on a Genie Blotter (Idea Scientific). Gel and membrane were sandwiched between two sheets of 3mm filter paper soaked in western transfer buffer, and transfer proceeded for 90 minutes.

After transfer, the membrane was blocked in 3% BSA (v/v) in TBS with 0.2% Tween 20 for 1 hour at room temperature or overnight at 4°C. α-FLAG primary antibody (Sigma cat#F7425) was diluted to 1:10,000 in 3% BSA in TBS-Tween, and membrane was incubated in primary antibody overnight at 4°C with agitation. Membrane was washed three times for ten minutes in TBS-tween at room temperature. Secondary antibody (α-Rabbit IgG conjugated to HRP, Sigma) was diluted to 1:10,000 in 3% BSA/TBS-Tween, and membrane was incubated with antibody with agitation for 1 hour at room temperature and washed as previously. Signals were detected with ChemiGlow West Chemiluminescence Substrate Kit (Protein Simple) according to manufacturer instructions, using Amersham Hyperfilm (GE Healthcare).

GST-tagged protein isolation

GST-tagged chromodomain of MOF was was amplified from MOF cDNA (clone MGC:30369 (IMAGE:5136304), clone sequence BC036284.1, in pCMV-SPORT6) by PCR with inclusion of BamHI and NotI restriction sites, and cloned into a overexpressed in bL21-RIPA competent bacteria in a pGEX6P1 plasmid as described in section 2.2.4. The bacteria were pelleted, and resuspended in 20mL 1x ice cold PBS with 1mg/mL lysozyme, 0.2mM PMSF, and Protease Inhibitor cocktail. The resuspended bacteria were sonicated on a Soniprep 150 sonicator for 20 minutes on ice with 30 second on off cycles. 1% Triton X-100 (v/v) was added, and the solution was mixed gently for 30 minutes at room temperature, then centrifuged at 12,000g for 10 minutes at 4ºC. 200µl of Glutathione Sepharose 4B beads (GE Healthcare, cat#17-0756-01) were washed in 1mL PBS, then the supernatant was incubated with the beads for 30 minutes at room temperature. Protein was eluted from beads

using four incubations with 200µl Glutathione Elution Buffer for 10 minutes at room temperature. A sample of elutions was checked by coomassie stain and western blot for specific expression of GST-tagged protein, and positive elutions were dialysed with membrane with a molecular weight cut off of 6-8000 (Spectrum Labs), against PBS and 20% glycerol (v/v) overnight at 4ºC. Protein concentration was checked using Bradford reagent (BioRad) according to manufacturers' directions on a spectrophotometer using a standard curve of BSA (New England Biolabs) to determine the correct concentration.

MODified Histone Peptide Array

MOFified Histone Peptide Array was purchased from Activ Motif (cat#13005) and used according to manufacturer instructions. Briefly, the slide was blocked in 5% MARVEL powdered milk at 4ºC overnight, then washed in TBS-Tween (10mM Tris-HCl pH8.3, 150mM NaCl, and 0.05% Tween-20 v/v) for 5 minutes. The slide was then incubated with 10nM GST-tagged MOF chromodomain protein, or GST tag alone at room temperature for 1 hour, then washed 3x5 minutes in TBS-Tween. Bound proteins were detected with α-GST primary antibody (Abcam, cat#ab6613, used at a dilution of 1:5000 for 1 hour at room temperature), washed as previously, then α-goat HRP (Sigma, used at a dilution of 1:12,000 for 1 hour at room temperature), washed again as previously, and signal was detected with ChemiGlow West Chemiluminescence Substrate Kit (Protein Simple) on an ImageQuant LAS4000 (GE Healthcare). Signal intensities and further analysis was performed with Array Analyser software (Activ Motif) as directed.

H4K16ac antibody specificity was calculated using MOFified Histone Peptide Array as above, with the exception that no protein was incubated with the array, and H4K16ac was used as primary antibody and α-Rabbit IgG HRP was used as secondary antibody.

SILAC IP

SILAC IP experiments were carried out in collaboration with Pradeepa Madapura Marulasiddappa (MRC HGU).

10 x 14cm dishes of OS25 or MEF cells were grown to approximately 80% confluence in SILAC DMEM (R0K0, R6K4, and R10K8, Dundee Cell Products) media supplemented with 10% FCS and penicillin (10,000 units/mL) and streptomycin (650µg/mL) for MEFs, or 10%

FCS, 0.1mM β-mercaptoethanol, penicillin (10,000 units/mL) and streptomycin (650µg/mL), 0.3mg/mL L-Glutamine, 1mM sodium pyruvate, non-essential amino acids (Sigma), 1000 units/mL human recombinant LIF and 0.1mg/mL hygromycin (Roche) for OS25s. Cells were trypsinised and pelleted, then resuspended in 5mL ice cold swelling buffer for 5 minutes, then dounced with a pre-chilled Dounce homogeniser (for 20 strokes with a tight pestle). The homogenised cells were centrifuged at 228g for 5 minutes at 4ºC, and the supernatant discarded. The nuclear pellet was resuspended in 5mL ice cold RIPA buffer and incubated for 30 minutes on ice, then sonicated in a Bioruptor Next Gen (Diagenode) for 10 minutes at full power with 30s on/off cycle at 4ºC. Extracts were then centrifuged at 16,000g for 10 minutes at 4ºC. Protein concentrations between the heavy, medium and light cell samples were tested using the Bradford assay.

Protein A dynabeads were prebound to 5µg of antibody (Rabbit IgG, αMOF – Bethyl labs cat#A300-99219, α-p75) for 1 hour at 4ºC, then equivalent amounts of nuclear extract from light, medium and heavy media was incubated with one of the antibodies for 1 hour at 4ºC. The beads were washed once separately in RIPA buffer, then combined and washed four more times (5 minutes at 4ºC). Bound proteins were eluted with 4x Invitrogen Loading Buffer with freshly added DTT.

SILAC mass spectrometry was carried out from this point by Dundee Proteomics. Briefly, the boiled proteins were separated by SDS-PAGE. The entire protein gel lane was cut into ten gel slices per fraction, and trypsin-digested peptides were separated on an Ultimate U3000 trap-enriched nanoflow LC-system (Dionex). An LTQ Orbitrap XL Mass Spectrometer with a nano-ES ion source (Proxeon Biosystems) was used for identification, and data acquired using Xcalibur software. Quantification was carried out using MaxQuant software (version 1.0.7.4). Conditions for mass spectrometry are as described in (Emmott et al., 2010).

2.2.12 Fluorescence Activated Cell Sorting (FACS)

Flow cytometric analysis of fluorescent cells was performed on a BD FACSAriaII SORP (Becton Dickinson) by Elisabeth Freyer (FACS Technician, MRC HGU). The instrument was controlled, and data was analysed in BD FACSDiva (Becton Dickson, Version 6.1.2).

2.2.13 Websites used for Computational Biology

www.ncbi.nlm.nih.gov/geo/*:* Gene Expression Omnibus

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/: FASTQC

http://www.perl.org/: Perl programming language for multiple custom scripts.

http://feb2011.archive.ensembl.org/index.html: Ensembl archive version used.

http://bioviz.org/igb/releases.html: Integrated Genome Browser (version 6.7.1 used)

http://liulab.dfci.harvard.edu/CEAS/: CEAS (version 1.0.2 used)

http://jtreeview.sourceforge.net/: Visualisation of enhancer heatmaps.

http://bifx-core.bio.ed.ac.uk:8080/: Edinburgh University Wellcome Trust Centre for Cell Biology Galaxy Server.

http://www.r-project.org/: R software (Version 2.14.0 used)

http://www.ebi.ac.uk/Tools/msa/clustalo/: ClustalOmega, for sequence alignment of cross-species MOF-chromodomains.

http://smart.embl-heidelberg.de/: SMART, for detection of chromodomains.

Chapter 3: H4K16 Acetylation is found on active promoters and enhancers in Embryonic Stem Cells

3.1 Introduction

Acetylation of H4K16 (H4K16ac) has been noted *in vitro*, for its role in upregulating transcription (Akhtar and Becker, 2000), for its ability to decompact nucleosome arrays, and to prevent their compaction (Shogren-Knaak et al., 2006; Robinson et al., 2008). *In vivo* H4K16ac has rarely been studied in mammalian cells, but has been examined in more detail in *D. melanogaster* where it forms an essential part of the X chromosome dosage compensation system, and in *S. cerevisae*, where targeting H4K16ac to a responsive promoter can trigger transcription(Akhtar and Becker, 2000).

Because of its direct role in chromatin structure and nucleosome-nucleosome interaction, I wanted to examine the role of H4K16ac *in vivo* in mammalian cells. Given that H4K16 has a potentially important role in murine development (Thomas et al., 2008), I made the decision to use embryonic stem cells (ES cells) to study the pattern of H4K16ac genome wide.

3.2 H4K16ac antibody verification, ChIP and ChIP-Seq QC

Antibody verification is an important step in any ChIP-sequencing experiment, as non-specific binding can give a false picture of the target profile. I therefore used the MODified Histone Peptide array (Activ Motif) which contains 59 modifications on N-terminal tails in combinations to form 384 targets for antibody testing. The H4K16ac antibody (Millipore) has previously been shown to be suitable for ChIP (Egelhofer et al., 2011), but batches of antibody can vary in efficiency. The results of the histone peptide array show that the antibody used is extremely specific to H4K16ac (Fig 3.1). The antibody binding to non-specific modifications gives very low signal (Fig 3.1A), and only histone tails which contain H4K16ac show visible signal (Fig 3.1B). H4K16ac is calculated to have a specificity factor approximately twenty times that of the highest background modification (Fig 3.1C).

The H4K16ac antibody used is therefore specific and suitable for ChIP.

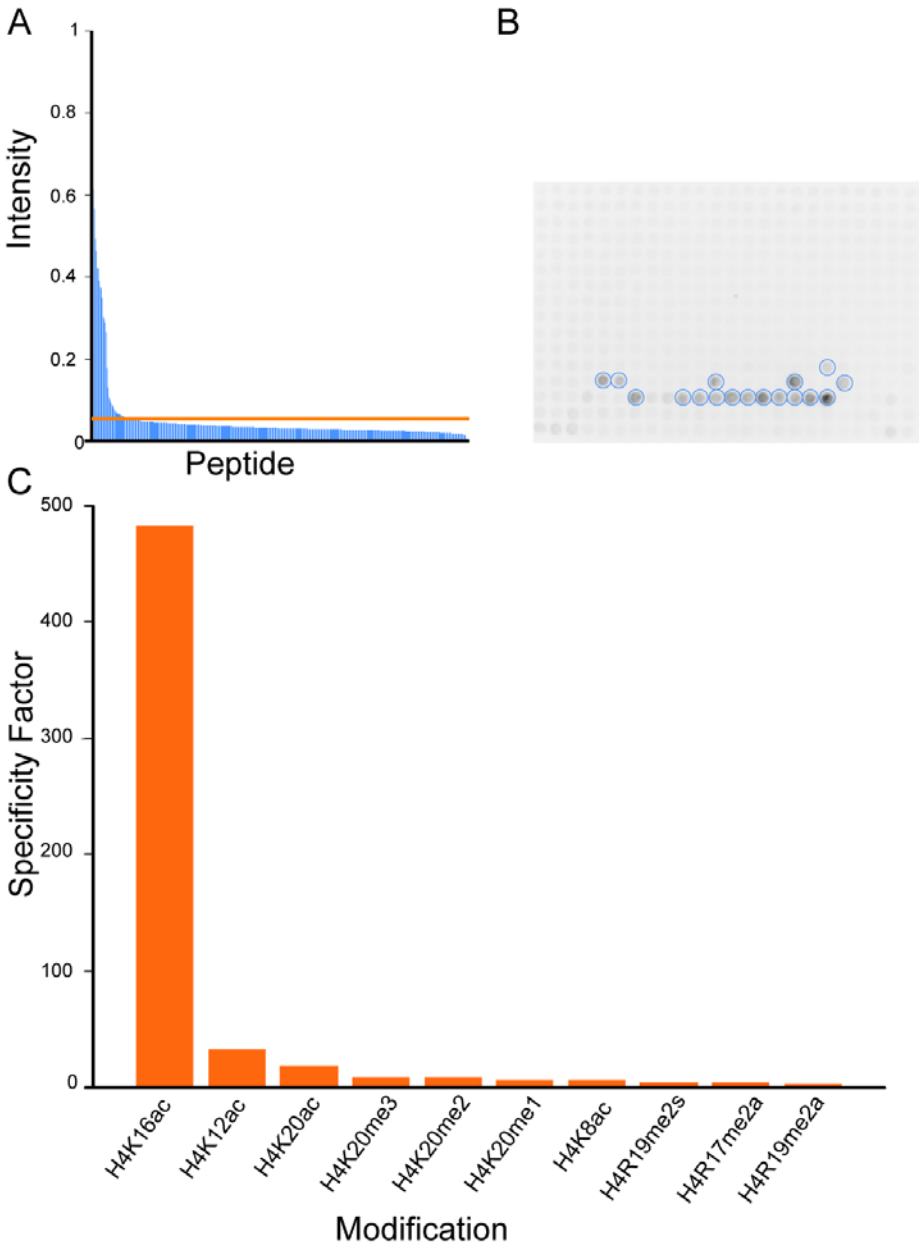


Figure 3.1 *H4K16ac antibody is specific*

A: Signal intensities for all spots on MODified Histone Peptide array. Orange line represents threshold for signal intensity designated as background. B: Example of H4K16ac antibody signal on MOFified Histone Peptide array. Spots containing H4K16ac modification are circled in blue. C: Specificity factors calculated by Array Analyse software (activ motif) given background designated in (A) for top 10 hits.

Native ChIP for H4K16 acetylation was carried out as shown in Figure 3.2A. Briefly, isolated chromatin from OS25 ES cells was digested with MNase to mono di and trinucleosomes (Fig 3.2B) that were then incubated with antibody-coupled beads. This was followed by washing and elution of bound DNA, which was purified and used in downstream analysis.

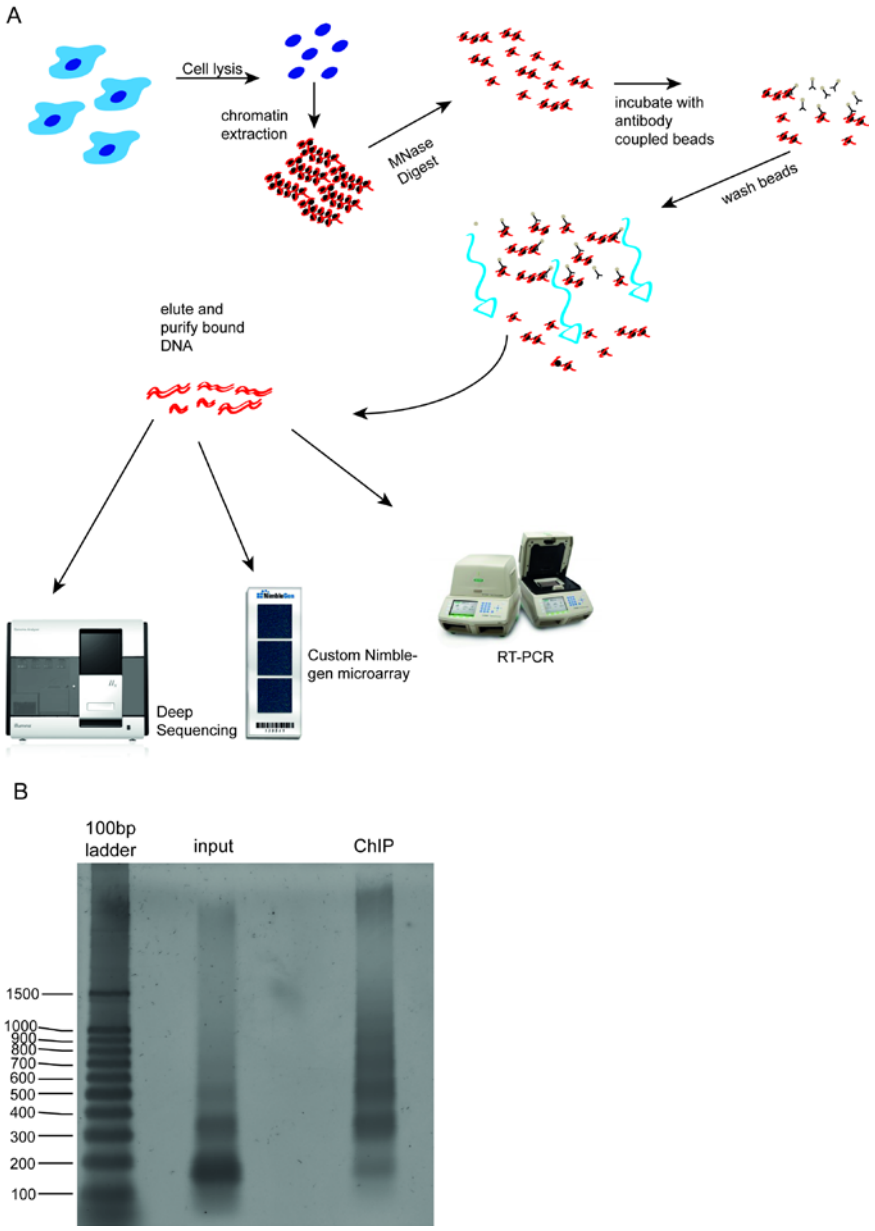


Figure 3.2 *ChIP-Seq workflow*

A: Schematic detailing Native ChIP workflow. Cells are lysed, chromatin extracted and digested with MNase to mono di and tri nucleosomes (which fall within the correct DNA length suitable for ChIP-seq library preparation). Buffers contain HDAC inhibitors throughout. The digested chromatin is then incubated with antibody coupled sepharose A beads, washed, and the bound DNA is eluted, purified, then used in downstream applications. First, qRT-PCR and microarrays for ChIP quality control, and then Solexa sequencing. B: MNase digested input and H4K16ac ChIP chromatin, digested with proteinase K, size separated on a 1.3% agarose gel, and stained with SYBR Gold. Size marker indicated on left (bp).

The initial step was to verify the ChIP specificity and efficiency by qRT-PCR. H4K16ac is known to be enriched on the promoters of active genes (Wang et al., 2008) and is not found on inactive genes. I therefore opted to look at the promoters of housekeeping and pluripotency genes (*β-actin* and *Oct4/Pou5f1*) as positive controls. Two lineage specific genes not expressed in ES cells (*Olig2* and *Hoxb1*) were used as negative controls.

As expected for a specific ChIP, H4K16ac is found on the promoters of the positive controls, but not those of the negative controls (Fig 3.3A). To further verify the ChIP across whole genes rather than just promoters, and to provide controls and verification for the subsequent ChIP-sequencing data, I hybridized the ChIP DNA to custom NimbleGen microarrays. These in house designs (Pradeepa et al., 2012) contain various pluripotency and control genes (array designs are detailed in Appendix 1), in addition to the four mouse *Hox* loci. The results show that, as expected, H4K16ac is found most densely on the promoters of the positive control genes, and not at all on the negative control gene (Fig 3.3B). Though H4K16ac is slightly more highly enriched on the promoter of *Pou5f1*, it is also found through the body of this gene. This is in contrast to the profile over *β-actin* and *Nanog*, where it is more obviously restricted to the promoter region. The profile of H4K16ac on these genes is inversely correlated with the silent-gene associated histone mark H3K27me3 (data used from (Illingworth et al., 2012)).

Across the *Hox* loci (*Hoxb* and *Hoxc* are shown as examples - Fig 3.4), H4K16ac is largely excluded from the *Hox* genes (again, in a roughly inverse relationship with H3K27me3), though interestingly *Gm53* (upstream of *Hoxb9*, Fig 3.4, top) is acetylated at H4K16, and trimethylated at H3K27. The overlap is unlikely to be due to poor resolution of the microarray, which is fully tiled. Each probe is approximately 50bp, with a median 45bp step.

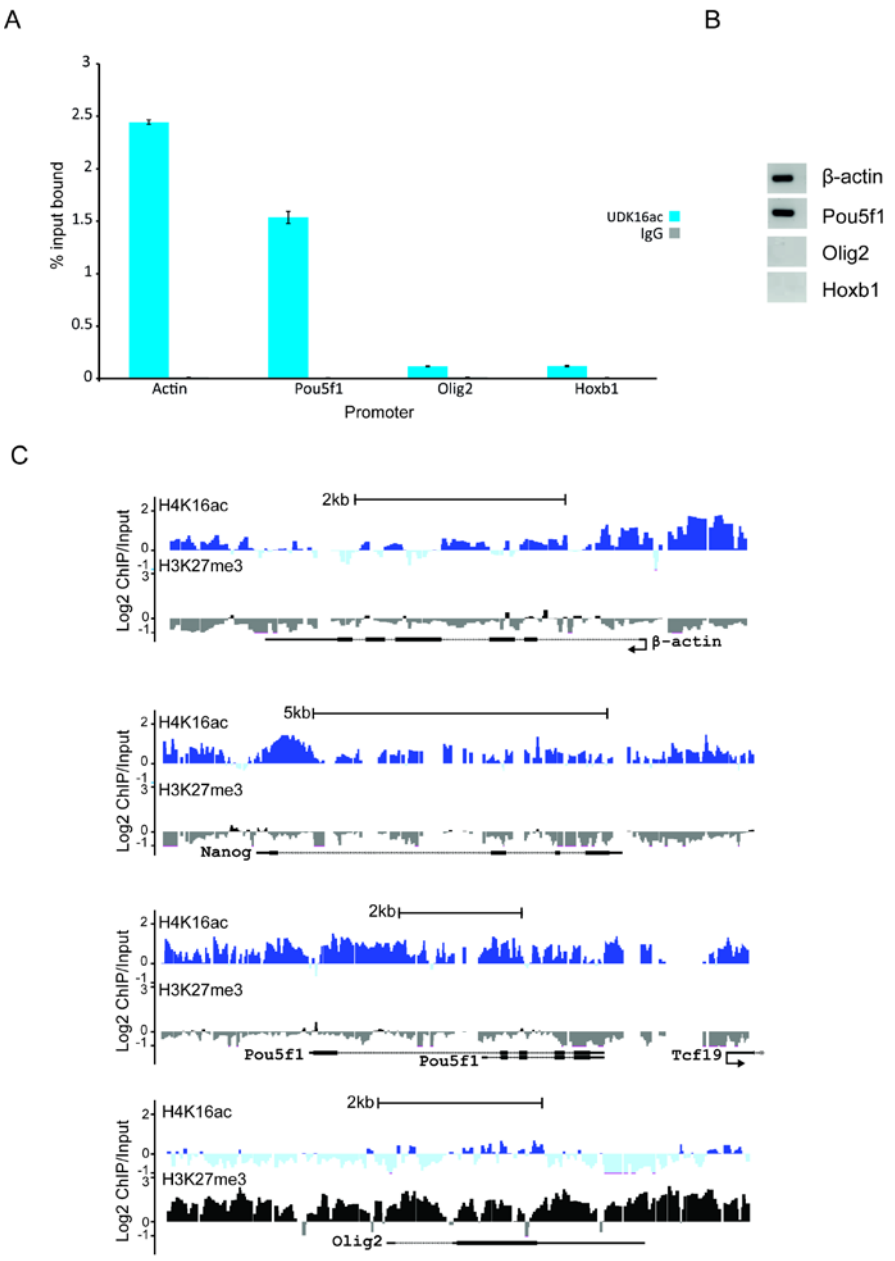


Figure 3.3 *H4K16ac ChIP quality control*

A: ChIP for H4K16ac at active or inactive promoters in undifferentiated (UD) ES cells assayed by qRT-PCR. Enrichment is shown as mean % input bound +/- SEM over three biological and three technical replicates. B: RT-PCR for expression of active/inactive genes in OS25 ES cells. C: Log_2H4K16ac/Input (top, blue) or Log_2H3K27me3/Input (bottom, black) in UD ES cells over housekeeping (*β-actin*), pluripotency (*Nanog, Pou5f1*), lineage specific (*Olig2*) genes on custom Nimblegen arrays. RefSeq gene annotations are from the July 2007 (mm9) Build 37 assembly of the mouse genome (http://genome.ucsc.edu) (Kent et al., 2002).

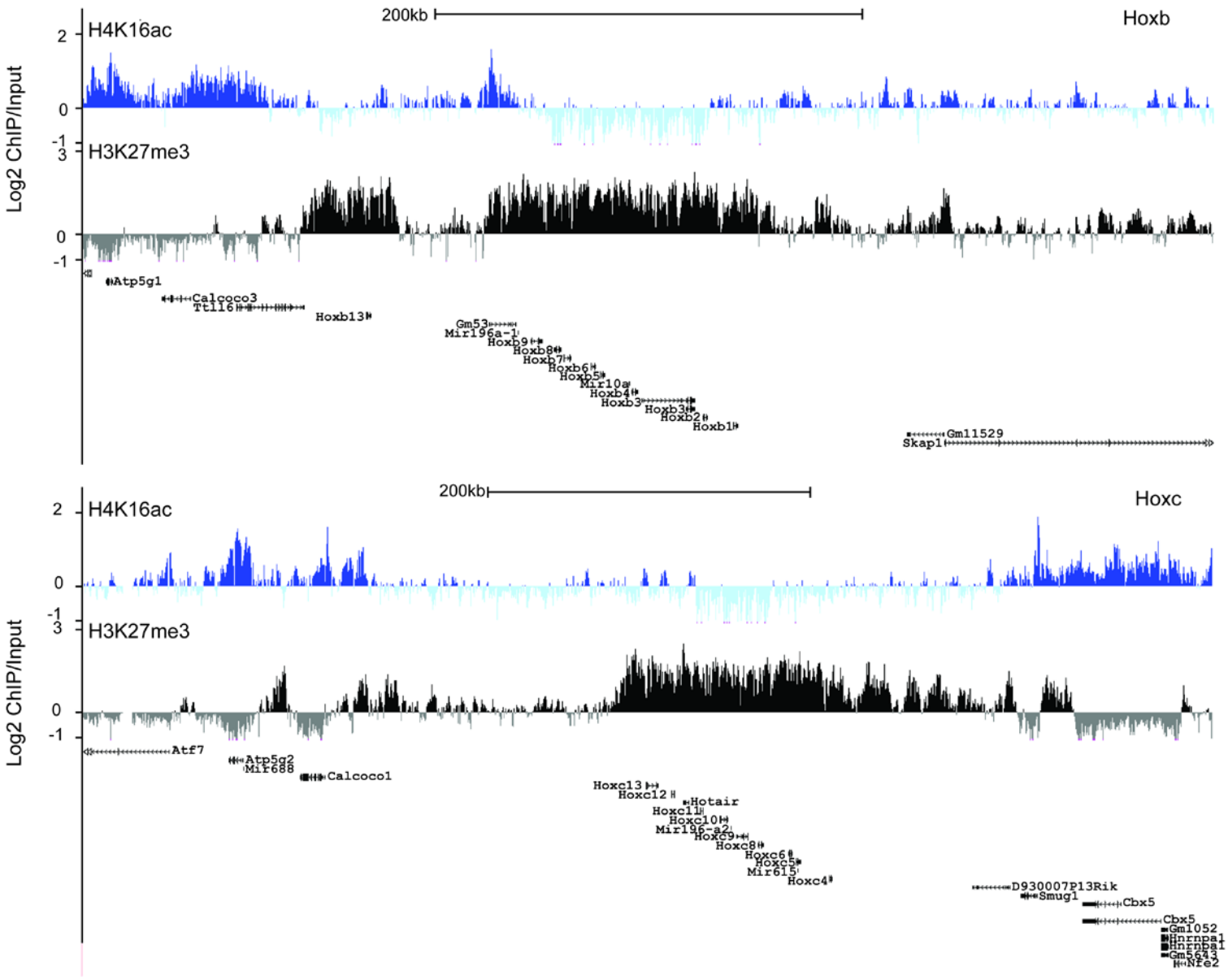

200kb
Hoxb
H4K16ac
H3K27me3
Log2 ChIP/Input
2
0
-1
3
0
-1
Atp5g1
Calcoco3
Ttll6
Hoxb13
Gm53
Mir196a-1
Hoxb9
Hoxb8
Hoxb7
Hoxb6
Hoxb5
Mir10a
Hoxb4
Hoxb3
Hoxb3
Hoxb2
Hoxb1
Gm11529
Skap1
200kb
Hoxc
H4K16ac
H3K27me3
Log2 ChIP/Input
2
0
-1
3
0
-1
Atf7
Atp5g2
Mir688
Calcoco1
Hoxc13
Hoxc12
Hotair
Hoxc11
Hoxc10
Mir196a-2
Hoxc9
Hoxc8
Hoxc6
Hoxc5
Mir615
Hoxc4
D930007P13Rik
Smug1
Cbx5
Cbx5
Gm1052
Hnrnpa1
Hnrnpa1
Gm5943
Nfe2

Figure 3.4: *H4K16ac ChIP quality control over Hox loci* (previous page)

Log_2H4K16ac/Input (top, blue) or Log_2H3K27me3/Input (bottom, black) in UD ES cells over *Hoxb* (top) and *Hoxc* (bottom) loci on custom Nimblegen arrays. RefSeq gene annotations are from the July 2007 (mm9) Build 37 assembly of the mouse genome (http://genome.ucsc.edu).

The H4K16ac ChIP was deemed to be specific and efficient, and therefore I proceeded with deep sequencing of H4K16ac bound DNA, and MNase digested input to be used as a control. Generating a sequenced profile of input DNA is important, firstly to account for biases in the chromatin digestion - not all chromatin is released after MNase digest, which could lead to designating false negative regions. Conversely, there may be some sequence bias in the MNase digestion (Dingwall et al., 1981). In addition, biases exist in the library preparation, the sequencing process itself, and the mapping process (reviewed in (Liu et al., 2010) and references therein). For example, whole cell extract DNA input has been shown to peak around transcription start sites (Vega et al., 2009).

The sequenced DNA was then subjected to quality control checks, before being mapped to the mouse genome, and used in downstream bioinformatics analysis.

I used FASTQC, a program designed to analyse the quality of deep sequencing data (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) to determine whether the data would be useable. The results indicated that some adapter sequences had been amplified and were overrepresented in the data (data not shown) but given that these do not map to the mouse genome, would not be present in the data after mapping, and represented a very small proportion of the reads, this was not a large concern. All other control checks were passed, and the Solexa Quality Scores (Fig 3.5) were acceptable. The Solexa 1.3 pipeline used in this experiment attaches a Phred score to each read to designate read quality. Phred scores range in value from 10 to 50; a Phred score of 10 means that the probability of an incorrect base call is 1 in 10, whilst a score of 50 means the probability of an incorrect base call in the read is 1 in 100,000. Scores in my ChIP-seq experiment are generally above 35, with some tail off towards the end of the read, meaning the base call accuracy is likely above 99.9%.

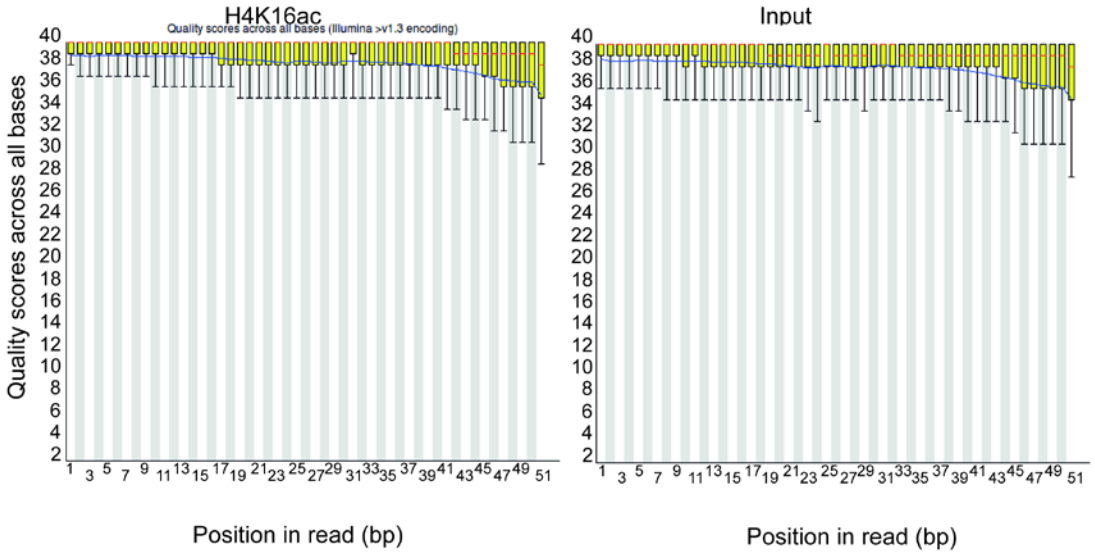


Figure 3.5 *Phred scores for deep sequencing quality control*

Quality scores generated in FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) for all reads in H4K16ac ChIP (*left)* and MNase digested input (*right*). Quality scores from Solexa 1.3 pipeline, maximum = 50

Having determined that the sequencing data was of sufficient quality, I mapped the sequences to the mm9 assembly of the mouse genome using Bowtie (Langmead et al., 2009). Table 3.1 shows the alignment statistics; the H4K16ac ChIP sample has nearly 75% of reads aligned, and the Input sample has close to 80%. Given that the sequencing quality of my reads declines after approximately 40bp, it is possible that trimming the reads before alignment could result in a slightly higher percentage of read alignment. However, the aligned read depth of my experiment was already considerably higher than most published datasets (with input and H4K16ac ChIP samples having 105 and 79.5 million reads aligned respectively), and I therefore decided that the read depth was likely sufficient.

To further investigate whether I had generated a sufficient depth of sequencing I decided to generate saturation curves by randomly sampling proportions of my data, and calling peaks using SICER (Zang et al., 2009)(peak calling will be discussed in more detail in Chapter 4). To compare my dataset saturation with that of published datasets, I used two published ChIP-seq profiles to perform the same analysis; H3K4me1, and H3K4me3 (Creyghton et al., 2010). H3K4me3 is known to form discreet peaks on active and bivalent promoters (Barski et al., 2007), whilst H3K4me1 forms broader, less defined peaks, which provides more of a challenge to peak calling algorithms (Zang et al., 2009). As expected, the H3K4me3 dataset saturates with a low proportion of reads – with only 30% of the reads, the majority of the total peaks detected with 100% of the reads have been detected. For H3K4me1 and H4K16ac the picture is more complicated, it seems likely that saturation has not been achieved, and that further sequencing would reveal further peaks, though in the case of H4K16ac the rate of peak discovery slows dramatically when approximately 50% of the data is used. At this same proportion of reads used from the H3K4me1 dataset, the number of peaks discovered is half that found with 100% of the data (Fig 3.5). Determining the appropriate depth necessary in ChIP-seq experiments is a matter of balancing the need to generate useful data against the cost of sequencing. Some datasets will never achieve saturation within a reasonable amount of sequencing, and it must be accepted that peaks determined are a proportion of what exists in the genome.

Table 3.1: Read alignment statistics.

	Total Reads	Reads Aligned to mm9	% reads aligned
UDH4K16ac	107027050	79531348	74.30957688
UDInput	134829138	105346646	78.13344175

Total number of H4K16ac and input and number of reads aligned to mm9 genome assembly using Bowtie

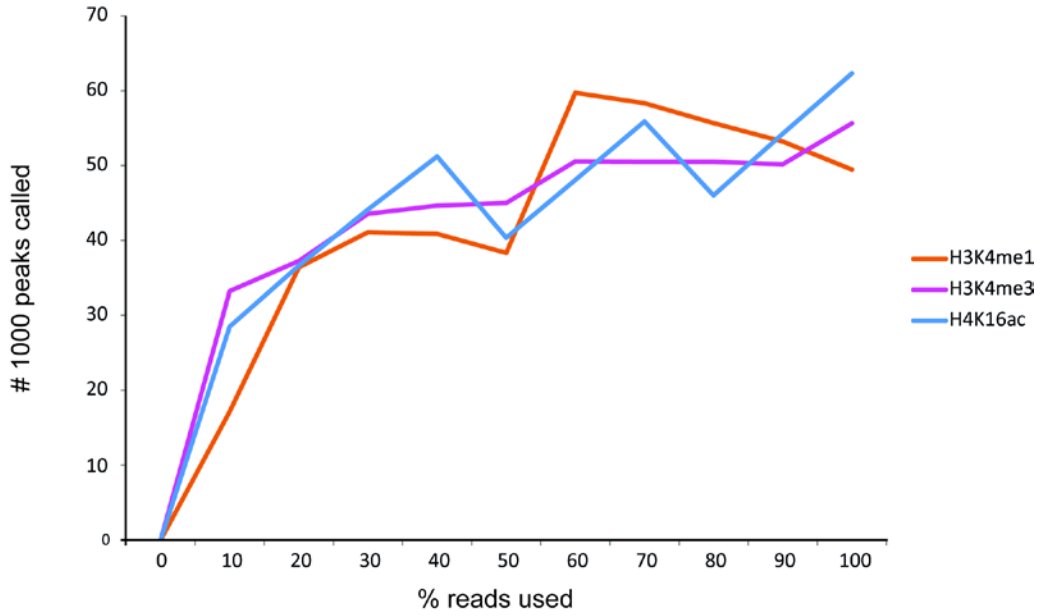


Figure 3.6 *Saturation curves*

Saturation curves for H4K16ac (this study) and H3K4me1 and H3K4me3 (Creyghton et al., 2010). Plotted as number of peaks found using SICER (Zang et al., 2009) with increasing percentage of reads randomly sampled from dataset.

3.3 Correlating gene expression and H4K16ac in undifferentiated ES cells

In order to examine the correlation between H4K16ac distribution and gene expression, I generated a profile of expressed and silent genes in undifferentiated OS25 ES cells. The expression profile was generated on the Illumina MouseWG-6 v2.0 Expression Bead Chip platform. I used this expression data to generate a sample of highly expressed or silent genes which could be examined for H4K16ac level, by taking the 500 genes with the highest average probe signal with a detection p value <0.005, and conversely, the 500 genes with the lowest signal, and a detection p value of >0.9. The highest and lowest genes in the sample are detailed in Tables 3.2 and 3.3, and the GO terms (Fig 3.7) most enriched in the active sample indicate that these genes are largely housekeeping genes and are not specific to ES cells, with the highest p-values belonging to such general GO terms as translation, and metabolic processes (H4K16ac profile for genes specific to undifferentiated and differentiated cells will be explored in more detail in Chapter 4).

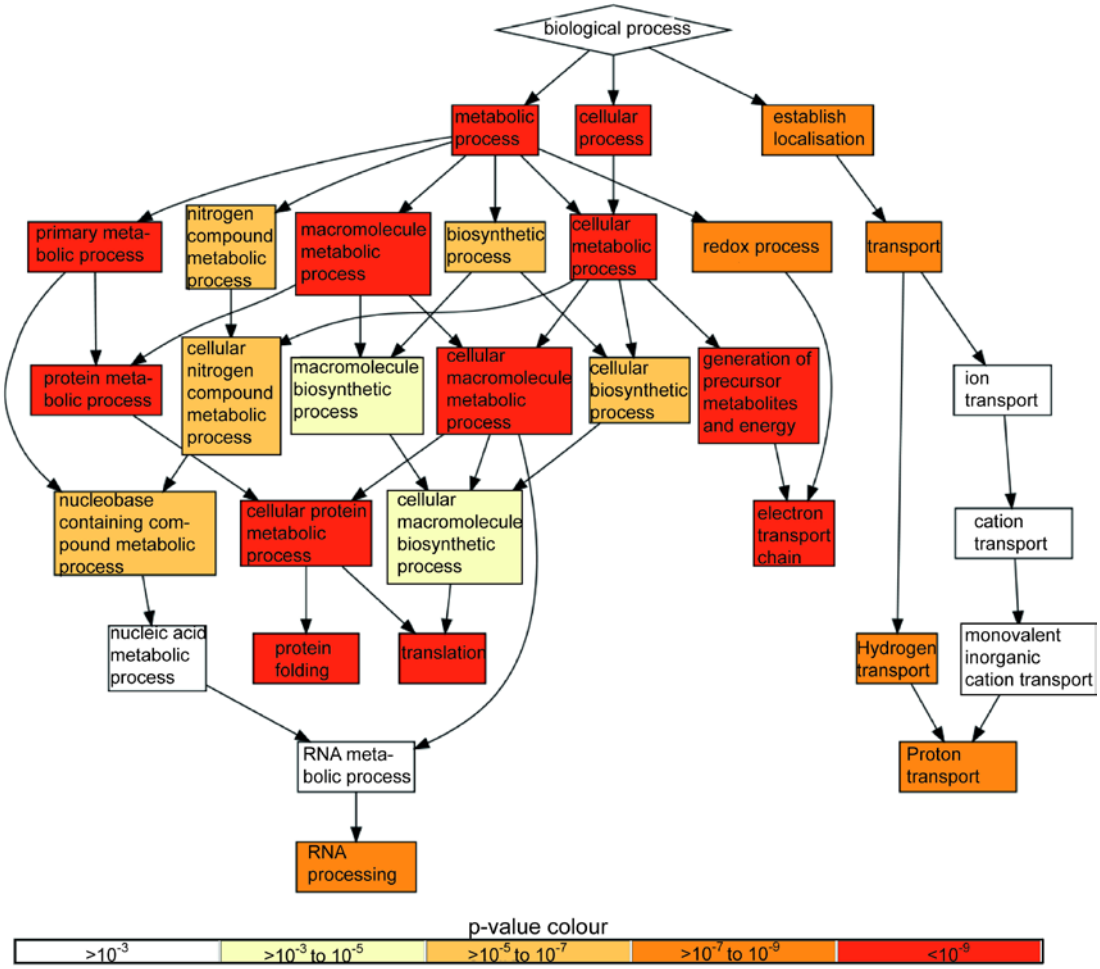


Figure 3.7 *Enriched GO terms in most active ES cell genes*

GO terms with pvalue $<10^{-8}$ for sample of active genes used in downstream analysis. Generated in GOrilla (Eden et al., 2009). Top 500 genes with detection pvalue <0.005 against array total probe background.

Table 3.2 Most active expressed genes in UD ES cells by Illumina bead array.

Gene Symbol	Average Signal				Standard Error	Detection pvalue
	UD 1	UD 2	UD 3	Mean		
EG667728	22847.61	19572.47	20257.38	20892.48667	575.8236481	0
Rplp1	22718.56	23245.13	22428.11	22797.26667	138.052516	0
LOC100047759	22159.66	21203.19	22394.61	21919.15333	210.3576311	0
Hist1h2ao	22134.52	20104.8	20739.11	20992.81	346.1235468	0
LOC623466	22031.25	21084.53	21128.81	21414.86333	178.0884835	0
LOC100043209	21801.67	19051.38	17977.16	19610.07	657.5053926	0
Tuba1b	21696.84	20637.79	21421.38	21252.00333	183.1553885	0
LOC677215	21661.02	20116.19	19070.31	20282.50667	434.4460715	0
Rps2	21429.98	22152.6	21830.65	21804.41	120.6746403	0
mtDNA_ND1	21367.66	21540.04	22622.7	21843.46667	226.7725627	0
Zyx	21284.27	21815.24	22736.39	21945.3	244.9149118	0
Atp5b	21272.15	18556.81	19419.05	19749.33667	462.4914162	0
Rps12	20598.31	20835.68	18849.96	20094.65	361.4824456	0
LOC381799	20425.81	20474.13	18314.4	19738.11333	411.0695328	0
Ubb	20337.17	18879.33	18330.82	19182.44	345.6502463	0
Rps29	20311.56	18975.64	19120.49	19469.23	244.3552178	0
Rpl10a	20155	16450.4	17210.41	17938.60333	652.2367019	0
mtDNA_COXIII	20136.06	19959.76	20189.67	20095.16333	40.09579506	0
Ppia	19920	18113.65	19285.13	19106.26	305.4542952	0

Table 3.3 Least expressed genes in UD ES cells by Illumina bead array.

Gene Symbol	Average Signal				Standard Error	Detection pvalue
	UD 1	UD 2	UD 3	Mean		
scl00093.1_86	48.53136	56.75934	52.78808	52.69292667	1.371605075	0.993233633
4930563O14Rik	48.76828	49.25138	49.29222	49.10396	0.097141233	0.999643867
LOC380676	48.95963	62.04847	48.37948	53.12919333	2.576588306	0.9761396
Prelid2	49.54593	60.67086	61.97208	57.39629	2.276557037	0.9280627
Sidt1	50.32195	63.25974	52.74836	55.44335	2.292350343	0.9643875
Olfr1419	50.46786	53.65835	55.67262	53.26627667	0.874812538	0.9910969
LOC100048814	50.71157	56.66877	49.91697	52.43243667	1.230073953	0.994302
LOC240131	50.7684	56.5957	56.61086	54.65832	1.122926022	0.985398867
4930430J02Rik	50.79748	49.36172	46.34458	48.83459333	0.757589488	0.9985755
LOC385462	50.90368	51.127	52.78337	51.60468333	0.342287184	0.996082633
Rdh13	50.91822	57.7257	57.4385	55.36080667	1.28335728	0.9804131
LOC329967	51.46046	57.56039	50.58362	53.20149	1.266763988	0.9928775
scl000460.1_17	51.47501	56.40339	54.72086	54.19975333	0.83505798	0.988247867
LOC384552	51.55732	60.30865	55.15291	55.67296	1.466260671	0.9786325
Shbg	51.58677	54.09728	52.05822	52.58075667	0.444778165	0.996082633
LOC674094	51.9087	50.46377	48.76921	50.38056	0.523799399	0.9985755
4930579F01Rik	51.97919	56.60028	56.05273	54.8774	0.841603539	0.9861111
LOC381427	52.00977	53.73677	56.69771	54.14808333	0.790293852	0.988247867
B930041G04	52.02512	58.13614	58.74847	56.30324333	1.239197386	0.97151
EG546894	52.0894	53.79928	57.41793	54.43553667	0.906882627	0.985042733

To examine the profile of H4K16ac over active and inactive genes, I used custom perl scripts (modified from scripts written by Duncan Sproul) to determine the tag density across the Transcription Start Sites (TSS), and Transcription End Site (TES) of the 500 active and 500 inactive genes (Fig 3.8A). The H4K16ac profile is considerably higher on the active genes than on the silent genes across the TSS (Wilcox test p-value for a 500bp window over the TSS, $<2.2x10^{-16}$), but the profile on the TES is unaffected by the gene expression level (p-value = 0.01927).

This result is expected, as it follows the previous H4K16ac ChIP-seq dataset conducted in $CD4^{+}T$ cells (Wang et al., 2008), where H4K16ac was determined to be increased on the promoters and the transcribed regions of active genes.

I also examined the promoters of several examples of active genes individually (a sample is shown in Figure 3.8B). The profile of H4K16ac over these genes shows a broad peak, beginning in the promoter, and at its highest approximately over the transcription start site (with a visible dip representing the +1 nucleosome - (Shivaswamy et al., 2008)) then reducing a short way into the gene body.

On the inactive genes, there is generally little H4K16ac beyond background, and no visible peak over the promoter in most genes (Fig 3.8C), however, some inactive genes do carry a small peak of H4K16ac. This peak is, on average, lower than peaks on active genes.

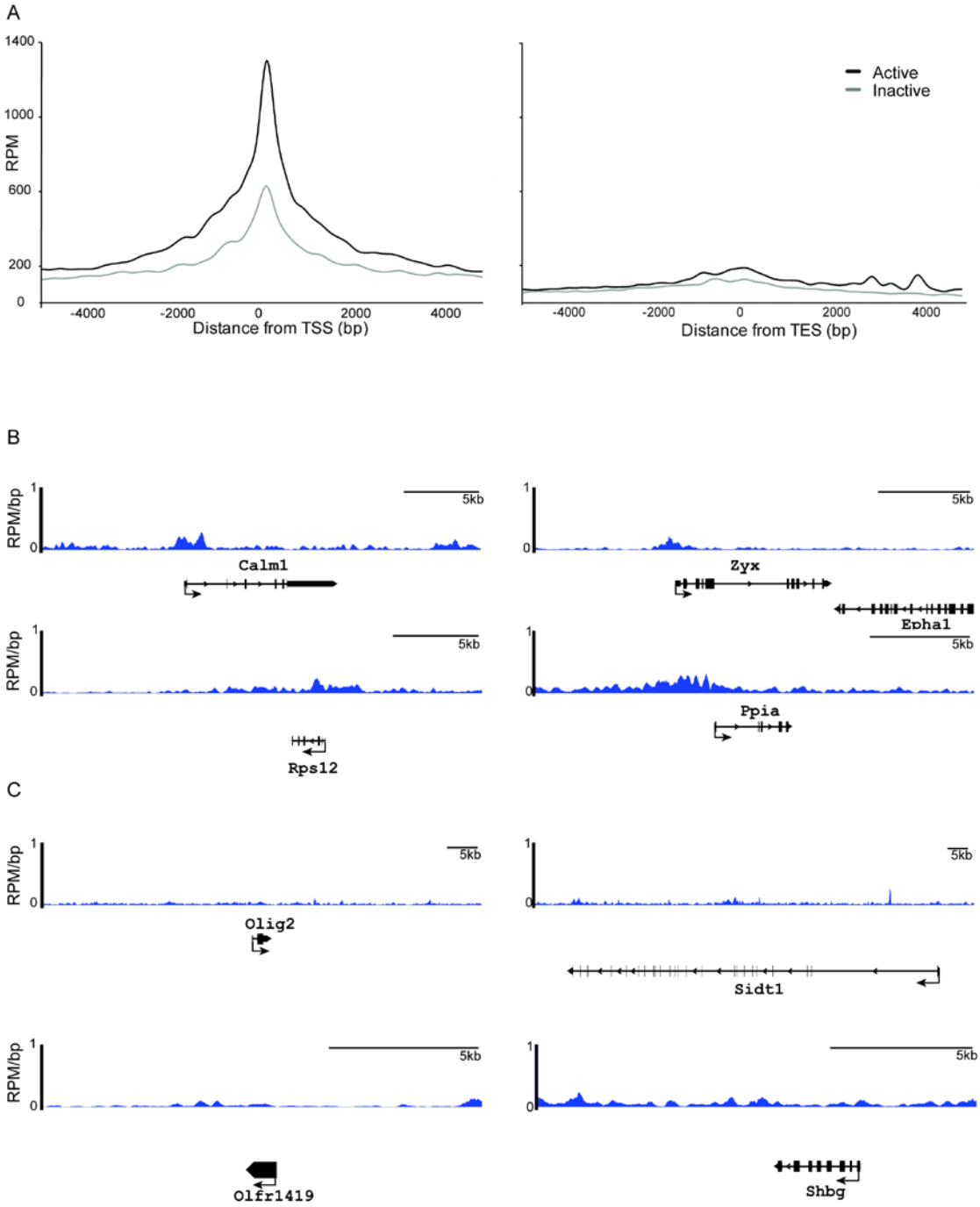

A
1400
1000
600
200
0
RPM
-4000
-2000
0
2000
4000
Distance from TSS (bp)
Active
Inactive
Distance from TES (bp)
B
RPM/bp
5kb
Calm1
Zyx
Epha1
Rps12
Ppia
C
Olig2
Sidt1
Olfr1419
Shbg

Figure 3.8 *H4K16ac has higher peaks on the promoters of active genes than inactive genes, and is less prevalent on TES of active and inactive genes.* (Previous page)

A: Normalised (RPM) tag counts of H4K16ac surrounding the TSS (*left*) and TES (*right*) for active and inactive genes. Each group represents the 500 genes with the highest or lowest expression measured by Illumina bead chip expression array.

B: Wiggle tracks plotted against the mm9 mouse genome build using IGB programme for a selection of active genes determined by expression array. Direction of transcription is indicated by arrow. Wiggle files show raw aligned sequencing data; height of signal represents the sequencing depth of ChIP'd fragments in reads per million per base pair. Data displayed over 200bp sliding windows with 20bp step, generated in WCB Edinburgh Galaxy server. C: As B, for unexpressed genes.

Given that H4K16ac is generally higher on transcriptionally active promoters (although there are small peaks on found on some inactive promoters), I decided to examine whether there is a difference in H4K16 acetylation levels on CpG promoters compared to non-CpG promoters. Approximately 70% of promoters in the mammalian genome are classified as CpG (reviewed in (Illingworth and Bird, 2009)), and CpG islands generally fall within transcriptionally permissive chromatin. Additionally, CpG rich chromatin is enriched for H4 and H3 acetylation in HeLa cells (Tazi and Bird, 1990). It would therefore be expected that CpG promoters would contain a higher level of H4K16ac, and this is the case for total promoters (Fig 3.9A) (p-value by wilcox test $<2.2x10^{-16}$). Interestingly the higher level of H4K16ac on CpG promoters is maintained, even when only the 1000 most active promoters in UD ES cells are selected (Fig 3.9B) (p-value by wilcox test $<2.2x10^{-16}$).

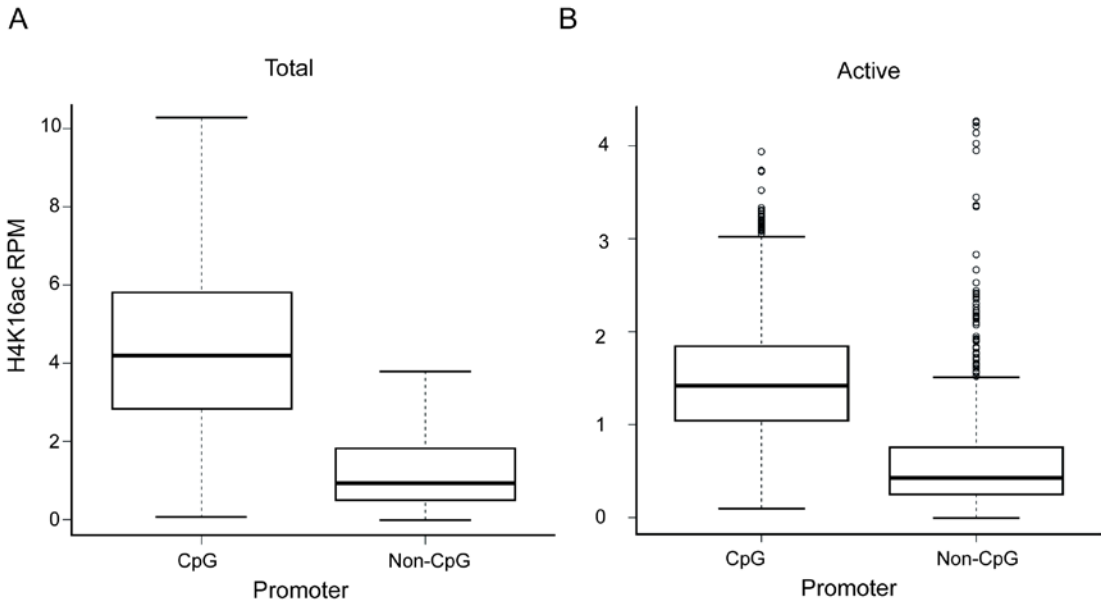


Figure 3.9 *H4K16ac is higher on CpG promoters than non-CpG promoters*

A: Boxplot representing H4K16ac depth over 500bp windows across TSSs which overlap mouse CpG islands (Illingworth et al., 2010) or non-overlapping TSSs. Sequencing depth normalised to reads per million reads. B: As A, for 1000 highest expressed genes in UD ES cells, according to mean probe intensity from Illumina MouseWG-6 v2.0 Expression Bead Chip. Circles outside boxplots represent outliers.

3.4 H4K16ac profile across promoters and distribution across the genome

To generate a picture of how H4K16ac is distributed across the gene, I produced tag density graphs across an average gene, by dividing each gene into percentage intervals and calculating the tag density within these intervals, and within 200bp windows to 2kb upstream of the TSS and downstream of the TES (Fig 3.10). I also performed this analysis for H3K27ac, H3K4me3, H3K36me3, H3K4me3 and H3K27me3 both as control for my analysis (H3K36me3 is known to accumulate at the 3' end of active genes in mammals (Bannister et al., 2005)) and to compare the H4K16ac profile to other "active" or "inactive" associated histone marks.

As in $CD4^+$T cells (Wang et al., 2008), H4K16ac has a much broader profile than H3K4me3, or H3K27ac, both of which were shown to mark the area surrounding the TSS of active genes (Barski et al., 2007; Wang et al., 2008). Both of these histone marks show a distinct peak directly over the transcription start site, with tag density beginning to increase less than 1kb from the TSS, and sharply decreasing again after it. H4K16ac, by contrast, shows a much broader peak, with tag density beginning to increase at least 2kb from the TSS, and only reaching background level again by approximately 40-50% of the way through the gene. In this way, H4K16 acetylation over active promoters seems to have a more similar profile to H3K27me3, which also shows a broad peak over the TSS – albeit for inactive genes. All histone marks found on active promoters show a dip at the TSS, representing the nucleosome free region found at these promoters (but not at inactive promoters) (Schones et al., 2008).

This profile presents a challenge for peak calling algorithms (to be discussed in Chapter 4), since most were developed with transcription factors in mind, and so expect a much narrower and higher peak than my H4K16ac data is likely to provide.

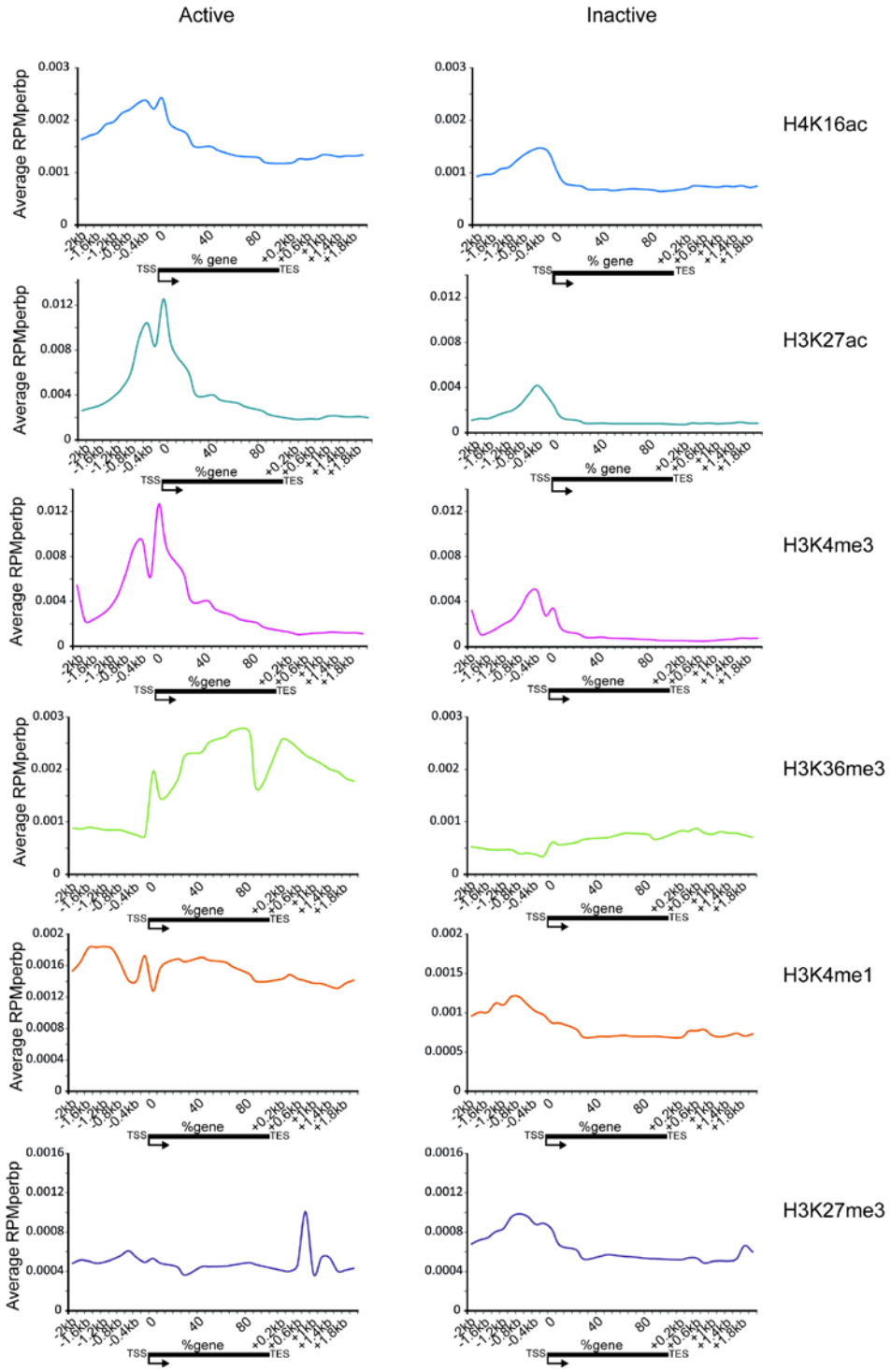
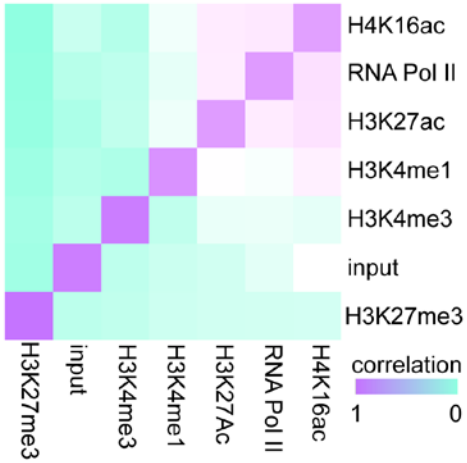


Figure 3.10 *Average profiles of histone marks across active/inactive genes*

Normalised (RPM/bp) tag counts of histone modifications in UD ES cells across gene bodies of 500 active or inactive genes, extending 2kb up and downstream of the TSS/TES. Data sources; H4K16ac (this study), H3K27ac, H3K4me3, H3K4me1 (Creyghton et al., 2010), H3K36me3 and H3K27me3 (Mikkelsen et al., 2007).

Having shown that in ES cells, H4K16ac falls on the promoters, I wanted to determine how it correlates with other histone marks which associate with promoters. I determined tag density within 2kb windows across TSSs (ensembl61, February 2011, mouse NCBI37 genome build) for seven ES cell ChIP-seq datasets, and MNase digested input DNA, then calculated the Spearman correlation of the distributions using R. I plotted the correlations as a heatmap (Fig 3.11A) clustering modifications which had a similar distribution. H4K16ac and RNApolymerase II (hypophosphorylated isoform) had the most similar distributions, followed by H3K27ac and surprisingly H3K4me1, which does not mark the region over the TSS but shows peaks either side of the H3K4me3 TSS peak (Fig 3.10). H3K4me3 showed a low correlation with H4K16ac distribution, potentially due to the fact that in ES cells, a proportion of the genes marked by H3K4me3 are in fact inactive, and bivalently marked with H3K27me3 (Bernstein et al., 2006). These genes are unlikely to be marked with H4K16ac, and this would therefore reduce the correlation between these two generally active chromatin associated histone marks in ES cells. Unsurprisingly, H4K16ac shows the lowest correlation with H3K27me3, a marker of polycomb repressed genes (Cao et al., 2002). I then wanted to determine how the peaks overlap genome wide, as opposed to only on promoters, between the four histone modifications, and used the bioconductor program ChIPpeakAnno (Zhu et al., 2010) to generate Venn diagrams of the peak overlap (Fig 3.11B). Of the ~60,000 H4K16ac peaks around a third did not overlap with any of the other marks analysed, whilst ~15,000 overlap with both H3K27ac and H3K4me1. A further ~11,000 peaks overlap with only H3K4me1, and another 10,000 overlap with H3K4me3, H3K4me1 and H3K27ac. This final group likely represents active promoters, but regions with a combination of H3K27ac and H3K4me1 without H3K4me3 have been shown to represent active enhancers (Creyghton et al., 2010), presenting the possibility that some H4K16ac peaks were falling on enhancer regions.

A



B

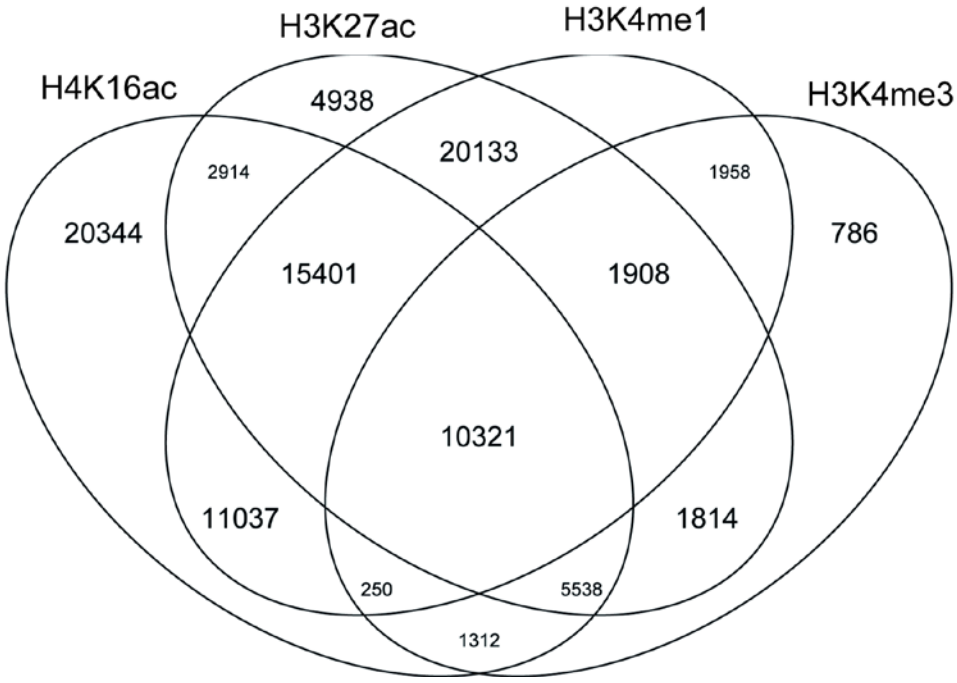


Figure 3.11 *Correlation of H4K16ac with other histone marks*

A: Clustered heatmap displaying Spearman correlation of distribution of tag densities over 2kb windows across all mouse TSSs (NCBI37) between histone modifications (and hypophosphorylated RNApolII). Correlations and heatmaps generated in R. B: Venn diagram for overlap of peaks between H416ac, H3K27ac, H3K4me1, and H3K4me3 calculated in ChIPpeakAnno (bioconductor).

Next, I wanted to determine where on the genome the majority of H4K16ac peaks fall. Previous analyses have focused around the TSS of known genes, but I wanted to use a less biased approach in order to determine whether H4K16ac has influence on other elements of the genome. I used a program called CEAS (Cis-Regulatory Element Annotation System) (http://liulab.dfci.harvard.edu/CEAS/index.html) which determines the percentage of ChIP-seq enriched regions which fall into various categories of the genome.

The bulk of H4K16 acetylation peaks (by comparison to the genome) fall within the promoters and the coding exons, though a significant percentage of peaks fall downstream of the genes, in the introns, and in the 3'UTR (Fig3.12A-C).

By comparison to H3K27ac, and H3K4me3, H4K16ac has a greater percentage of peaks in the region between 1kb and 10,000kb upstream of the TSS (Fig3.12D), though it has more similarity with H3K27ac than H3K4me3, which has a considerably greater percentage of enriched regions which fall within the area immediately surrounding the TSS – 8.3% fall within the 5'UTR and 1kb promoter region, compared with 2.2 % in the H3K27ac dataset and just 0.8% for H4K16ac.

This large percentage of H4K16ac peaks which fell without the promoter/coding region/downstream area offered the interesting prospect that there was a sub population of H4K16ac which marked something other than the promoters of active genes. Given that these were unlikely to overlap with H3K4me3, they potentially represented the population of H4K16ac peaks which showed overlap with H3K4me1 but not H3K4me3 (Fig 3.11B).

Over promoters, H4K16ac showed the greatest correlation with H3K4me1 and H3K27ac. Both these marks are well known for marking enhancers and active cell type specific enhancers respectively (Heintzman et al., 2007, 2009; Creyghton et al., 2010), and it therefore seemed logical that if H4K16ac shows a correlation with these marks around promoters, it might also correlate with them over enhancers.

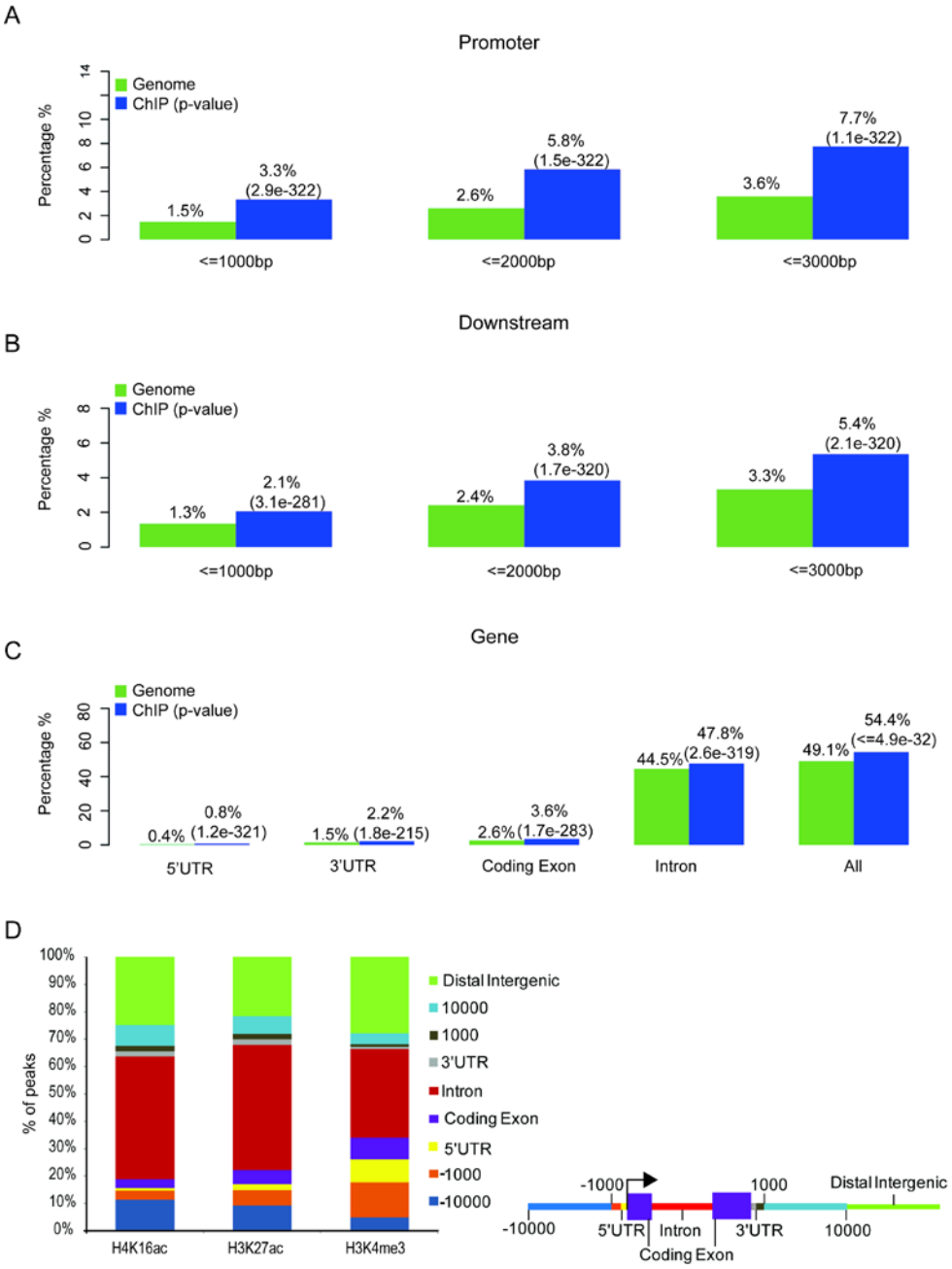


Figure 3.12 *Distribution of H4K16ac peaks*

A: Percentage of H4K16ac peaks which fall on promoters compared to expected background (mm9 genome assembly), generated in CEAS (http://liulab.dfci.harvard.edu/CEAS/index.html). Purple bars represent the percentage of mappable regions located in the specified regions. Green bars represent the percentage of ChIP peaks in located in such regions. B/C: As A, for downstream regions and coding regions. D: Percentage distribution of H4K16ac/H3K27ac/H3K4me3 peaks relative to mm9 genes (*left*), schematic for the categories into which peaks are placed (*right*).

3.5 H4K16ac on active Enhancers

The presence of H4K16ac over enhancers has been investigated before (Wang et al., 2008), but it was shown to be present on only a very small fraction (approximately 0.04) of enhancers, compared with H3K4me1, present on a fraction of 0.3 of the enhancers. The depth of H4K16ac sequencing in that early ChIP-seq experiment was very low however (a total of 7.02 million reads obtained).

In addition, there was the possibility that H4K16ac in ES cells may have a different profile to H4K16ac in $CD4^{+}$T cells from adults, marking enhancers associated with pluripotency, but not in terminally differentiated cell types.

As an initial exploration, I looked to the literature to find genetically defined enhancers known to be active in ES cells (Fig 3.13A) or enhancers that are known to be cell type specific, and looked at the profile of H4K16ac across these known enhancers.

I found that H4K16ac has a similar profile to H3K4me1 over an upstream *Nanog* enhancer (Jiang et al., 2008), which is also heavily marked by H3K27ac. Similarly, the enhancer located downstream of the *Vpreb1* locus is known to be marked by acetyl-H3 and methylated-H4 (Szutorisz et al., 2005), and primed for subsequent activation by Sox2 in undifferentiated ES cells (Liber et al., 2010). This region is marked by a domain of H4K16ac, as well as the known enhancer marks H3K27ac and H3K4me1.

I also looked at cell-type specific enhancers for differentiated cells (Fig 3.13B) and found that an enhancer which can drive *Satb2* expression specifically in callosal projection neurons (Tashiro et al., 2011) is marked, as expected, by H3K4me1, but is marked by neither H4K16ac nor H3K27ac. A neuronal specific enhancer which drives transcription of *Mnx1* (also known as *Hb9*) (Nakano et al., 2005) shows the same pattern.

These analyses indicated that H4K16ac was potentially a novel mark for (at least) ES cell specific enhancers. The next step was to determine whether H4K16ac was present on active enhancers genome wide.

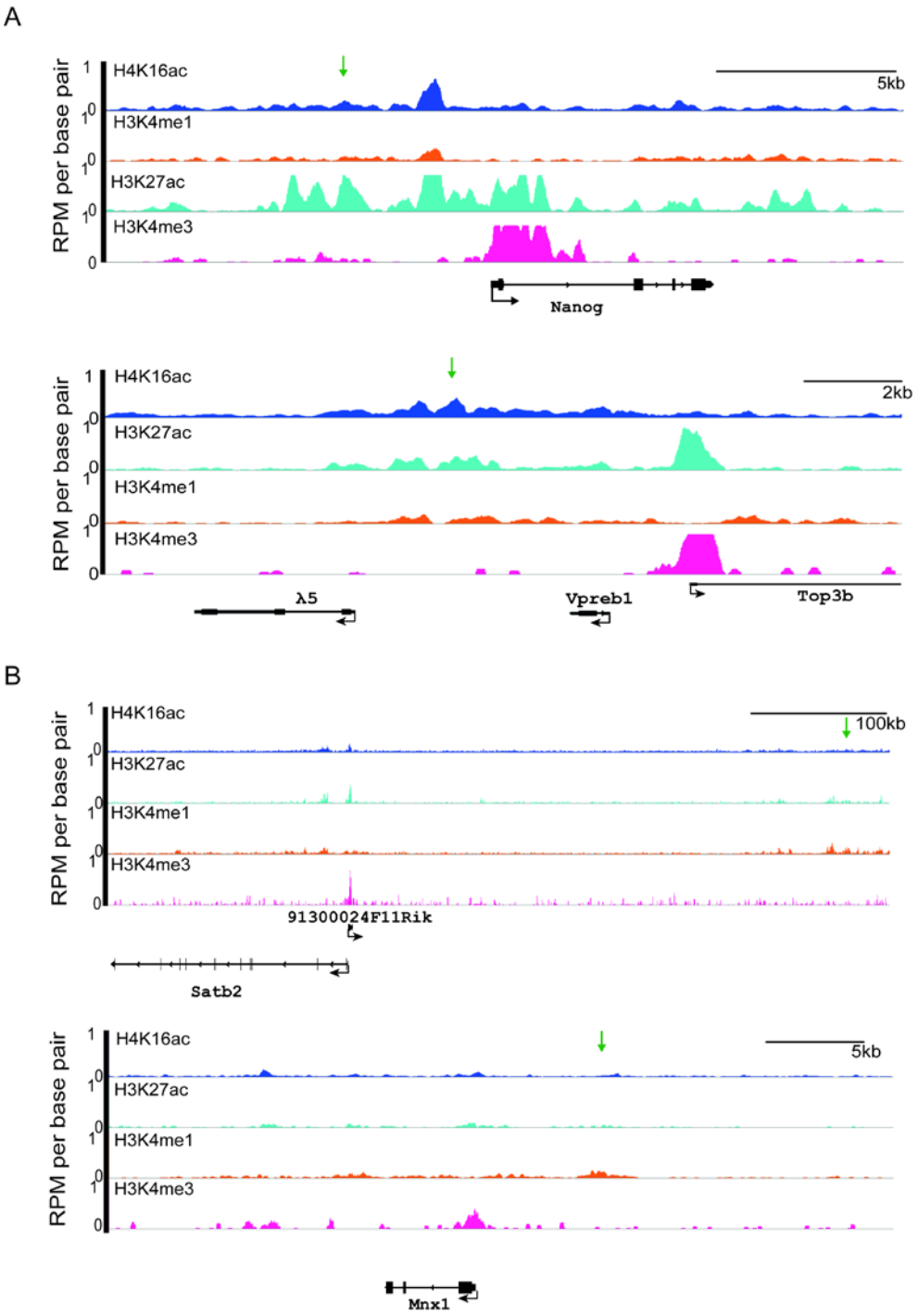


Figure 3.13 *H4K16ac profile over defined ES cell active enhancers*

Wiggle files show raw aligned sequencing data; height of signal represents the sequencing depth of ChIP'd fragments in reads per million per base pair. Data displayed over 200bp sliding windows with 20bp step, generated in WCB Edinburgh Galaxy server, for A: active ES cell enhancers and B: Lineage specific enhancers.

To test this, I generated a list of putative enhancers using publicly available ES cell histone modification data. I first used SICER to call peaks for H3K4me1 and H3K27ac, and also generated a window 2kb around the TSSs. I excluded these TSS windows, from the H3K4me1 peaks, in order to restrict my analysis to distal regions, and to exclude core promoters. Including core would confuse the analysis since H4K16ac is expected to correlate with H3K27ac in these locations.

Of these H3K4me1 peaks, I then selected those which had at least 50bp overlap with a peak of H3K27ac, or those which showed no overlap, in order to distinguish between active enhancers and inactive enhancers. I took the midpoint and generated 500bp windows extending +/- 5kb then calculated the tag numbers normalised to reads per million within them.

I then generated heatmaps clustered according to H4K16ac level over active or inactive enhancers for H4K16ac, H3K27ac, H3K4me1, H3K4me3 and Input (Fig 3.14A). The pattern of H4K16ac on enhancers follows the pattern of H3K27ac; enhancers with more H4K16ac at the top of the heatmap show an increased amount of H3K27ac, whilst those at the bottom of the heatmap, with the lowest overall amount of H4K16ac show the lowest amount of H3K27ac. Some enhancers with lower levels of H3K27ac lack H4K16ac showing that there is not complete overlap between the two marks. The profile of H3K4me1 remains relatively constant (though the broadest peaks of H4K16ac also correlate to the broadest peaks of H3K4me1) throughout the enhancers, whilst H3K4me3 and Input samples, as expected, do not show any coverage.

On the inactive enhancers (Fig 3.14A, lower) there is a lack of H3K27ac (by necessity, since that is how they are defined). The majority of these enhancers show a lack of H4K16ac too, though some few enhancers (at the top of the inactive heatmap) show H4K16ac and H3K4me1, but not H3K27ac, presenting the possibility that some enhancers are marked as active with H4K16ac instead of H3K27ac. Again, H3K4me3 and Input as negative controls show very low coverage over all enhancers.

H3K27ac marked enhancers are generally located near genes with higher expression than those which are marked only by H3K4me1 in that specific cell type (Zentner et al., 2011). I replicated this result with my enhancer set in UD ES cells, and also wanted to see whether classifying active enhancers using H4K16ac instead of H3K27ac would show this same result. There is no significant difference between the expression level of $H4K16ac^{+}/H3K4me1^{+}$ enhancers and $H3K27ac^{+}/H3K4me1^{+}$ enhancers, but there is a

significant difference between the expression level of both these sets of active enhancers and inactive enhancers with neither of the active marks (Fig 3.14B).

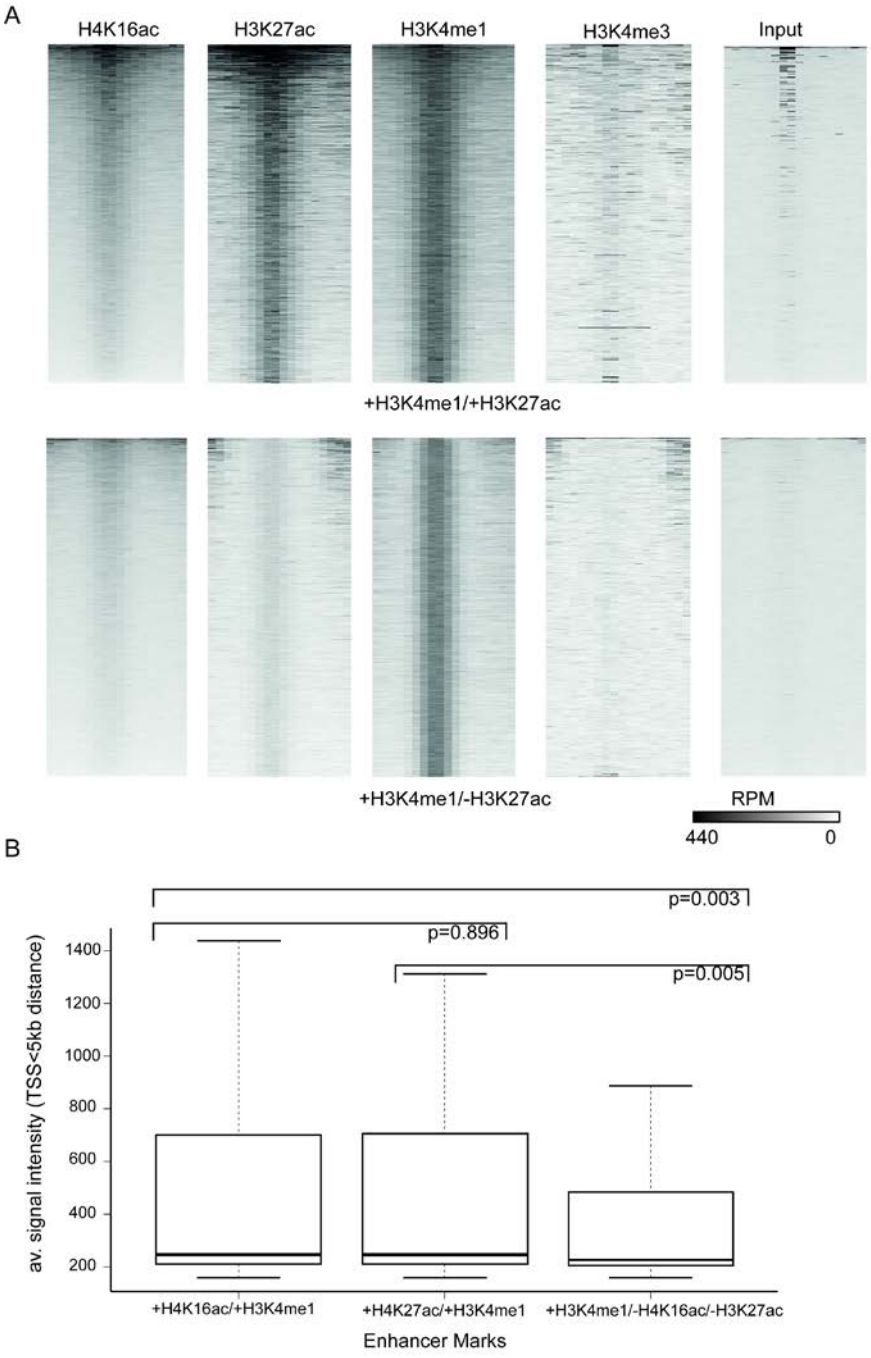


Figure 3.14 *H4K16ac profile over total ES cell enhancers*

A: Heatmap showing levels of H4K16ac, H3K27ac, H3K4me1 and H3K4me3 over active ($H3K4me1^{+}/H3K27ac^{+}$/distal from TSS) or inactive ($H3K4me1^{+}/H3K27ac^{-}$ /distal from TSS). Intensities are determined by RPM tag density in 500bp windows +/- 5kb from enhancer midpoint. B: Boxplot of expression levels of closet gene with TSS <5kb from enhancers of different classes. P-values calculated by Wilcox test using R.

To further investigate the presence of H4K16ac on enhancers, I defined active and inactive ES cell enhancers in a different way, without relying on the presence or absence of H3K27ac. Active enhancers are not only marked by histone modifications, but are also bound by transcription factors. For example in ES cells, combinatorial binding of *Pou5f1*, *Sox2* and *Nanog* can predict enhancers which are associated with actively transcribed genes (Göke et al., 2011). I therefore took a group of ES cell transcription factor/transcription mediator (*Pou5f1/Med1/Sox2/Nanog*) binding sites (Whyte et al., 2012) and overlapped them with H3K4me1 peaks from the same study to generate a set of ES cell transcription factor bound enhancers. For comparison, I also generated a list of unbound enhancers, and plotted the average tag density around the enhancer midpoint (Fig 3.15). For TF bound enhancers, as expected there is a peak of H3K4me1 and H3K27ac, but no H3K4me3, and there is a clear peak of H4K16ac over these enhancers also. For the non-TF bound set, there is still a clear peak of H3K4me1, but the peaks of H3K27ac and H4K16ac are flattened.

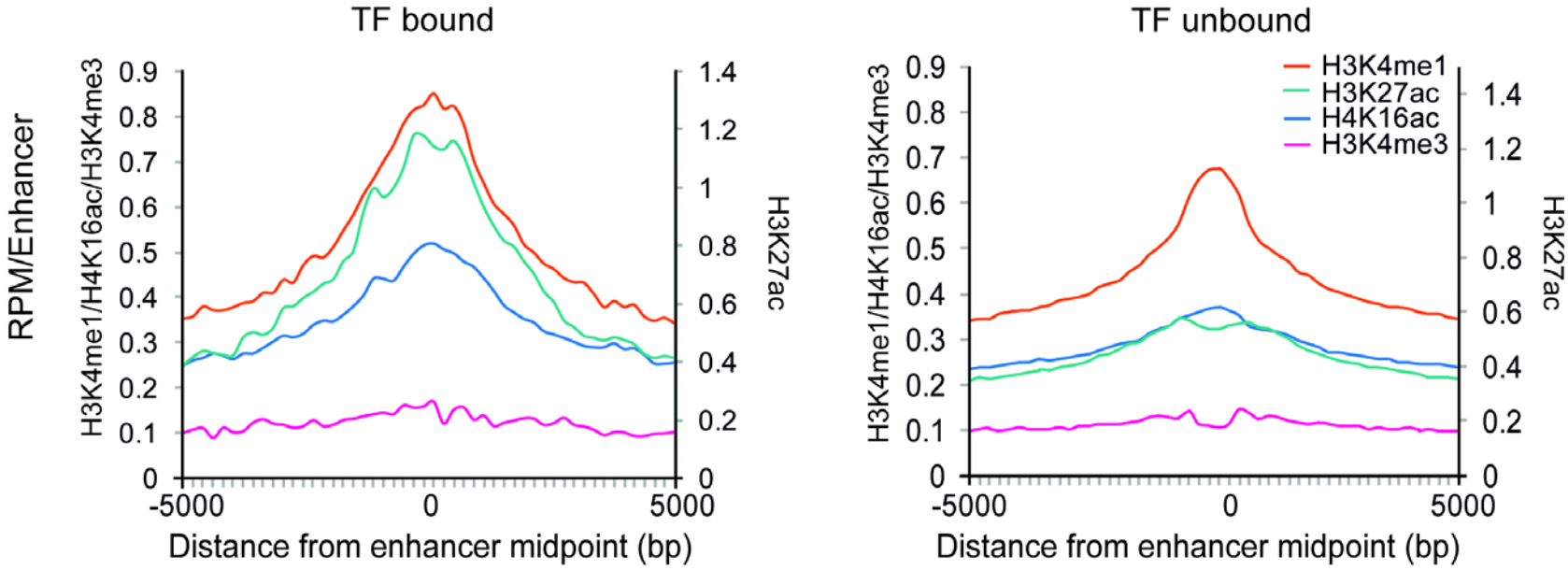


Figure 3.15 *H4K16ac profile over TF bound or unbound enhancers*

Tag density for H4K16ac, H3K4me1, H3K27ac, and H3K4me3, across 10kb region around enhancer midpoint for distal peaks of H3K4me1 (Whyte et al., 2012) which overlap peaks of Pou5f1, Sox2, Nanog and Med1 from the same study (*left*) or distal H3K4me1 peaks which do not overlap any TF (*right*).

I had noticed that on some genes, the enhancer peak of H4K16ac was much higher than the promoter peak of the nearest active gene. I therefore wanted to investigate whether this was a global pattern. To do this, I took the list of defined H4K16ac peaks in UD ES cells, and defined those which overlapped a peak of H3K4me3 as promoter, and those which overlapped peaks of H3K4me1 and H3K27ac as enhancer peaks. I then plotted the tag density of the peak in reads per base pair (Fig 3.16A). Tag density is generally higher for promoter peaks, and overall, promoter peaks are wider than enhancer peaks, meaning that peaks of H4K16ac which overlap a promoter are generally higher than those which mark an enhancer. This is illustrated in Figure 3.16B, the enhancer found towards the 3' end of *Mgst3* has a much lower peak of H4K16ac than the two flanking promoters of *Aldh9a1* and *Mgst3*. Similarly, there is a small peak of H4K16ac over an enhancer between *Lyve1* and *Mrvi1*, which is considerably smaller than the peak over the nearby gene *Rnf141*.

Though this is overall the case, it is not the case for every enhancer/promoter pairing. For example, the enhancer region downstream of *Klf4* (Shen et al., 2012), shows a broad region of H4K16ac/H3K27ac/H3K4me1 approximately 20kb wide, and which peaks much higher than the H4K16ac peak over the gene itself. As can be seen by the broad pattern in Figure 3.14A, this type of profile is the exception, rather than the rule, and it remains to be seen whether there is any significance to this type of enhancer marking.

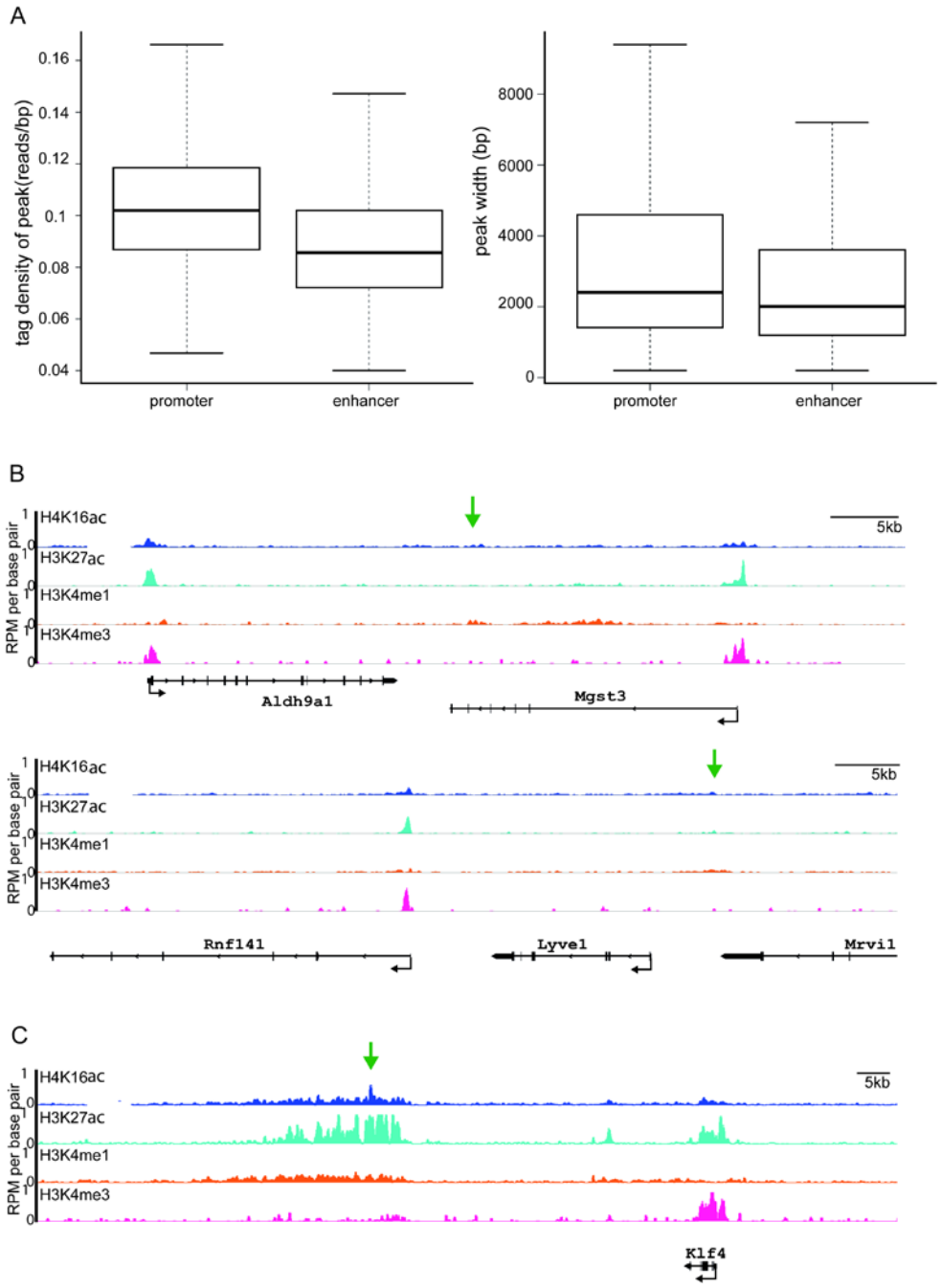


Figure 3.16 *Tag density of promoter and enhancer peaks*

A: Boxplot representing tag density (RPM/bp) of H4K16ac peaks which overlap promoter or enhancer (*left*). Boxplot representing peak width (bp) of H4K16ac peaks which overlap promoter or enhancer (*right*). B, C: Wiggle files show raw aligned sequencing data; height of signal represents the sequencing depth of ChIP'd fragments in reads per million per base pair. Data displayed over 200bp sliding windows with 20bp step, generated in WCB Edinburgh Galaxy server, for B: Promoter/Enhancer pairs which show the same pattern as genome wide average (lower peak on enhancer than promoter for H4K16ac) and C: Inverse promoter/enhancer pair with higher H4K16ac peak on enhancer.

3.6 MOF/Myst1 is also found at active promoters and enhancers

MOF is the specific histone acetyl transferase (HAT) responsible for H4K16ac (Thomas et al., 2008), so it made sense to compare its profile across TSS and enhancer regions with p300, which is one of the major HATs which catalyse acetylation of H3K27 in ES cells (Fig 3.17) (Pasini et al., 2010) . I downloaded publicly available ES cell ChIP datasets for MOF (Li et al., 2012) and p300 (Creyghton et al., 2010), and generated tag density graphs around the TSS of active and inactive genes (as in Figure 3.8A). As expected from the profiles of their respective histone marks, binding of both HATs is increased around the TSS of active genes compared with inactive genes (Fig 3.17A) (p-value by Wilcox test for a 500bp region around the TSS $<2.2x10^{-16}$), though p300 peaks are higher than those of MOF. This is expected, since the profile of H3K27ac shows higher, narrower peaks around the TSS of active genes than H4K16ac, which has broader, lower peaks over promoters (see section 3.5). I then went on to perform the same analysis over active and inactive enhancers. The results mirror those of H4K16ac and H3K27ac; both HATs show a distinct peak over active but not inactive enhancers (p-value by Wilcox test for a 600bp region around the TSS $<2.2x10^{-16}$). The enhancer upstream of *Nanog* is shown as an example in Figure 3.17C. A broad domain of H3K4me1, approximately 10kb wide, overlaps with peaks of H3K27ac and H4K16ac. The profile of MOF over this region matches the profile of H4K16ac very well. H3K27ac distribution is similar to that of p300 though there is a peak of H3K27ac approximately 12.5kb upstream of the *Nanog* TSS, which is not accompanied by a peak of p300. This suggests that a different H3K27ac HAT may be acting there. Whilst a small peak of MOF overlaps with this H3K27ac peak, there is no evidence that MOF can acetylate any residues on the tail of H3.

The presence of MOF over active enhancers from an independent dataset adds further evidence for the presence of H4K16ac on active enhancers.

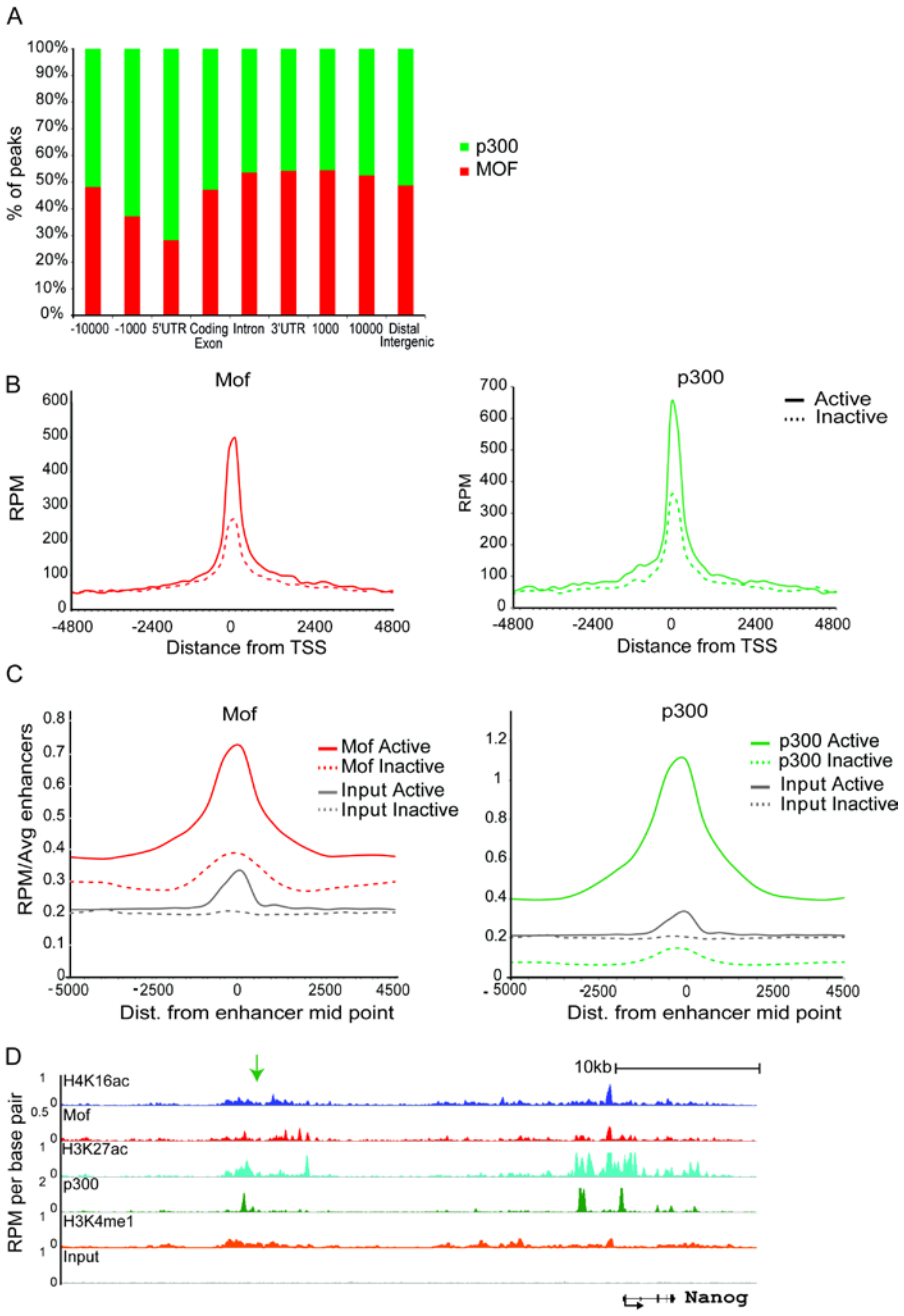


Figure 3.17 *MOF/Myst1 marks active promoters and enhancers in ES cells*

A: Percentage distribution of MOF peaks relative to p300 peaks across genome divided into 9 categories. B: Normalised (RPM) tag counts of MOF (*left*) and p300 (*right*) surrounding the TSS for active and inactive genes. Each group represents 500 genes with high or low expression. C: Normalised (RPM) tag counts of MOF (*left*) and p300 (*right*) with input (grey) +/- 5kb over active/inactive enhancer midpoints. D: Wiggle files at *Nanog* locus, for H4K16ac and MOF, H3K27ac and p300, H3K4me1, and input normalised to RPM/bp for 200bp windows with 20bp step.

3.7 Summary of H4K16ac profile in UD ES cells

From these results I conclude that in ES cells, H4K16ac marks the promoters of highly expressed genes, and not those of silenced genes, as was previously determined in human $CD4^{+}$T cells. In addition, I have found that H4K16ac marks the active enhancers of ES cells.

It would be interesting to investigate whether H4 tail acetylation on enhancers is specific to H4K16ac, or whether the other H4 acetylations are present there as well. Unfortunately, this cannot be investigated until ChIP-seq profiles are available for these marks in ES cells.

Next, to look at how changes in expression were associated with changes in H4K16ac profile, I looked at differentiation of ES cells along the neuronal pathway using retinoic acid.

Chapter 4: Profile of H4K16ac changes upon ES cell differentiation

4.1 Introduction

Having found that H4K16ac is found on the promoters of active genes and on active enhancers in undifferentiated embryonic stem cells (UD ES cells), I decided to investigate how the profile of H4K16ac changes genome-wide upon ES cell differentiation.

Retinoic acid (RA) treatment induces differentiation along the neuronal pathway (Bain et al., 1995), and causes expression of the *Hox* genes in temporal colinearity in mouse ES cells (Chambeyron and Bickmore, 2004). I used three days of RA differentiation, by which stage there is polarised activation of the 3' *Hox* genes (Morey et al., 2009), loss of H3K27me3 over the early *Hoxb* and *Hoxd* genes, along with chromatin decompation in the region (Eskeland et al., 2010).

At later stages of RA induced differentiation, *Mof* expression is lowered and subsequently eliminated, but by day 3, expression is still at levels approximately equivalent to UD ES cells (Li et al., 2012) (and this study) and could therefore still be important for the changes in expression which characterise the shift into the neuronal lineage.

In addition to determining whether H4K16ac presence on promoters correlates with gene activation in differentiated ES cells, I also wanted to observe the behaviour of the active enhancers upon differentiation. This was in order to determine whether H4K16ac, like H3K27ac shows cell type dependent presence on active enhancers. H3K27ac is lost from the enhancers of genes which are silenced upon differentiation, and placed onto enhancers close to genes which gain expression (Zentner et al., 2011). H4K16ac has not been shown to mark enhancers in another differentiated cell type (Wang et al., 2008) and MOF is particularly important for gene expression in UD ES cells, and H4K16ac levels gradually diminish during retinoic acid (Li et al., 2012). It was therefore a possibility that H4K16ac may be important only for active enhancers in UD ES cells.

4.2 RA induced differentiation of ES cells

I induced differentiation with RA for 3 days (Fig 4.1A), which causes a change in morphology of ES cells from rounded colonies into more flattened and elongated cells. The changes occur across the majority of the cell population and it can be easily verified under light microscope, indicating a successful differentiation prior to usage of the cells in further experiments (Fig 4.1B and C).

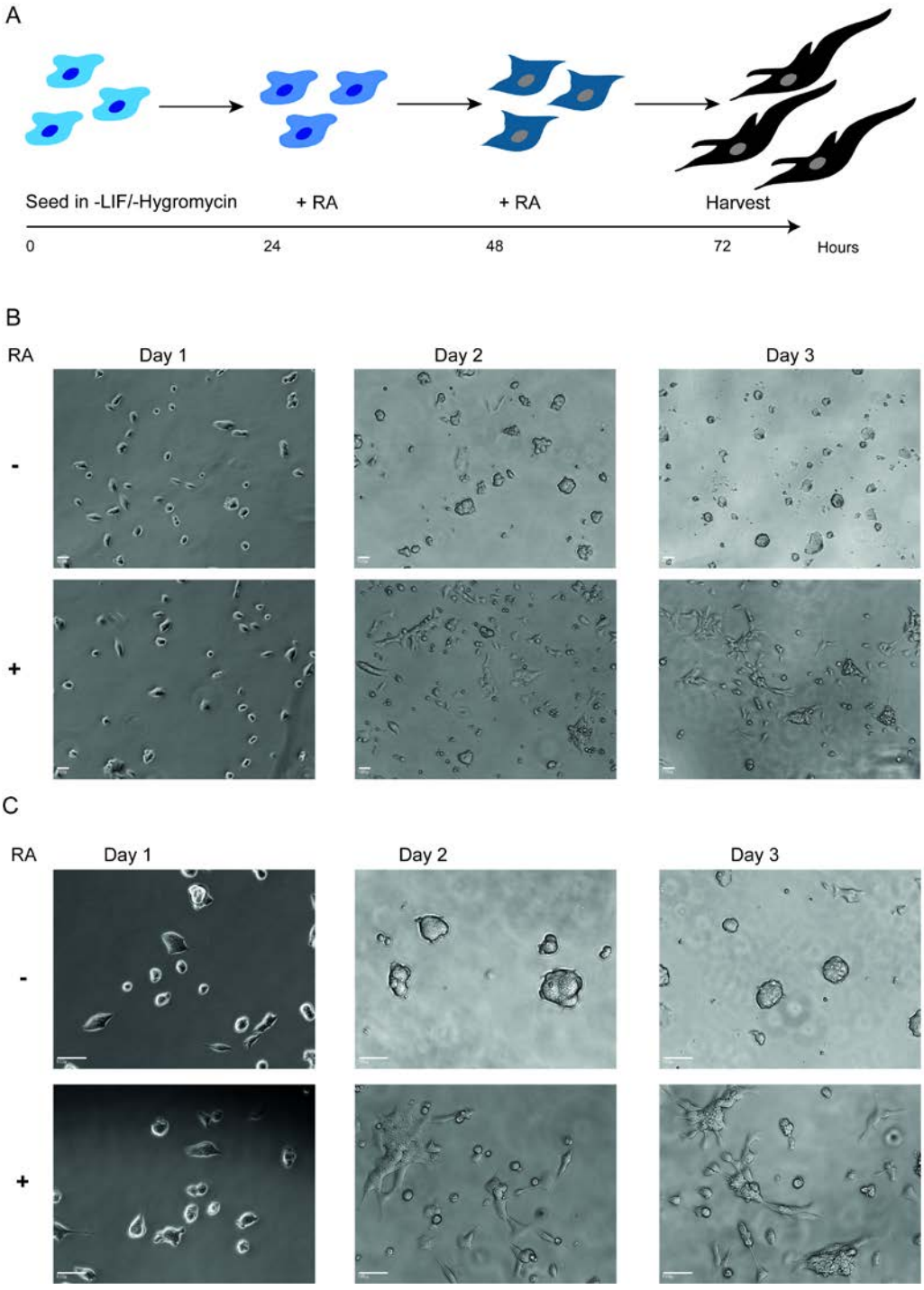


Figure 4.1: ES cell differentiation using retinoic acid (RA)

A: Schematic representing differentiation protocol. B: Morphological changes associated with RA differentiation in mouse OS25 cells at 4x magnification by phase contrast microscopy C: Morphological changes associated with RA differentiation in mouse OS25 cells at 10x magnification. (scale bar 100µm)

To determine whether the H4K16ac profile changes in line with differentiation induced changes in gene expression, I extracted RNAs from undifferentiated ES cells and from cells after 3 days of RA differentiation (D3). I then labelled them with Cy3 before hybridizing them to Agilent whole mouse genome (4x44k) oligo microarrays.

After data extraction and quantile normalisation, I validated genes upregulated in UD or D3 by RT-PCR (Fig. 4.2A), then took genes which were significantly increased in expression in either UD or D3 samples (adjusted p-value of <0.005). I then used GOrilla (Eden et al., 2009) to determine enriched GO terms compared with total probes on the array. As expected from previous studies of RA differentiation, the enriched GO terms in genes comparatively upregulated in UD ES cells included stem cell maintenance and negative regulation of cell differentiation. Conversely, in the D3 sample, GO terms derived from the expression of *Hox* genes such as skeletal system morphogenesis and anatomical structure morphogenesis were highly enriched (Fig 4.2B).

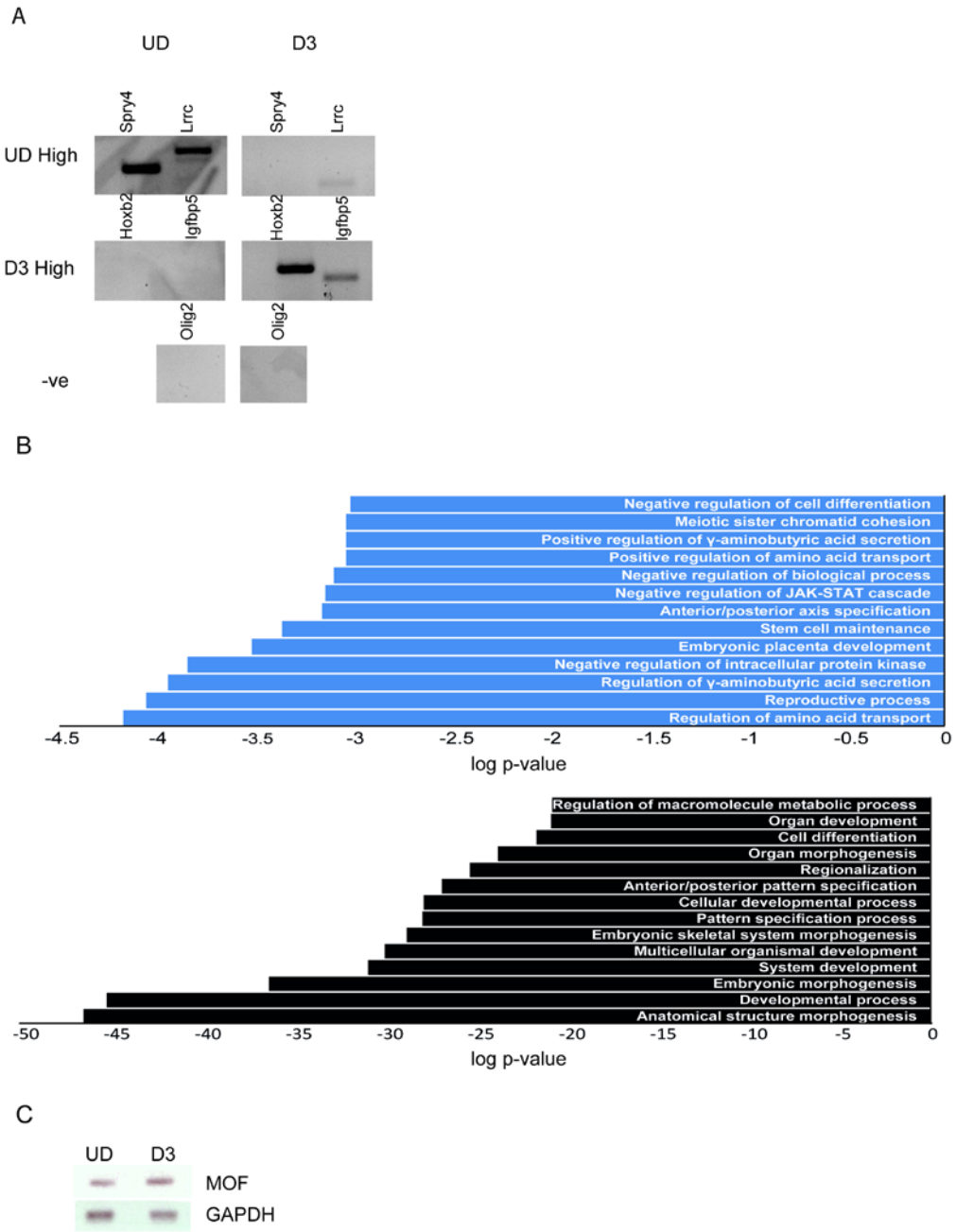


Figure 4.2*: Expression changes upon RA differentiation*

A: RT-PCR for expression of genes with highest p-value upregulated in UD/D3 cells. B: Functional annotation of genes upregulated in UD (top) or D3 (bottom) using GOrilla for biological process GO terms. Genes upregulated with adjusted p-value of <0.005, analysed against total array probes. Log_{10} of enrichment p-value is displayed. C: RT-PCR for expression of *Mof* in OS25 UD and D3 ES cells.

4.3 H4K16ac profile changes upon ES cell differentiation

I performed ChIP for H4K16ac as previously, in UD and D3 ES cell populations. I then looked for its presence or absence by qRT-PCR, at the promoters of genes which were either newly expressed (*Hoxb1*, *Hoxb13*, *Hoxd1*), or reduced in expression (*Pou5f1*) in D3 cells. I also examined *β-actin* (positive control) and *Olig2* (negative control). Testing of these genes by RT-PCR verified their expected expression status (Fig 4.3A). Upon differentiation, H4K16ac remains largely unchanged on the promoters of *β-actin*, *Olig2*, and, surprisingly, *Pou5f1*. Given the reduction in expression of *Pou5f1*, I expected that it would lose acetylation on the promoter, but this is not the case (Fig 4.3B).

These results were verified across larger areas using custom Nimblegen arrays. *β-actin* is acetylated in both cell types, whilst *Olig2* is acetylated in neither (Fig 4.3C). Other pluripotency genes (*Nanog* and *Sox2*) show the expected pattern of loss of H4K16ac upon differentiation which was not seen in *Pou5f1* (Fig 4.3D).

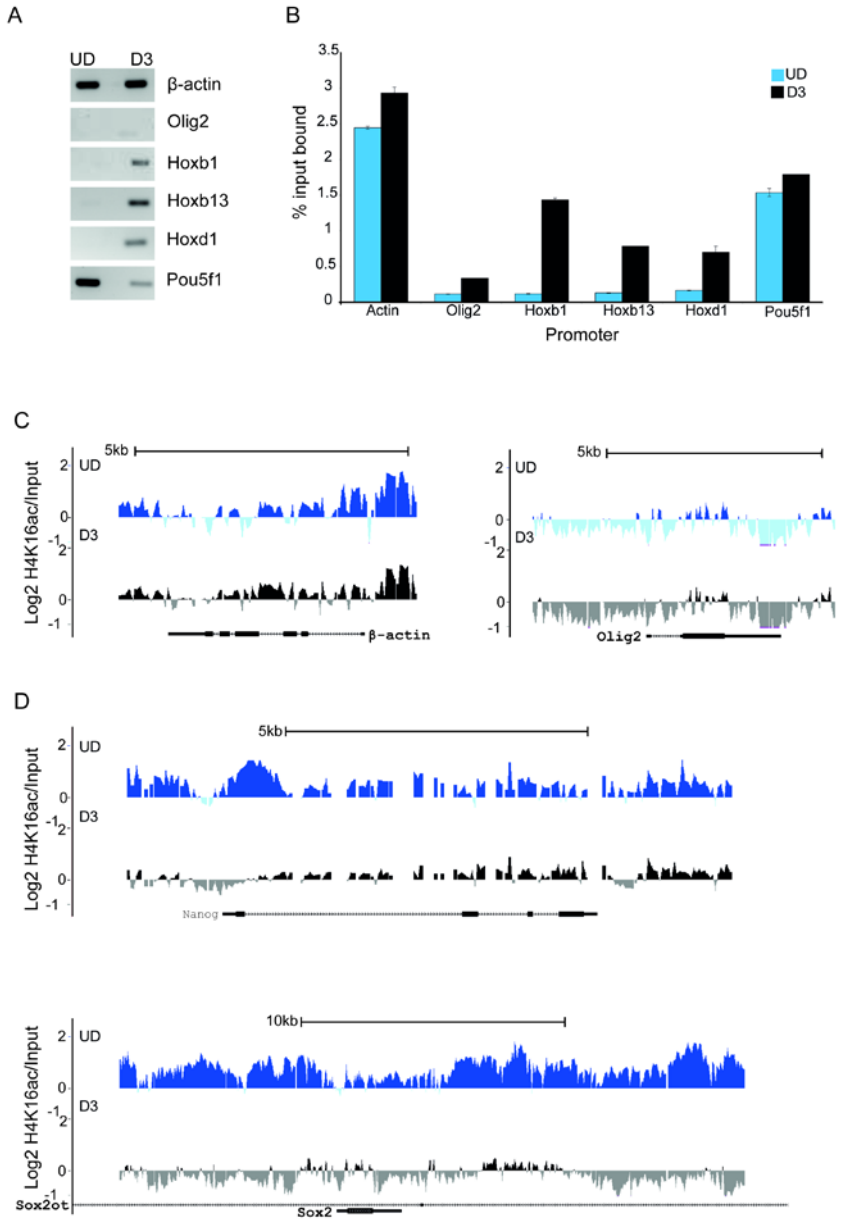


Figure 4.3*: H4K16ac follows changes in expression of cell-type specific genes upon differentiation*

A: RT-PCR for expression of *β-actin, Olig2, Hoxb1, Hoxb13, Hoxd1, Pou5f1* in UD and D3 OS25 cells. B: ChIP for H4K16ac at active or inactive promoters in UD (blue) and D3 (black) OS25s assayed by qRT-PCR. Enrichment is shown as mean % input bound +/- SEM over three biological and three technical replicates. C: Log_2H4K16ac/Input in UD/D3 ES cells over *β-actin* and *Olig2* on custom Nimblegen arrays. RefSeq gene annotations are from the July 2007 (mm9) Build 37 assembly of the mouse genome (http://genome.ucsc.edu). D: As C, for pluripotency factors *Sox2* (top) and *Nanog* (bottom).

The patterns seen over the four *Hox* loci (Fig 4.4) are broadly similar. H4K16ac is absent from *Hox* loci in undifferentiated ES cells, but upon differentiation, the early *Hox* genes become acetylated. By contrast, the later *Hox* genes remain unacetylated. For example, in the *Hox* a locus, the novel acetylation in the D3 sample reaches to *Hoxa7*. There is a peak of H4K16ac in both samples in the centre of the locus, the function of which is unclear, and after which, H4K16ac is excluded in both samples.

In the *Hoxb* locus, novel acetylation covers the region from *Hoxb1* to *Hoxb6*. Interestingly, this is the same region which loses H3K27me3 in D3 samples also (Eskeland et al., 2010). The *Hoxc* locus shows a clear region of H4K16ac in the D3 sample, which stops abruptly upstream of *Hoxc6*.

The *Hoxd* locus shows the smallest change in H4K16ac, with a low amount of acetylation over the early *Hoxd* genes.

Broadly speaking, I found that H4K16ac follows changes in expression. I then wanted to look at the association with expression on a genome wide level. I therefore generated a ChIP-seq dataset for H4K16ac in D3 cells.

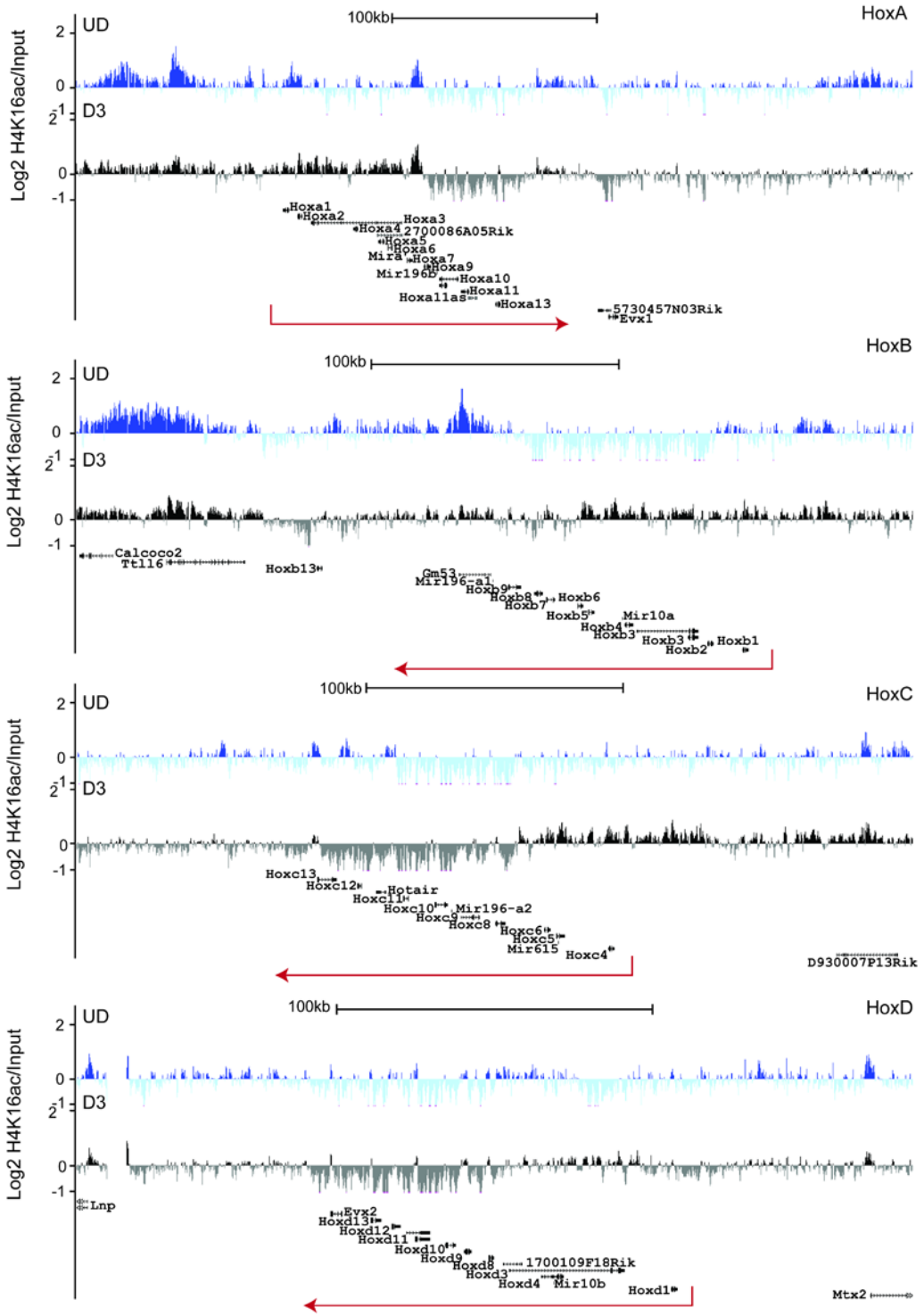


Figure 4.4: *H4K16ac is gained over early Hox loci upon differentiation*

Log_2H4K16ac/Input in UD/D3 ES cells over the four mouse Hox loci on custom Nimblegen arrays. Direction of temporal colinearity is indicated with red arrow. RefSeq gene annotations are from the July 2007 (mm9) Build 37 assembly of the mouse genome (http://genome.ucsc.edu).

4.4 Quality Control for H4K16ac ChIP-seq in D3 cells

As with the UD sample (Fig 3.4) I used FASTQC to determine sequence quality of my ChIP-seq data. I found that in the initial sample the sequencing was of poor quality, with many of the reads falling below a Phred score of 30, indicating base call accuracy of ≤99.9% in these reads (Fig 4.5A). This is not unuseable, previously published ChIP-seq datasets have scores of this level and lower (data not shown), but I made the decision to repeat the sequencing for the ChIP sample, and the second lane of sequencing yielded better results (Fig 4.5C). I combined the two samples for alignment. Those which were aligned, and therefore included in downstream analysis were of high quality (Fig 4.5D), showing that poor quality sequences had been discarded.

The number of aligned reads for the D3 sample was considerably lower than the UD samples (36 million for the ChIP sample and 57 million for input, compared with 79 million and 105 million respectively for the UD sample). Therefore normalisation would be challenging, and some results must be interpreted with caution before conclusions can be drawn.

The D3 sample requires a higher percentage of its reads for the peak calling algorithm to find the majority of the peaks it finds with 100% of the reads (approximately 60-70%) than the UD sample (around 40%), but still appears to be saturated.

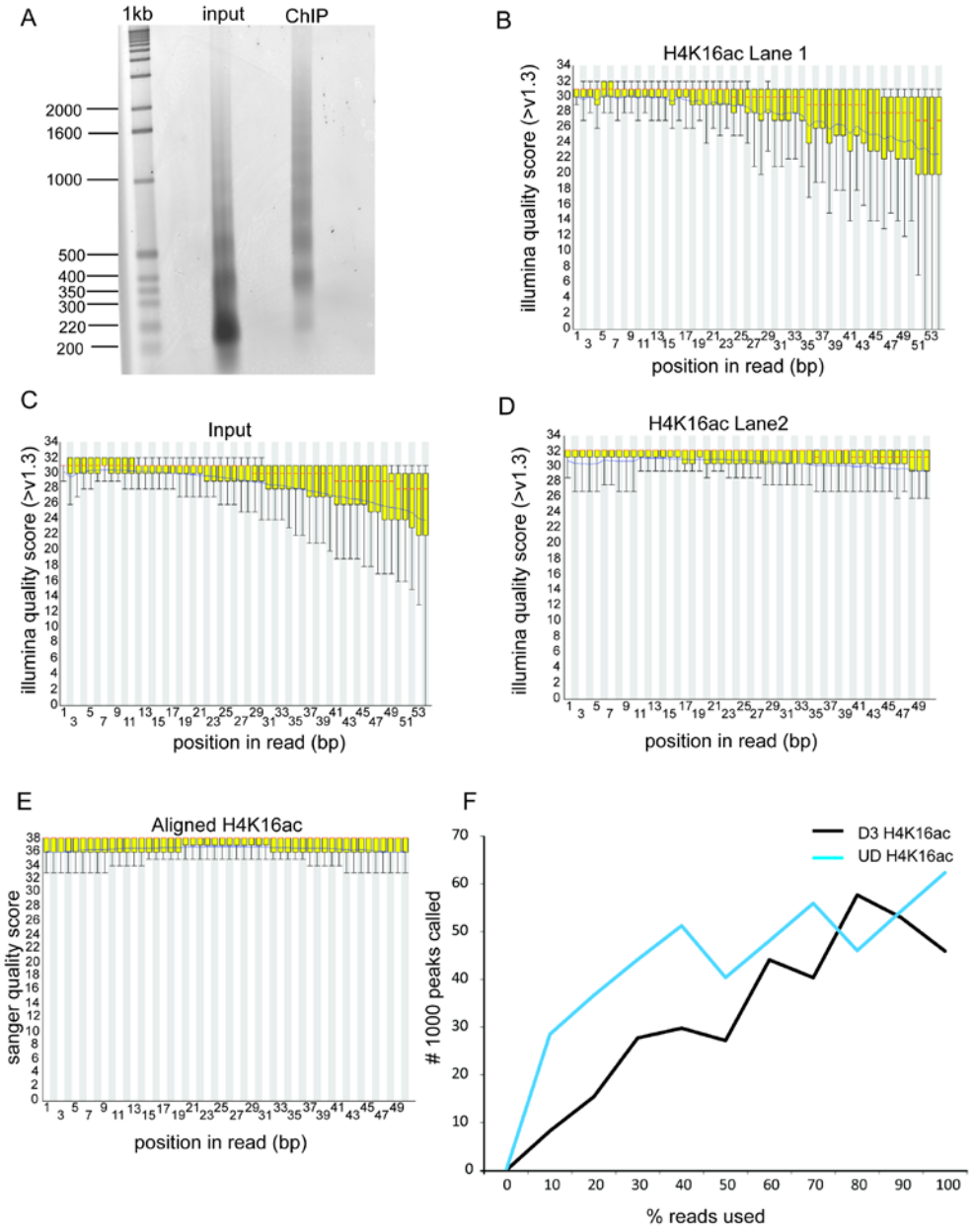


Figure 4.5*: D3 H4K16ac Sequencing Quality Control*

A: MNase digested input and H4K16ac ChIP digested with proteinase K and separated on a 1.3% Agarose gel, stained with SYBR Gold. B: Quality scores generated in FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) for all reads in first lane of D3 H4K16ac ChIP. C: As B, for D3 input DNA. D: As B, for second lane of D3 H4K16ac ChIP. E: Sanger quality scores for FASTQ output of bowtie alignment for all D3 ChIP reads. F: Saturation curves for D3 H4K16ac and UD H4K16ac (this study) . Plotted as number of peaks found using SICER (Zang et al., 2009) with increasing percentage of reads randomly sampled from dataset.

Table 4.1: Read alignment statistics.

	Total Reads	Reads Aligned to mm9	% reads aligned
D3H4K16ac	70012050	36032370	51.46595479
D3Input	70295210	57283750	81.49026086

Total number of H4K16ac and input and number of reads aligned to mm9 genome assembly using Bowtie.

4.5 Normalisation of ChIP-seq datasets

Normalisation of ChIP-seq datasets remains an open challenge. In most publications, ChIP-seq data is normalised either to input or, more commonly, by scaling tag density to account for the differences in sequencing depth. This is the simplest way to normalise across multiple datasets from different experiments which may or may not carry appropriate controls. I used a variety of normalisation strategies and visualised the effect on the appearance of H4K16ac around the positive control *β-actin* promoter, and *Hoxb1*, which by ChIP on chip shows an increase of H4K16ac upon differentiation.

In all cases, the gain of H4K16ac at *Hoxb1* in the differentiated sample (even when visualising raw data) is clear. On the *β-actin* promoter however, normalisation affects whether a difference is detected in the level of H4K16ac between UD and D3 samples. In the raw sample, the level of UD H4K16ac is over 3x that of the D3 sample (reflecting the difference in sequencing depth). After normalising to sequencing depth (Fig 4.6B), the difference is reduced to approximately 2 fold, and is eliminated entirely by quantile normalisation (Fig 4.6C). After the quantile normalised sample is normalised to input (Log_2 ChIP/Input), the difference remains at approximately 2 fold.

It was unclear, however, whether the broad reduction of H4K16ac over the *β-actin* promoter might represent a biological global reduction in H4K16ac upon differentiation (excluding those positions which gain it), given the reduction in MOF and H4K16ac seen by Li and colleagues (2012). Therefore I chose to use simple normalisation to sequencing depth for all visualisations. Recent work has shown a novel way to normalise ChIP-sequencing data which takes into account the expectations of peaks in ChIP samples, versus the lack in input samples (Diaz et al., 2012), which could be used future analysis.

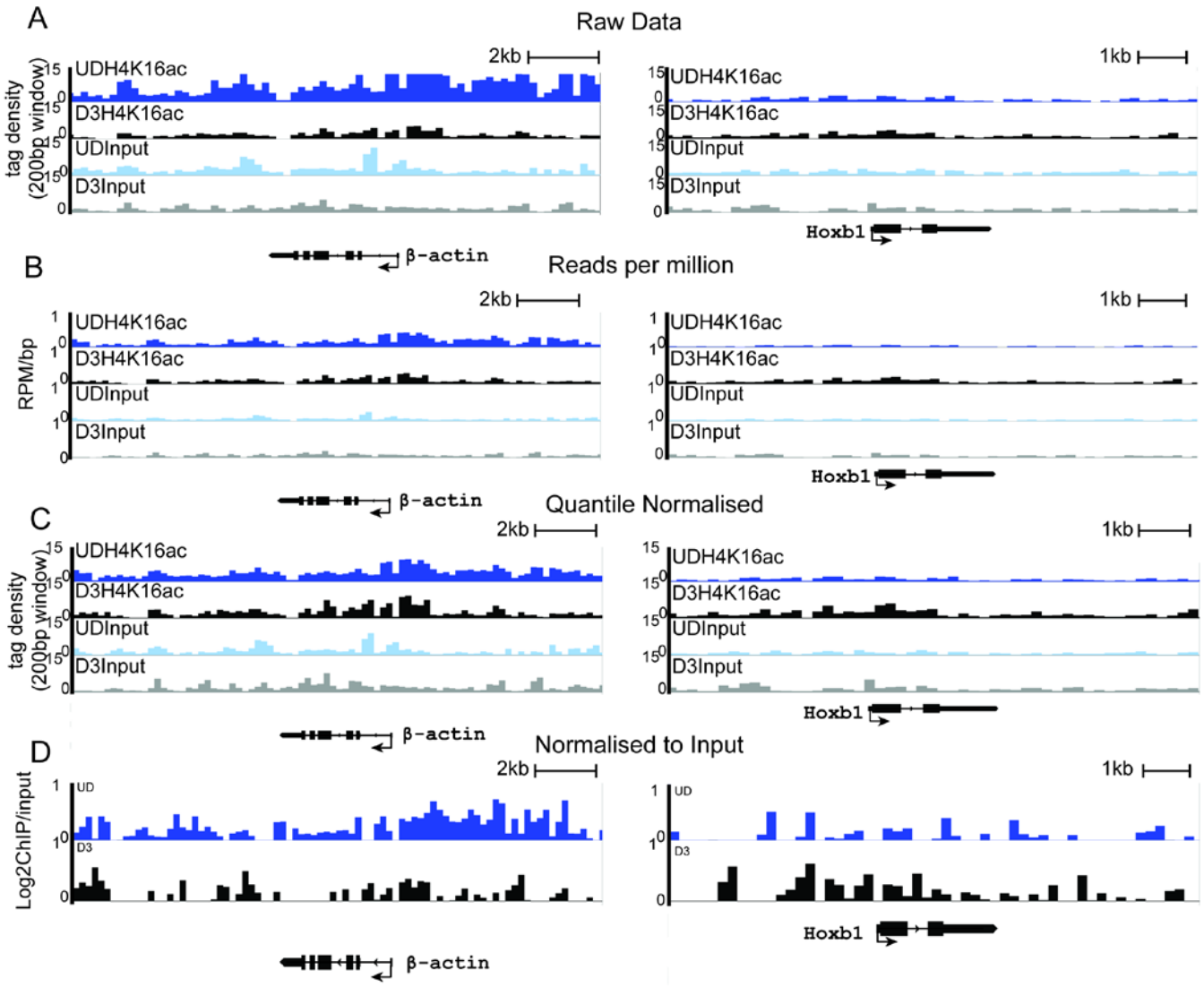

A
Raw Data
2kb
1kb
tag density (200bp window)
UDH4K16ac
D3H4K16ac
UDInput
D3Input
Hoxb1
β-actin
B
Reads per million
RPM/bp
C
Quantile Normalised
D
Normalised to Input
Log2ChIP/input
UD
D3

Figure 4.6: *Three types of normalisation for ChIP-seq data* (previous page)

Wiggle tracks plotted against the mm9 mouse genome build using IGB programme in 200bp windows with 200bp step over *β-actin* (left) and *Hoxb1* (right) for A: Un-normalised data, B: Data normalised by division by #million tags in dataset, C: Quantile normalisation between UD and D3 datasets, and D: Log ratios of quantile normalised ChIP data to respective input samples (and see materials and methods for further details of normalisation).

4.6 Peak calling using NPS, MACS and SICER

Peak calling can be a difficult problem for certain histone modifications which do not form discreet peaks in the way that transcription factors generally do. I therefore tried three different methods of peak calling on my data to determine which was most suited to the broad peaks formed by H4K16ac.

Firstly, I used NPS (Nuclear Positioning from Sequencing) (Zhang et al., 2008b) to position nucleosomes from both datasets, then calculated tag numbers from each dataset within those nucleosomes. I took those which were significantly different between UD and D3 datasets, taking into account differences in sequencing depth.

Secondly, I used two published algorithms, MACS (Zhang et al., 2008a) and SICER (Zang et al., 2009). MACS is very widely used (cited by over one hundred articles), whilst SICER was designed with the variable peak width of histone modifications in mind. Both utilise input DNA to select peaks significantly above background and account for differences in sequencing depth. Numbers of peaks called are shown in Table 4.2.

Table 4.2: Total numbers of peaks generated from NPS, MACS or SICER and number which are specific to UD or D3 dataset.

Program Name	Total Peaks			Cell Type Specific	
	UD	D3	Merged	# UD Specific	# D3 Specific
NPS	N/A	N/A	200000	2372	7915
MACS	95528	44817	110000	77555	21598
SICER	67117	41325	74623	40903	11451

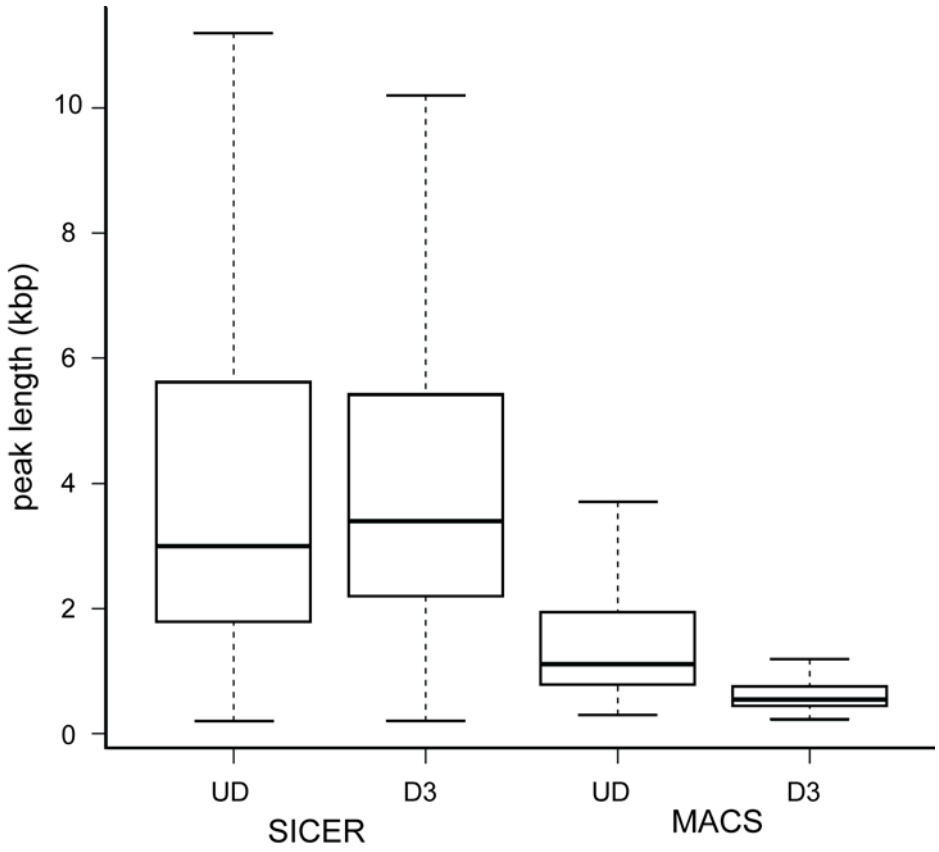


Figure 4.7*: Average width of peaks called with SICER or MACS algorithms*

Boxplot of peak width distribution (kbp) for SICER and MACS peaks in UD and D3 datasets.

After calling peaks, to try to determine whether the peaks called were relevant I took the closest TSS within 2kb for each peak, and ran GO term analysis. Knowing that H4K16ac marks the promoters of active genes, it would be expected that the closest TSS to UD H4K16ac peaks would be genes active in UD ES cells (and the same for the differentiated sample). None of the algorithms were especially skilled at picking up peaks from the D3 sample. The NPS method found only one GO term enriched for each dataset, which did not seem relevant to UD or D3 cell types. MACS and SICER performed better, both found UD specific GO terms enriched (highlighted in red – Table 4.3), but neither found any GO terms associated with RA induced differentiation in the D3 sample.

Table4.3: Functional annotation of closest TSS to UD dataset peaks or D3 dataset peaks (<2kb) using GOrilla for biological process GO terms.

Program	GO term (Biological Process)	P-value	GO term (Biological Process)	P-value
	UD		D3	
NPS	negative regulation of protein metabolic process	9.59E-04	regulation of Notch signaling pathway	2.28E-04
MACS	metabolic process	2.32E-07	aromatic compound catabolic process	4.01E-04
	primary metabolic process	1.59E-06	glutamine family amino acid metabolic process	8.20E-04
	macromolecule metabolic process	7.56E-06	regulation of T-helper 1 type immune response	8.91E-04
	cellular macromolecule metabolic process	1.18E-05		
	cellular metabolic process	1.55E-05		
	regulation of primary metabolic process	1.68E-05		
	developmental process	2.73E-05		
	negative regulation of cellular process	3.03E-05		
	reproductive process	3.70E-05		
	regulation of cellular metabolic process	3.73E-05		
	negative regulation of biological process	4.84E-05		
	regulation of nitrogen compound metabolic process	1.18E-04		
	regulation of nucleobase-containing compound metabolic process	1.23E-04		
	regulation of metabolic process	1.27E-04		
	negative regulation of cellular component organization	1.44E-04		
	gamete generation	1.68E-04		
	regulation of cellular component organization	1.70E-04		
	positive regulation of cellular component organization	1.85E-04		
	regulation of RNA biosynthetic process	1.88E-04		

Program	GO term (Biological Process)	P-value	GO term (Biological Process)	P-value
	UD		D3	
SICER	regulation of cellular metabolic process	1.86E-08		
	regulation of primary metabolic process	9.73E-08		
	regulation of macromolecule metabolic process	1.59E-07		
	regulation of metabolic process	1.79E-07		
	negative regulation of biological process	3.73E-06		
	regulation of response to stimulus	4.33E-06		
	regulation of gene expression	4.73E-06		
	regulation of nucleobase-containing compound metabolic process	4.77E-06		
	regulation of nitrogen compound metabolic process	6.22E-06		
	regulation of cellular biosynthetic process	6.33E-06		
	developmental process	7.18E-06		
	positive regulation of cellular process	7.27E-06		
	regulation of signal transduction	8.19E-06		
	regulation of biosynthetic process	8.97E-06		
	negative regulation of cellular process	9.49E-06		
	positive regulation of biological process	1.04E-05		
	negative regulation of signaling	1.27E-05		
	negative regulation of developmental process	1.50E-05		
	regulation of cellular macromolecule biosynthetic process	1.57E-05		

Knowing that H4K16ac peaks are generally wider than most other active histone modifications (Fig 3.10) - promoter peaks are frequently around 2kb wide - I analysed the width of peaks picked up by SICER and MACS. MACS peaks had an average width of around 1kb in UD, and around 500bp in the D3 sample. I was not expecting any difference between peaks in the two samples, and indeed, the peaks found in SICER do not differ in size profile. In addition, SICER peaks are generally wider than peaks found by MACS.

Finally, I visualised a selection of UD or D3 only peaks determined by the three methods (Fig 4.8A). NPS called fewer peaks than the other two methods, and peaks detected were the width of a single nucleosome. The method did not take into account the surrounding peaks. This was not useful for my analysis, and so I decided not to use this data in further work. MACS peaks were generally good for UD only H4K16ac peaks, but the very large number found indicated that regions of H4K16ac may have been split into separate peaks, or potentially that the peaks generated contain false positives. The peaks called by SICER fell in line with my observations for H4K16ac, and I therefore opted to use this algorithm in further analysis.

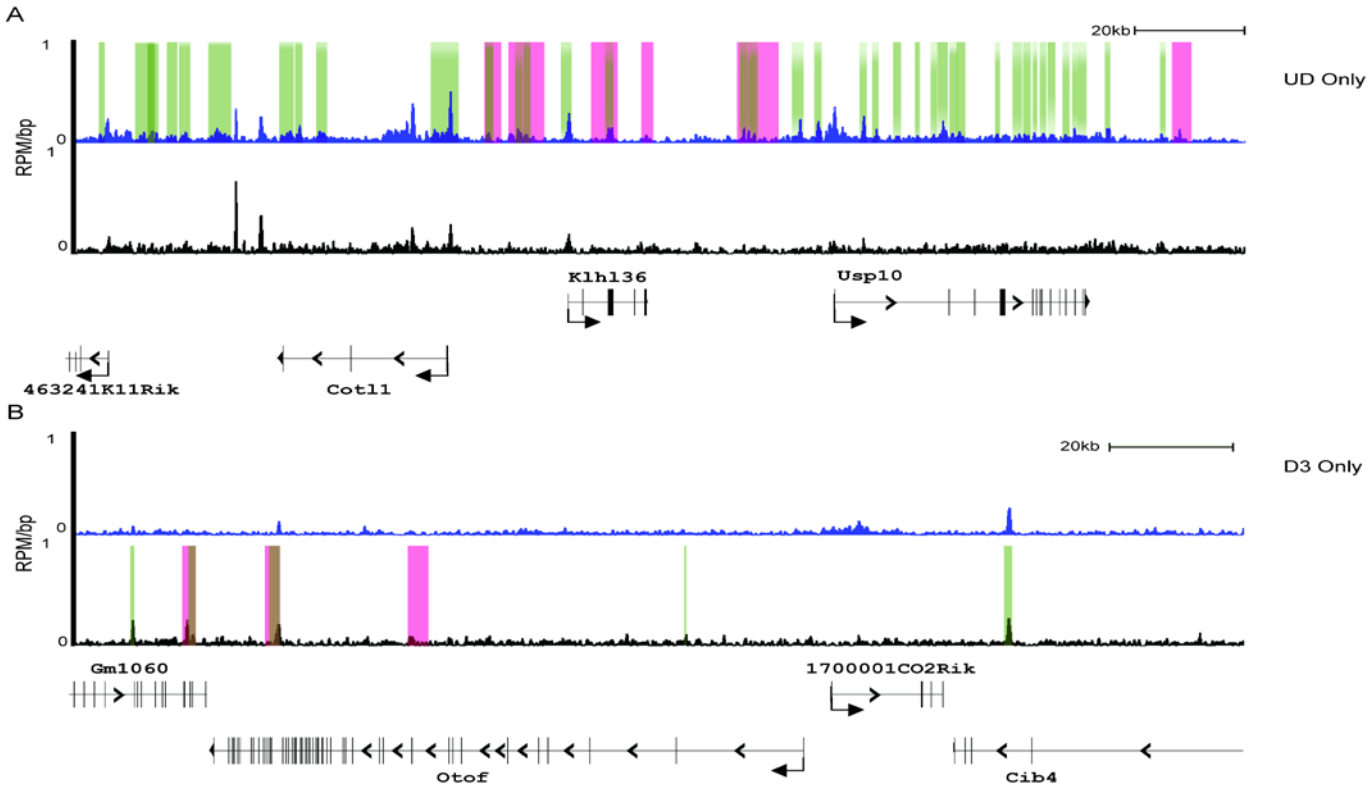


Figure 4.8*: Visualisation of peak calling results over two regions of the genome.*

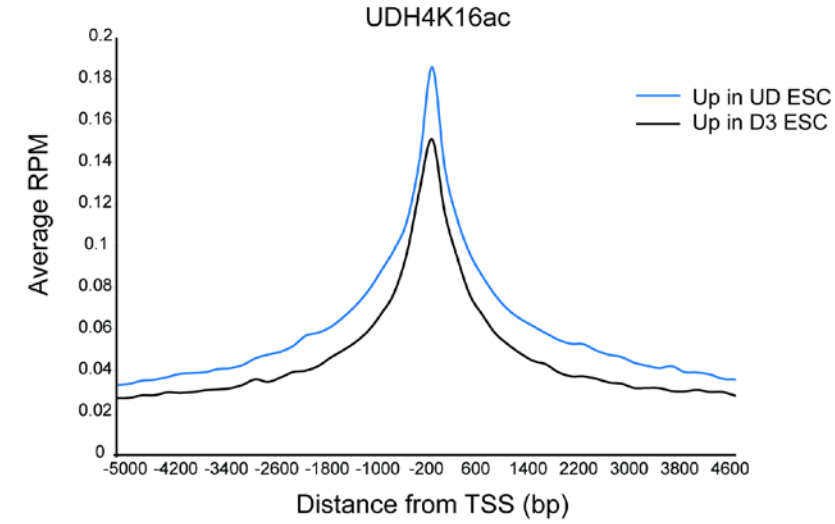
Wiggle tracks plotted against the mm9 mouse genome build using IGB programme in 200bp windows with 20bp step for UD specific (A) or D3 specific (B) peaks called by MACS (green highlight) or SICER (pink highlight). No peaks called by NPS were located in the 200kb regions selected.

4.7 H4K16ac changes with gene expression upon differentiation

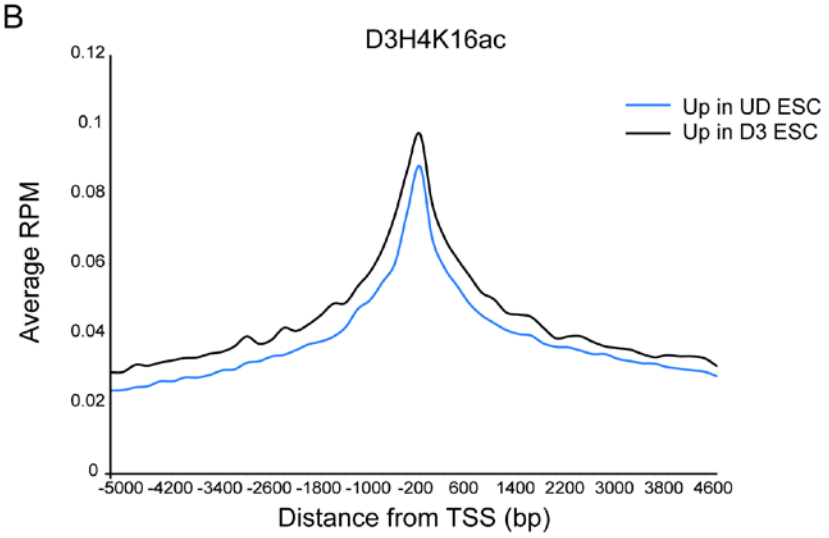
To test whether genes upregulated in the different samples also saw an increase in H4K16ac around the TSS, I took genes which were determined as upregulated in UD and D3 samples (with an adjusted p value of <0.005) and mapped the UD and D3 H4K16ac tag density around both these sets of genes. There is an increase of UD H4K16ac around the TSS of genes upregulated in UD cells (Fig 4.9A) (p-value by Wilcox test $<2.2x10^{-16}$). Conversely, there is an increase of D3 H4K16ac around the TSS of genes upregulated in D3 cells (Fig. 4.9B) (p-value by Wilcox test $<2.2x10^{-16}$). The difference remains significant if ChIP tag values are normalised to input (Log_2ChIP/input, p-value by Wilcox test <0.0002 for both gene sets). Interestingly, the increase over upregulated UD genes is approximately two fold higher on average than the increase over upregulated D3 genes, which may reflect the declining global levels of H4K16ac during differentiation. I also wanted to look at the genes which go from silent or very lowly expressed in UD ES cells to highly active in D3 cells. I divided the gene expression profile into quartiles, and took genes which fell into the lowest quartile in UD expression, and the highest quartile in D3 expression (n=34). I then mapped the tag density of UD H4K16ac and D3 H4K16ac around these genes. Although the H4K16ac peak on these genes is much higher in the D3 sample (Fig.4.9C), this difference is not significant (Wilcox test, p=0.4). Therefore the difference is likely due to a subset of genes within the set, and on other genes H4K16ac state does not correlate with expression level.

From this data I conclude that H4K16ac follows changes in expression during differentiation, and is found on active cell-type specific genes. These peaks of H4K16ac are then generally lost upon gene silencing. I did however note the retention of H4K16ac at *Pou5f1* during differentiation, which requires further study (Fig 4.3B). Retention of H4K16ac on the promoter of *Pou5f1* is noted in both the qRT-PCR and ChIP-chip datasets, but is not seen in the ChIP-seq dataset (for other pluripotency genes, loss of H4K16ac upon differentiation is consistent between ChIP-chip and ChIP-seq). This discrepancy could reflect the lower sequencing depth of the D3 ChIP-seq sample, which may result in some false negatives (Fig 4.10).

A


UDH4K16ac
Average RPM
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-5000 -4200 -3400 -2600 -1800 -1000 -200 600 1400 2200 3000 3800 4600
Distance from TSS (bp)
Up in UD ESC
Up in D3 ESC

B


D3H4K16ac
Average RPM
0.12
0.1
0.08
0.06
0.04
0.02
0
-5000 -4200 -3400 -2600 -1800 -1000 -200 600 1400 2200 3000 3800 4600
Distance from TSS (bp)
Up in UD ESC
Up in D3 ESC

C

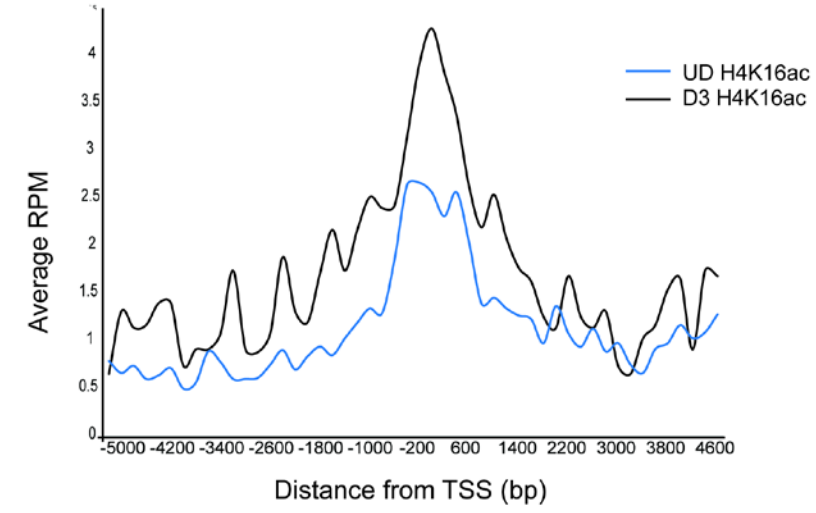

Average RPM
4
3.5
3
2.5
2
1.5
1
0.5
0
-5000 -4200 -3400 -2600 -1800 -1000 -200 600 1400 2200 3000 3800 4600
Distance from TSS (bp)
UD H4K16ac
D3 H4K16ac

Figure 4.9: H4K16ac increases around TSS with increase in gene expression in UD and D3 cells (Previous page)

A: Normalised (average RPM) tag counts of UD H4K16ac surrounding the TSS of genes significantly upregulated in UD ES cells (blue) or D3 cells (black). B: Normalised (average RPM) tag counts of D3 H4K16ac surrounding the TSS of genes significantly upregulated in UD ES cells (blue) or D3 cells (black). C: Normalised (RPM) tag counts of D3 H4K16ac (black) and UD H4K16ac (blue) surrounding TSS of genes in lowest quartile of expression for UD, and highest quartile for D3.

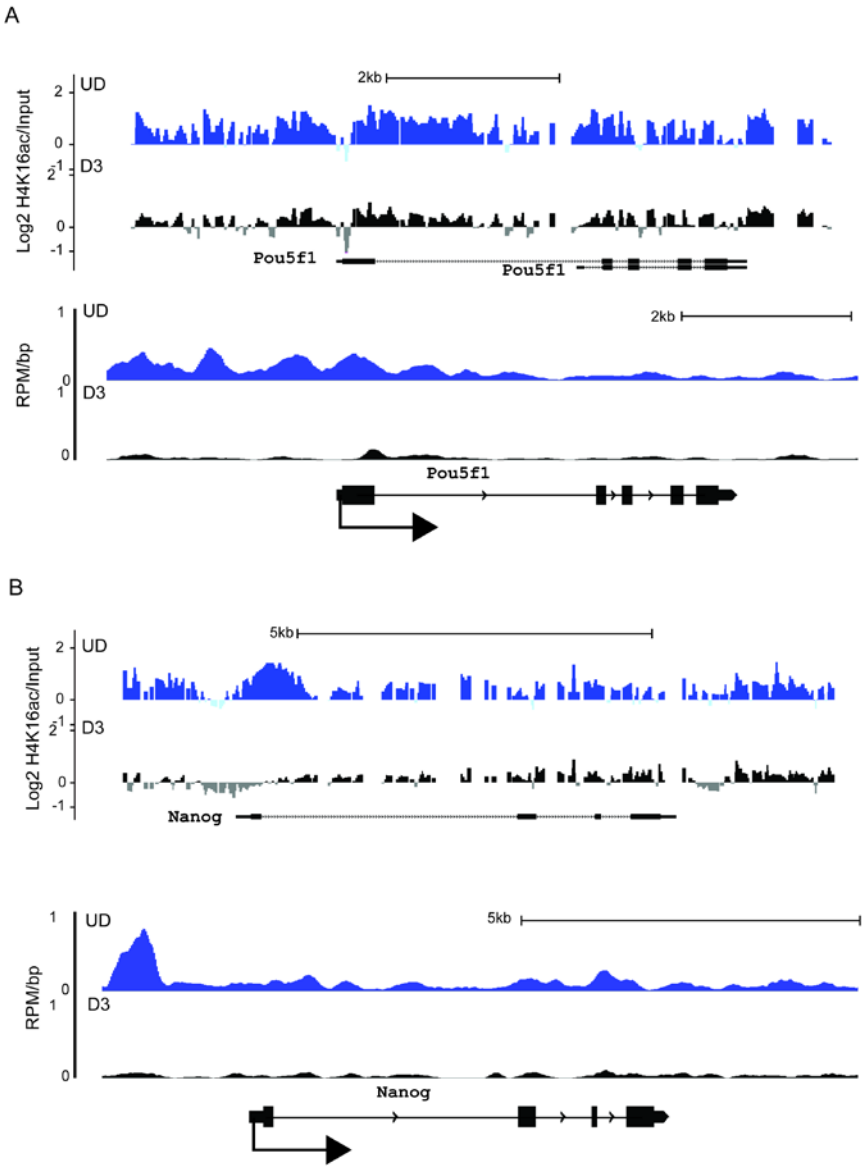


Fig 4.10: H4K16ac is retained on Pou5f1 promoter upon differentiation, but this is not detected in ChIP-seq dataset.

Log_2H4K16ac/Input in UD/D3 ES cells over *Pou5f1* (A) and *Nanog* (B) on custom Nimblegen arrays RefSeq gene annotations are from the July 2007 (mm9) Build 37 assembly of the mouse genome (http://genome.ucsc.edu) (top), and wiggle tracks of H4K16ac ChIP-seq plotted against the mm9 mouse genome build using IGB programme in 200bp windows with 20bp step.

However, having noticed that the increase of H4K16ac was lower on the D3 upregulated genes, I wanted to look at how numbers of H4K16ac peaks differed between the samples genome wide without restriction to transcription sites.

I therefore called peaks using SICER, which takes into account input background, and determined the number of UD specific, D3 specific, and overlapping peaks (Fig 4.11). There are almost four times as many UD specific peaks as there are D3 specific ones, and around 20,000 that are shared between UD and D3, and so may not be cell type specific. Some of the UD specific peaks may be due to false positive peak calls from the UD sample, or false negative peak calls from the D3 sample; always a potential difficulty in ChIP-seq analysis. UD ES cells have more specific H4K16ac, however, which would be consistent with the specific role of MOF in these cells (Li et al., 2012).

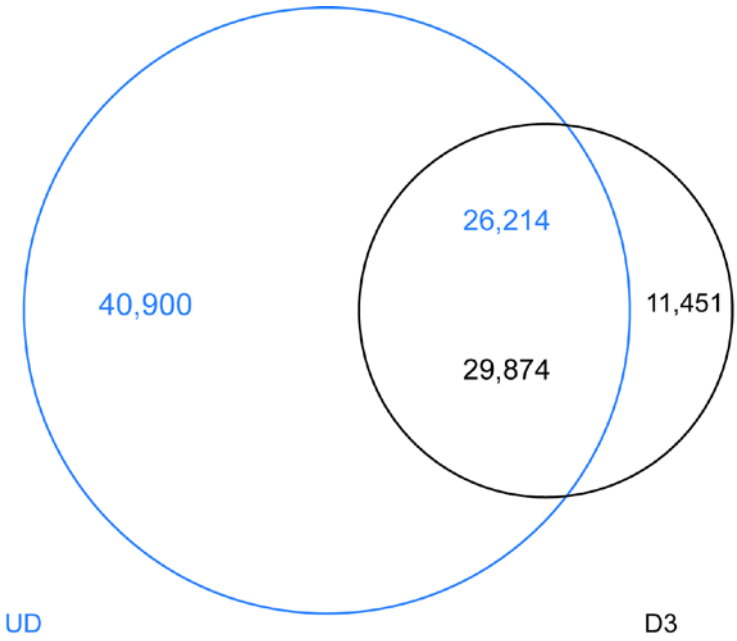


Figure 4.11*: Number of H4K16ac peaks which overlap between UD and D3.*

Venn diagram of peaks unique to UD, or to D3 ES cells. Within overlapping circle, number of UD peaks which overlap D3 peaks (blue) and number of D3 peaks which overlap UD (black).

Having shown that an increase in expression generally leads to an increase in H4K16ac, I wanted to find out whether gain or loss of an H4K16ac peak over the promoter during differentiation would predict a change in expression.

Overall, there is a large increase in expression level between genes with no H4K16ac, and genes which have an H4K16ac peak (Fig 4.12A), so presence or absence of a H4K16ac peak is a general predictor of gene expression, but I also wanted to look at how this pattern holds over differentiation induced changes in H4K16ac peaks and gene expression.

I therefore generated a list of putative promoters around mouse TSSs (2kb upstream, and 0.1kb downstream of the TSS), and determined a list of those overlapped (≥ 150bp) by a peak unique to UD, or unique to D3, then determined the log fold UD:D3 change in expression of those genes. Surprisingly, there was only a very small (though statistically significant, $p < 0.001$ by Mann-Whitney test) shift towards UD expression in the genes overlapped by a UD peak, and towards D3 expression in genes overlapped by a D3 peak. Genes overlapped by a peak in neither sample, or by a peak in both, have a log fold change value close to 0, indicating generally there is no change in expression in that set of genes (Fig 4.12B). The result was interesting, as if H4K16ac presence is determined simply by changes in chromatin environment induced by transcription, I would expect that there would be a large change shown by this experiment, since presence of an H4K16ac peak followed by loss upon differentiation would place the gene in the category shown in Figure 4.9C. This is not the case however; some genes with a peak in UD H4K16ac are lowly expressed in that cell type.

To investigate this further, I visualised the peaks over some genes which fell into the lower quartile of expression in undifferentiated cells (Fig 4.11C), to see whether the peaks were false positives (their presence is not due to sequencing bias, as that is accounted for in SICER peak calling against MNase digested input DNA). As shown in the representative examples, the peaks found fall broadly over the promoters, as on active genes, and are lost upon differentiation, though some, as in the example *Shfm1*, show a very small peak, which could be a false positive background peak. This indicates that there at least some situations where H4K16ac presence can be decoupled from activity of genes. This situation is reminiscent of that seen at polycomb regulated developmental genes where the active histone mark H3K4me3 falls in combination with H3K27me3 over the promoters of silent genes poised for expression (Bernstein et al., 2006; Zhao et al., 2007). H4K16ac is not marking this particular set of genes, however, since the *Hox* genes are polycomb regulated, and H4K16ac is notably excluded from all four *Hox* loci in undifferentiated cells (Fig 4.4).

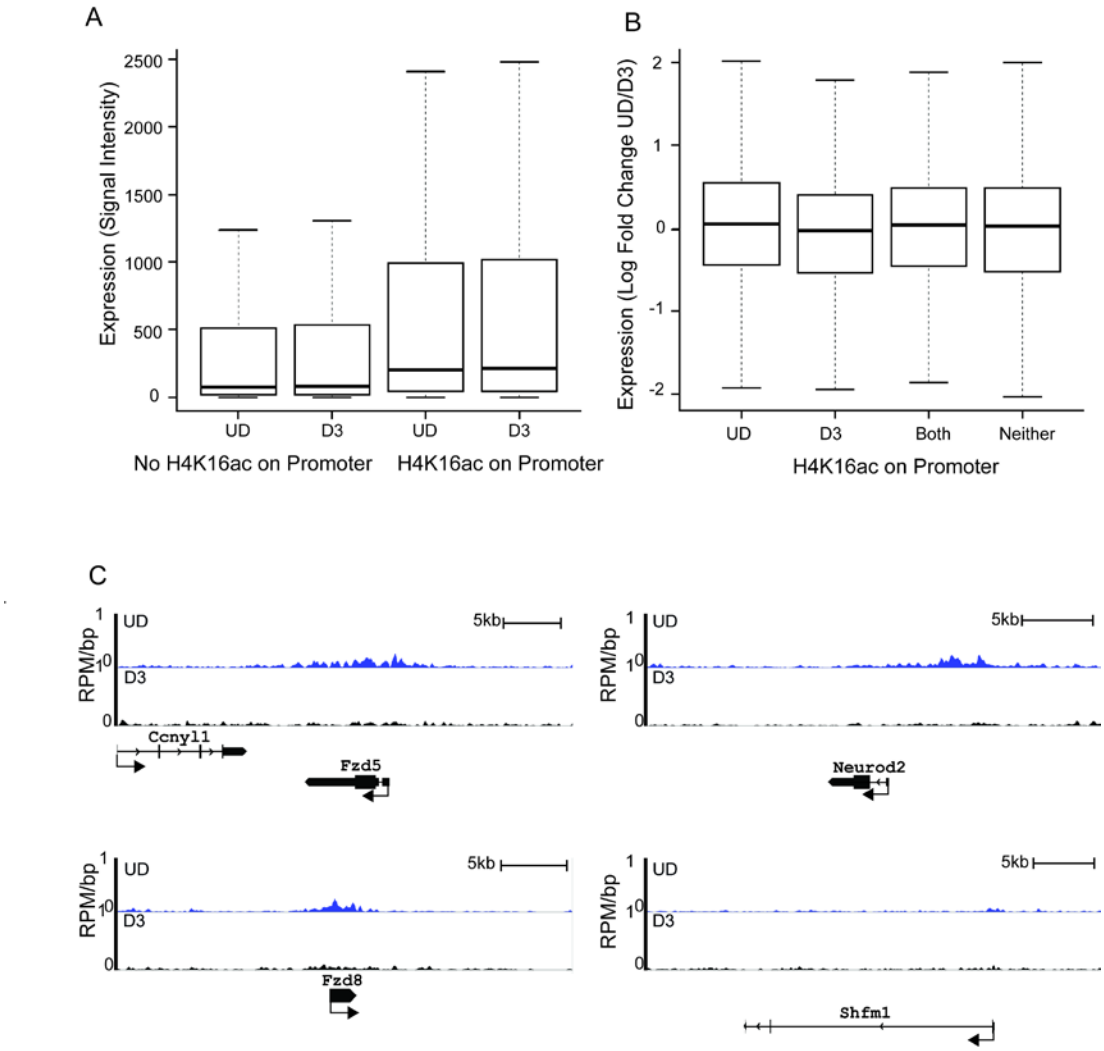


Figure 4.12*: H4K16ac does not always correlate with expression.*

A: Boxplot for distribution of expression level (mean probe signal intensity) in UD and D3 cells, for genes with a TSS overlapped by a peak of H4K16ac in both UD and D3 samples, or in neither. Outliers were excluded. B: Boxplot for distribution of log fold change UD/D3 (i.e. –ve numbers = more expression in D3 cells) for genes with H4K16ac on the promoter only in UD or D3 cells, in both, or in neither. C: Four examples of genes with H4K16ac peak defined by SICER in UD cells, which have expression in lowest quartile for UD cells. Normalised (RPM) wiggle tracks plotted against the mm9 mouse genome build using IGB programme in 200bp windows with 20bp step.

I looked at GO term enrichment of genes which have a peak of H4K16ac in UD cells only, and those which have a peak in D3 cells only (Table 4.2). In the D3 gene set, the enriched GO terms are as expected for genes which are expressed only in D3 cells. Whilst the UD set does show pluripotency associated GO terms (eg negative regulation of cell differentiation), there are also significant GO terms associated with development (which is not present when GO terms are analysed for UD expression data (Fig. 4.2). Therefore, whilst, broadly speaking, H4K16ac is associated with expression, and increases around the TSS of genes which have differentiation induced increases in expression, there are some cases where H4K16ac is present on promoters of silent (or very lowly expressed) genes. This is similar to H3K14ac, which marks a subset of inactive promoters in mouse ES cells (Karmodiya et al., 2012), and it was also noticed that MOF marks a subset of development associated proteins in undifferentiated ES cells, separate from Polycomb regulated bivalent genes (Li et al., 2012).

Table 4.4: Functional annotation of promoters with H4K16ac peak in UD sample only (top) or in D3 sample only (bottom), using GOrilla for biological process GO terms.

GO Term	Description	P-value	FDR q-value
GO:0032502	developmental process	7.09E-28	7.57E-24
GO:0048869	cellular developmental process	4.41E-23	2.36E-19
GO:0048856	anatomical structure development	1.40E-22	4.99E-19
GO:0051239	regulation of multicellular organismal process	1.78E-21	4.76E-18
GO:0030154	cell differentiation	3.07E-18	6.55E-15
GO:0023051	regulation of signaling	3.44E-18	6.12E-15
GO:0007610	behavior	4.29E-17	6.54E-14
GO:0048731	system development	7.19E-17	9.6E-14
GO:0048518	positive regulation of biological process	1.41E-16	1.68E-13
GO:2000026	regulation of multicellular organismal development	4.25E-16	4.54E-13
GO:0007275	multicellular organismal development	6.60E-16	6.40E-13
GO:0048522	positive regulation of cellular process	1.3E-14	1.16E-11
GO:0050793	regulation of developmental process	2.70E-14	2.22E-11
GO:0032879	regulation of localization	1.06E-13	8.1E-11
GO:0010646	regulation of cell communication	1.25E-13	8.89E-11
GO:0009966	regulation of signal transduction	1.66E-13	1.11E-10
GO:0048583	regulation of response to stimulus	2.87E-13	1.80E-10
GO:0051094	positive regulation of developmental process	4.06E-13	2.41E-10
GO:0044057	regulation of system process	6.42E-13	3.61E-10
GO:0045595	regulation of cell differentiation	8.97E-13	4.79E-10
GO:0007399	nervous system development	3.47E-12	1.76E-09
GO:0051093	negative regulation of developmental process	1.42E-08	0.00000271
GO:0045596	negative regulation of cell differentiation	1.76E-07	2.21E-05

GO Term	Description	P-value	FDR q-value
GO:0003002	regionalization	2.7E-10	0.00000288
GO:0007389	pattern specification process	2.8E-10	0.0000015
GO:0009952	anterior/posterior pattern specification	4.94E-10	0.00000176
GO:0048704	embryonic skeletal system morphogenesis	0.000000357	0.000952
GO:0032502	developmental process	0.000000378	0.000808
GO:0048705	skeletal system morphogenesis	0.000000776	0.00138
GO:0030154	cell differentiation	0.00000383	0.00584
GO:0007275	multicellular organismal development	0.00000604	0.00806

4.8 H4K16ac on enhancers is cell-type specific

In the previous chapter, I showed that H4K16ac marks the active enhancers of pluripotent ES cells. Here, I wanted to determine whether these active enhancers are still marked after 3 days of differentiation or whether they are specific only to pluripotent ES cells. Also, I wanted to determine whether there are any putative enhancers which become H4K16ac in the differentiated samples.

Given that there is no data for H3K4me1 or H3K27ac in D3 ES cells publicly available, I used the universal enhancer marking property of H3K4me1 to identify a set of candidate enhancers. H3K4me1 is found on both active and inactive/poised enhancers; therefore, H3K4me1 marked enhancers in ES cells may be inactive in ES cells, but activate upon differentiation along a specific lineage (Creyghton et al., 2010). Therefore, the locations of these candidate enhancers (defined using an ES cell H3K4me1 ChIP-seq dataset) may be used to determine putative enhancers (which may be active or inactive) in differentiated cells. I then used H4K16ac ChIP-seq datasets to separate this list of enhancers to those which are active (defined by presence of H4K16ac) in UD ES cells or D3 cells (Fig 4.13A). The bulk of enhancers were either only marked by H3K4me1 (and therefore active in neither cell type) or marked by H4K16ac in both cell types (totalling ~18,000 enhancers). However, a large number were specific to UD H4K16ac, whilst a much smaller proportion (only around 1000 regions) were putatively active only in D3 cells (an example is shown in Fig 4.13B). There were no significantly enriched GO terms (p-value $\leq 10^{-5}$) found from a list of the closest TSSs (< 10kb distance from enhancer midpoint) to these enhancers.

Therefore, it seems that H4K16ac on enhancers is more important for UD specific enhancers than across different cell types, although without functional data this cannot be proven. This agrees with data from previous studies which have not shown H4K16ac to be associated with enhancers in CD4+T cells.

Secondly, I looked at the enhancer marks surrounding the core pluripotency genes, *Nanog*, *Klf4*, and *Sox2*. All these pluripotency genes have a broad region of H3K4me1+/H3K27ac+/H4K16ac+ in their vicinity, indicating the presence of one or more enhancers active in ES cells. Upon differentiation, the H4K16ac on all these regions is removed (Fig 4.12C), showing its cell type specificity.

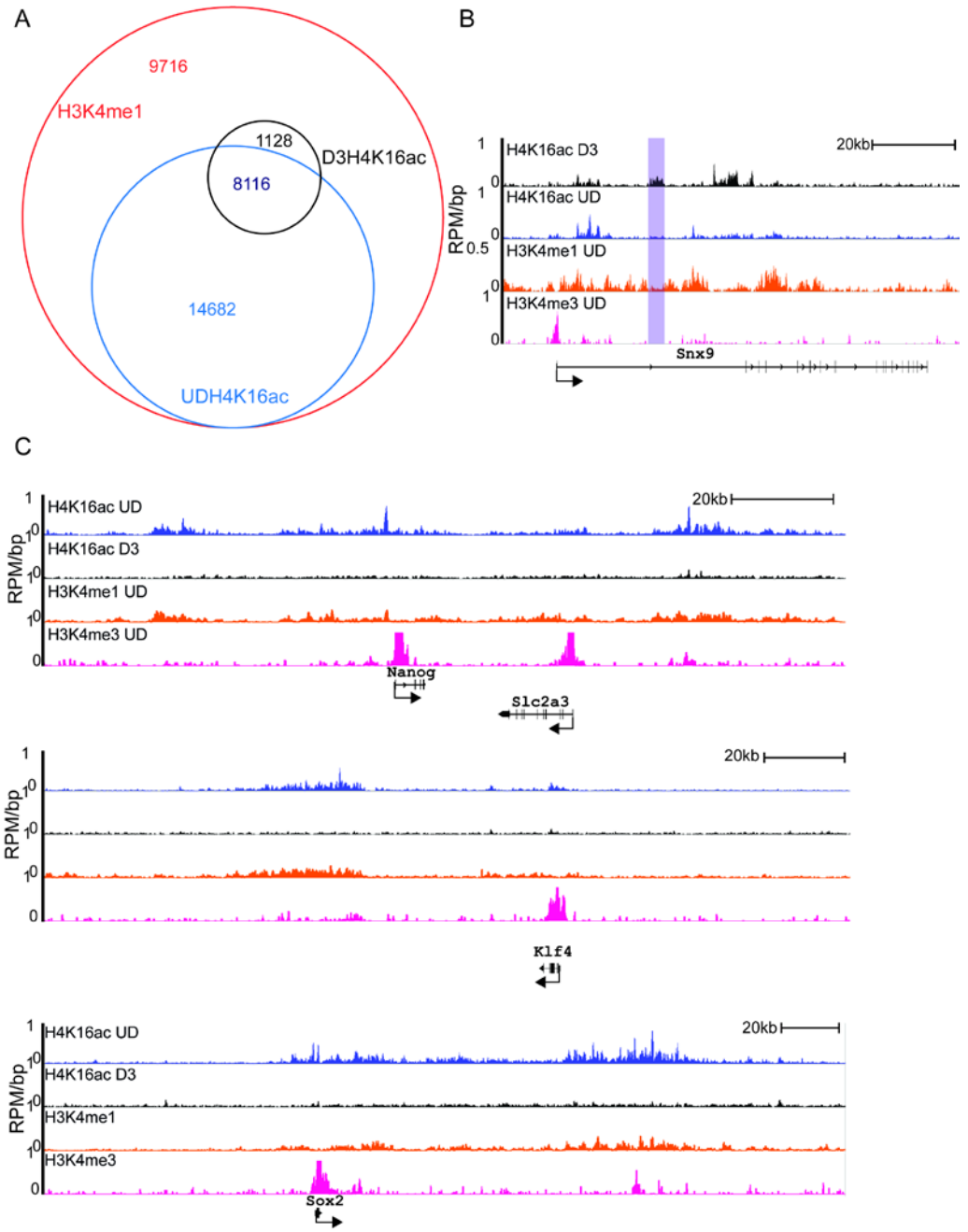


Figure 4.13: *Most H4K16ac enhancers are unique to UD ES cells*

A: Venn diagram showing enhancers (H3K4me1+/H3K4me3-) that are overlapped by UD H4K16ac compared to D3 H4K16ac. B: Normalised (RPM) wiggle tracks plotted against the mm9 mouse genome build using IGB programme in 200bp windows with 20bp step for UDH4K16ac/D3H3K16ac/UDH3K4me1/UDH3K4me3 at a putative D3 specific enhancer (location highlighted in purple).

C: As B, at regions surrounding core pluripotency genes *Nanog, Klf4* and *Sox2*.

To investigate whether other known enhancers of other cell types might carry H4K16ac, I used immortalised mesenchymal cells. These are derived from the posterior distal limb of E11.5 embryos, and which show expression of genes which correlate with their developmental origin (Williamson et al., 2012).

A "regulatory archipelago" was identified in a 600kb gene desert upstream of *Hoxd*, between *Lnp* and *Atp5g3* (Montavon et al., 2011) (Fig 4.14A), which acts on the *Hoxd* genes in limb cells. The term "regulatory archipelago" in this instance refers to a group of regulatory elements which are distributed throughout the gene desert upstream of the *Hoxd* locus, and which have an effect on the transcription of the *Hox* genes in the developing limb. I therefore used custom limb arrays (Lettice et al., 2012) and existing datasets for H3K27ac and H3K4me1 (Silvia Peluso, unpublished) to determine the epigenetic landscape on this archipelago in posterior derived limb cells.

There is a low level of H3K27ac and H3K4me1 throughout this region, indicating the five regions known to be enhancer elements in the developing limb (Montavon et al., 2011) are also candidate enhancers in the limb cells used. However H4K16ac is excluded from the region entirely. By contrast, H4K16ac is present over the ubiquitously expressed *Mtx2* and *Lnp* genes, along with the more 3' *Hoxd* genes which are active in the posterior limb.

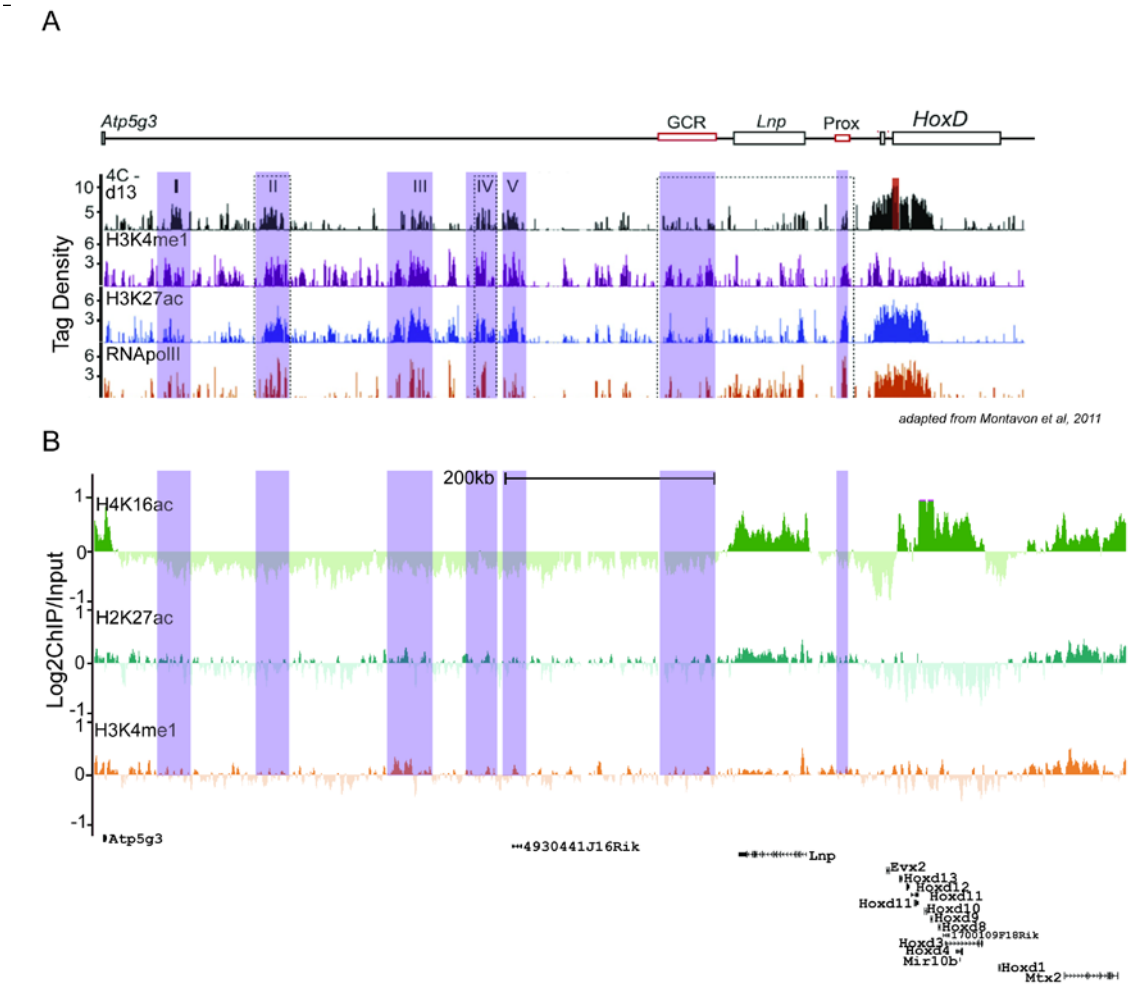


Figure 4.14: *H4K16 is not acetylated in 14fp limb cells on a regulatory region active in limb cells*

A: Interactions between DNA fragments in gene desert and *Hoxd13* (black), H3K4me1 (purple), H3K27ac (blue), and RNApolII (orange) in developing digits. Regulatory regions are highlighted in purple. Adapted from (Montavon et al., 2011).B: H4K16ac (green), H3K27ac (teal) and H3K4me1 (orange) in 14fp limb cells over *Hoxd* region and distal gene desert. Regulatory regions defined in A are highlighted in purple. H3K27ac and H3K4me1 data provided by Silvia Peluso. Log_2ChIP/Input on custom Nimblegen arrays. RefSeq gene annotations are from the July 2007 (mm9) Build 37 assembly of the mouse genome (http://genome.ucsc.edu). H4K16ac loess normalised across two biological and one technical replicate.

4.9 Summary of H4K16ac profile upon differentiation

I conclude that H4K16ac is found on transcribed genes upon differentiation, but presence of H4K16ac on the promoter of a gene is not necessarily predictive of its expression level, as some lowly expressed/silent genes show a distinct peak of H4K16ac. Acetylation of inactive genes has also been found specifically for H3K14ac (Karmodiya et al., 2012), though unlike H3K14ac (and H3K9ac from the same study), H4K16ac is not found over bivalent promoters such as the *Hox* genes in pluripotent ES cells.

Upon differentiation, the H4K16ac which is found on many enhancers is lost; by contrast very few enhancers gain H4K16ac in D3 cells. All enhancers located close to pluripotency genes lose their H4K16ac, indicating that their acetylation is cell type specific. I failed to detect any H4K16ac on a known enhancer region in immortalised limb cells. H3K4me1 and H3K27ac can be detected here, indicating that H4K16 acetylation of enhancers may be largely specific to UD ES cells.

Chapter 5: Loss of H4K16ac does not predict changes in chromatin compaction

5.1 Introduction

Histone lysine hyperacetylation neutralises the positive charge carried by an unmodified lysine, and generally correlates with euchromatin rather than compact heterochromatin (Shahbazian and Grunstein 2007). Within specific euchromatic loci, histone acetylation colocalises with DNAseI hypersensitive sites (Hebbes et al. 1994), showing its potential ability to determine chromatin structure. Additionally, nucleosomes with acetylated histones have decreased strength of interaction with the DNA (Hong et al. 1993), and increased accessibility of DNA to transcription factors and transcriptional machinery (Vettese-Dadey et al. 1996).

Amongst the numerous potentially acetylated lysine residues, H4K16ac is specifically important for the process of chromatin decompaction. *In vitro*, addition of H4K16ac must be removed prior to decompaction (Robinson et al. 2008). Addition of H4K16ac to nucleosome arrays causes them to decompact from a 30nm fibre like structure and inhibits their ability for self-association (an *in vitro* approximation of chromatin fibre interactions) (Shogren-Knaak et al. 2006; Allahverdi et al. 2010). The interaction between H4K16 and H2A/H2B on the neighbouring nucleosome (Luger et al. 1997; Liu et al. 2011) is important for nucleosome-nucleosome interaction; acetylation of H4K16 abrogates this (whilst mutation of the lysine to an uncharged glutamine residue does not), demonstrating a direct effect of H4K16ac on nucleosome folding.

I used the system of Retinoic Acid (RA) differentiation and the associated changes in H4K16ac to assay whether a change in H4K16ac state alone was a predicting factor for changes in chromatin compaction state *in vivo*. I tested this using 2D Fluorescence In-Situ Hybridisation (FISH), which measures changes in interprobe distance between two points in a region, which can be inferred to be a result of changes in chromatin compaction state (Chambeyron and Bickmore 2004).

5.2 Detecting Changes in Chromatin Compaction using 2D FISH

2D FISH gives a measure of chromatin compaction over specific genomic regions. For example, between different regions in the same cell type (Fig 5.1A)(Yokota et al. 1997), the same regions between wild type and mutant cells (Eskeland et al. 2010). Most relevant for this study, is measurement of changes during differentiation (Chambeyron and Bickmore 2004).

To measure chromatin compaction changes by FISH, fosmid pairs flanking the region of interest are selected using Ensembl (Ensembl61, NCBIM37, July 2007 genome assembly). The probes are labelled with DIG-11-dUTP or biotin-16-dUTP and hybridized to MAA fixed ES cells or D3 cells, then detected with fluorescein or Texas Red-conjugated antibodies to DIG and biotin respectively. The physical distance between the two probes in the different cell types can then be measured. Chromatin behaves according to a random walk model, meaning there is a linear relationship between the interphase interprobe distance^2 and the genomic distance (Fig 5.1B) at distances from 100kbp to 1.5Mbp. Therefore larger values at the same genomic distance suggest a less compacted chromatin state. At greater genomic distances (1.5Mbp-200Mbp), chromatin behaviour deviates from the random walk model; physical interprobe distance does not increase at the same rate (as a function of genomic distance) as at smaller genomic distances. This is explained with a “giant loop” model, of large loops on a flexible backbone (Sachs et al. 1995).

Therefore, at genomic distances lower than 1Mbp, it is possible to use 2D FISH to measure compaction. For example the gene enriched Giemsa-negative bands of mitotic chromatin show greater interprobe distances/genomic distance than Giemsa-positive (less enriched in genes) (Yokota et al. 1997) (Fig. 5.1A).

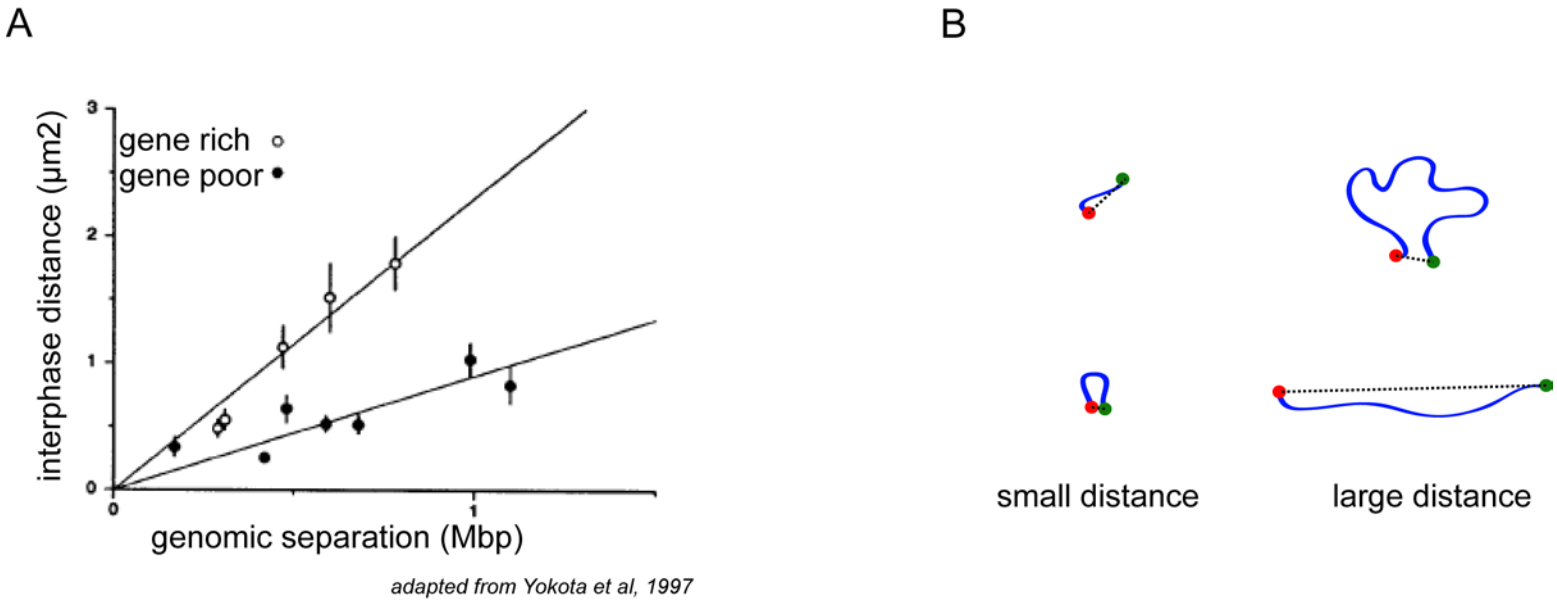


Figure 5.1*: Mean physical interprobe distance^2 is proportional to genomic interprobe distance*

A: Relationship between interprobe distance and genomic distance in gene rich – band Xq28 (open circles) and gene poor – band Xp21.3 (black circles) regions. Adapted from (Yokota et al. 1997). B: Illustation of how mean interprobe distance^2 is proportional to genomic distance. Since small interprobe distances occur over regions with large genomic separation (due to random walk model behaviour of the chromatin fibre), raw interprobe distance mean is reduced. Squaring the dataset means that large interprobe distances which occur only across regions with large genomic distances are high enough to increase the mean. Across short genomic distances, squaring the interprobe distance has little effect on the mean.

5.3 Loss of H4K16ac does not predict compaction of chromatin

To analyse the relationship between H4K16ac and chromatin compaction in ES cells, I first looked for regions of the genome which showed large (>100kb) domains of H4K16ac in UD ES cells which were subsequently lost upon differentiation. I therefore centred my search around genes associated with pluripotency, starting with the core pluripotency factors, *Pou5f1*, *Nanog*, and *Sox2*. I visualised the H4K16ac profile (along with MNase digested input DNA, to exclude regions with high background) in the Integrated Genome Browser (IGB) (Fig 5.2). All the loci encoding core pluripotency factors had large domains of H4K16ac in their vicinity, which, as expected, were lost upon differentiation. There was little background, evidenced by low amounts of input DNA in both samples. I therefore selected *Sox2* and *Nanog* for further investigation, and designed probes surrounding the H4K16ac domains at these loci for 2D FISH (Table 5.1).

For initial analysis, I looked at raw interprobe distance squared (d^2), and for both *Nanog* and *Sox2*, along with the control α-globin locus, and found no significant difference between UD and D3 cells ($p>0.05$, Wilcox test) (Fig 5.3). Since nuclear area can alter with cell type, and can also be affected by humidity on the day of slide preparation, it is important to normalise to area (Fig 5.4), but this does not reveal any significant change in chromatin compaction ($p>0.05$ for all probe pairs, Wilcox test).

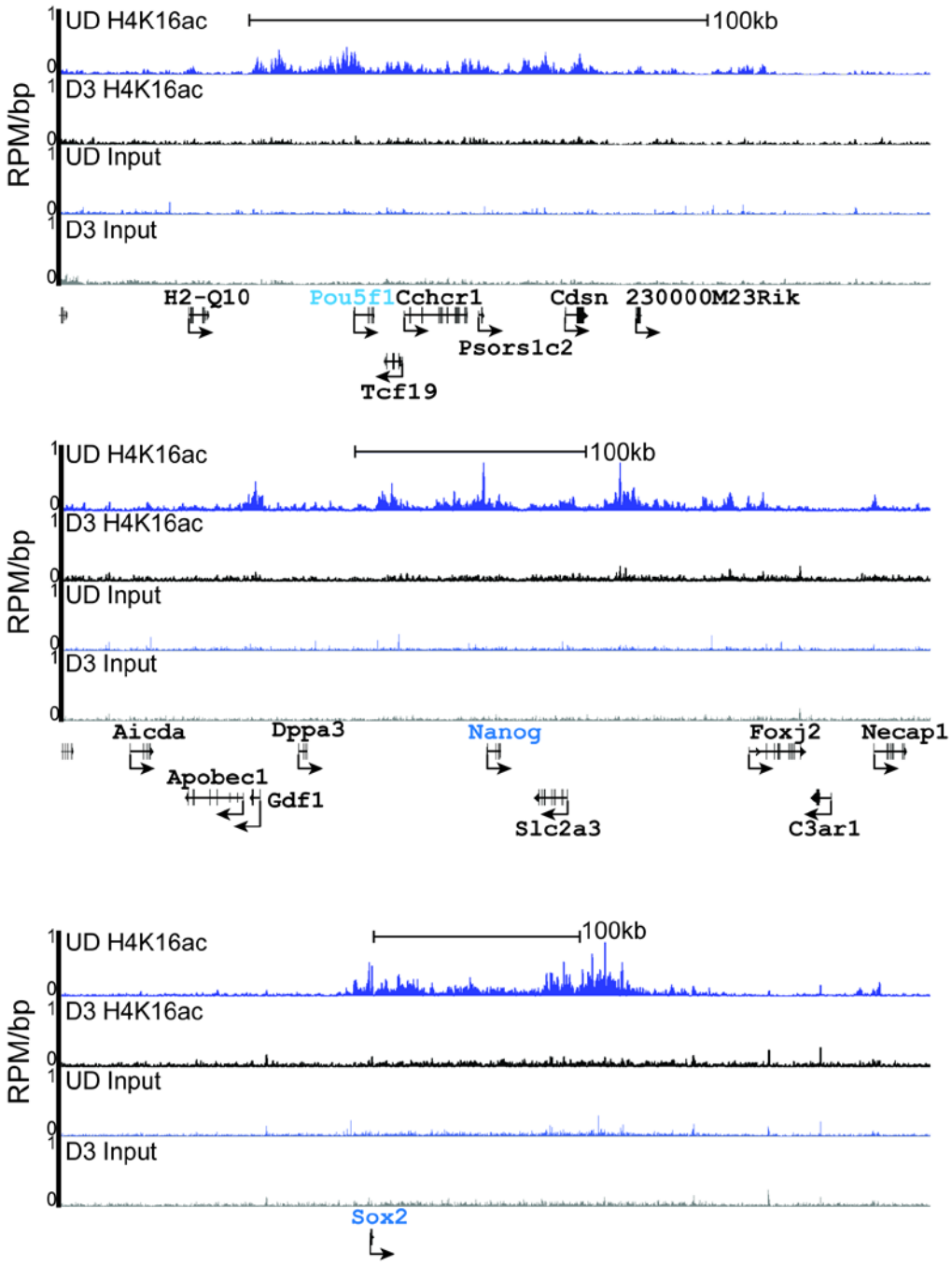


Figure 5.2: *Large domains of H4K16ac in UD ES cells*

Three locations in UD ES cells with a large (>100kb) domain of H4K16ac lost upon differentiation. Normalised (RPM) wiggle tracks plotted against the mm9 mouse genome build using IGB programme in 200bp windows with 20bp step.

Table 5.1: Statistics of d^2 and d^2/area for all datasets from 2D FISH.

Region	Probe 1	Probe 2	Separation (bp)	Cell Type	Mean d^2	Median d^2	Mean d	Median d	Median d/Mean d	St.Dev d	St.Dev. d/Mean d	p-value Wilcox test d, UD to D3
Nanog	WI1-2049I9	WI1-196C18	305863	UD	0.521	0.334	0.629	0.578	0.918	0.356	0.566	0.34
				D3	0.549	0.289	0.610	0.537	0.881	0.423	0.693	
Sox2	WI1-1766B8	WI1-1185J11	142984	UD	0.222	0.109	0.394	0.330	0.837	0.260	0.660	0.06
				D3	0.155	0.076	0.331	0.275	0.832	0.215	0.650	
α-globin	WI1-2837A17	WI1-2903N21	95495	UD	0.325	0.085	0.420	0.291	0.693	0.387	0.923	0.06
				D3	0.236	0.049	0.347	0.222	0.640	0.342	0.987	

Region	Probe 1	Probe 2	Separation (bp)	Cell Type	Mean d^2/r^2	Median d^2/r^2	p-value Wilcox test d^2/r^2, UD to D3
				UD	0.004	0.003	
Nanog	WI1-2049I9	WI1-196C18	305863	D3	0.006	0.003	0.91
				UD	0.002	0.001	
Sox2	WI1-1766B8	WI1-1185J11	142984	D3	0.002	0.001	0.35
				UD	0.003	0.001	
β-globin	WI1-2837A17	WI1-2903N21	95495	D3	0.002	0.001	0.12

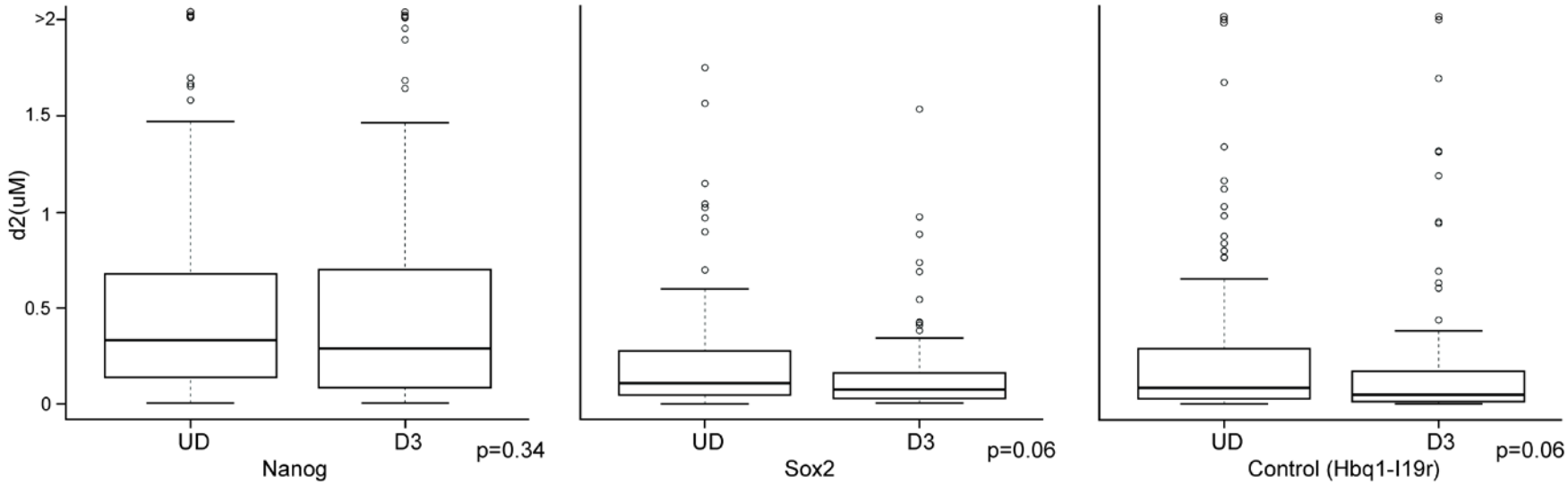


Figure 5.3 *Loss of H4K16 acetylation does not correlate to decrease in interprobe distance*

Boxplots indicating the distribution of unnormalised squared interprobe distances (d^2) for UD and D3 cells. Boxes show the median and interquartile range of the data; circles indicate outliers. n=50 nuclei. Statistical significance of differences were examined by Mann-Whitney U test in R version 2.14.0.

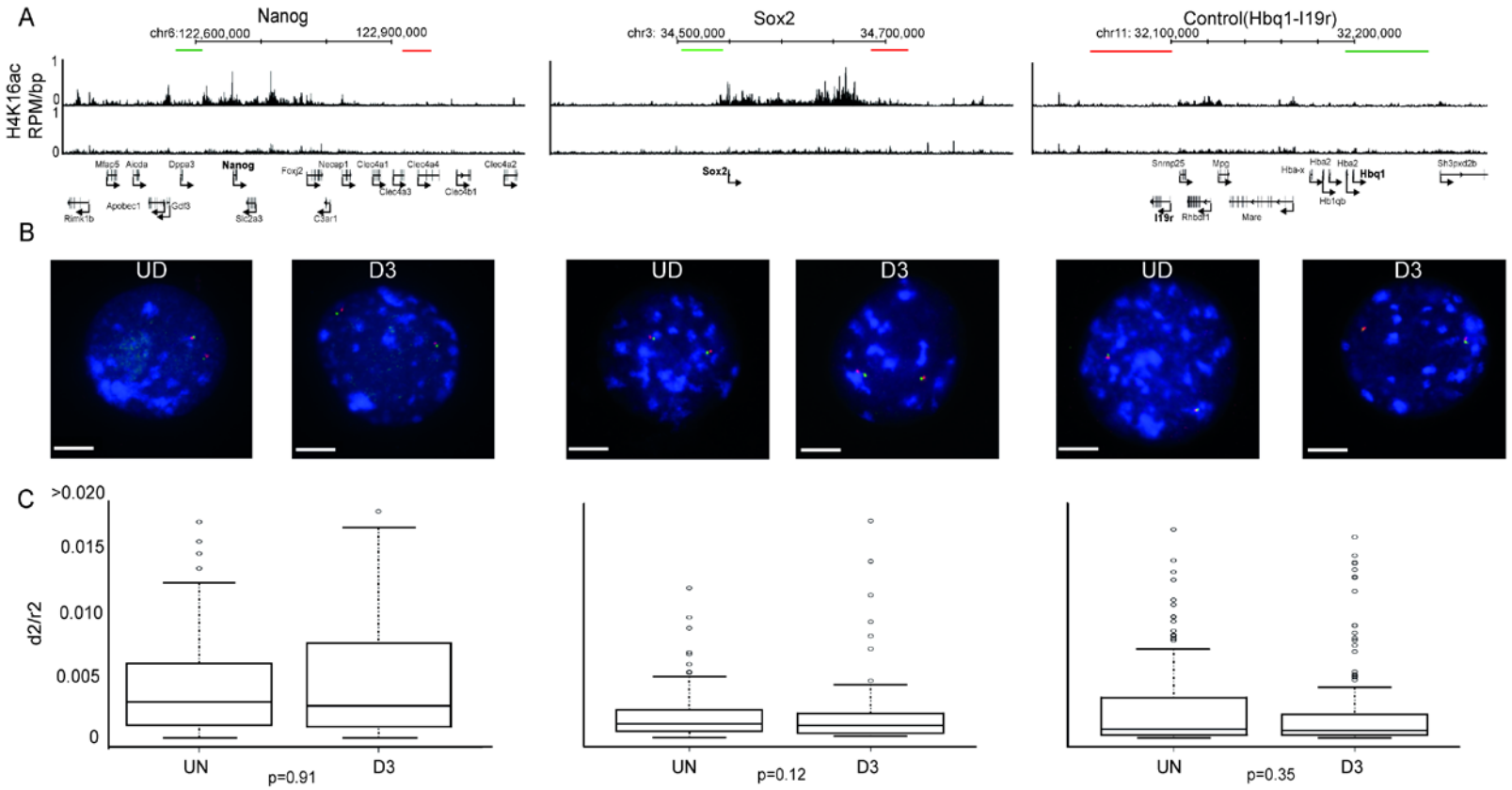

A
Nanog
chr6:122,600,000
122,900,000
Sox2
chr3: 34,500,000
34,700,000
Control(Hbq1-I19r)
chr11: 32,100,000
32,200,000
H4K16ac
RPM/bp
Nanog
Sox2
Hbq1
I19r
B
UD
D3
UD
D3
UD
D3
C
>0.020
0.015
0.010
0.005
0
d2/r2
UN
p=0.91
D3
UN
p=0.12
D3
UN
p=0.35
D3

Figure 5.4*: Loss of H4K16 acetylation does not correlate to chromatin compaction in vivo*. (Previous page)

A: H4K16ac (RPM/bp in 200bp sliding window with 20bp step) across the *Nanog, Sox2* loci and control (Hbq-Il9r) loci in undifferentiated ES cells (UN, top row) and in differentiated cells (D3, bottom row). The position of fosmid probes (green and red boxes) used in FISH is indicated below genomic maps from the mm9 assembly of the mouse genome. B: Example FISH images of nuclei from undifferentiated (UN; left) and differentiated (D3; left) ES cells, hybridized with probe pairs cross the *Nanog, Sox2* and Hbq-Il9r loci. Nuclei were counterstained with DAPI (blue). Scale bar = 20µm. C: Boxplots indicating the distribution of squared interprobe distances (d^2) normalised to nuclear radius^2 (r^2) for UD and D3 cells. Boxes show the median and interquartile range of the data; circles indicate outliers. n=50 nuclei. Statistical significance of differences were examined by Mann-Whitney U test in R version 2.14.0.

5.4 Summary of association with chromatin compaction

Loss of a large domain of H4K16ac during ES cell differentiation does not lead to a change in chromatin compaction over that domain at a level which can be detected by FISH. These experiments do not rule out the possibility that chromatin decompaction occurs on a nucleosomal level which cannot be detected by FISH, or that removal of H4K16ac may be necessary for a subsequent decompaction by another mechanism

Chapter 6: Preliminary investigation into the MOF complex in ES cells

6.1 Introduction

The interaction partners of MOF are extremely important for its activity. Although on free histones, MOF alone displays the ability to acetylate H4K16, on nucleosome arrays it can only show HAT activity when its interaction partners are also present. Even on free histones, the interaction partners increase HAT activity and are required for specificity (Morales et al., 2004).

In Drosophila, MOF is vital for the process of dosage compensation in male flies, which carry only one copy of the X chromosome to the females' two, but which require equal dosage of many of the genes contain therein. However, MOF is also expressed in female flies, implying a division of function. In males, MOF localisation to the X chromosome is mediated by binding to MSL1 in association with MSL3; this binding also increases HAT activity. MOF is also found at the promoters of autosomal genes in both males and females. This binding is independent of the MSL complex and instead occurs with MOF in the context of the NSL, or Non-Specific Lethal complex (Raja et al., 2010; Feller et al., 2011).

In mammalian cells, MOF has also been documented to be a member of several complexes, including the mammalian homolog of the MSL complex (Smith et al., 2005), potentially the MLL complex (Dou et al., 2005), and the mammalian homolog of the NSL complex. As in Drosophila, the complexes are functionally distinct; within the MSL complex, MOF has specific substrate specificity to H4K16ac, whilst when part of the NSL complex, MOF is reported to exhibit acetylation activity on H4K5 and H4K8 as well as H4K16 (Cai et al., 2009).

As shown in previous chapters, MOF is of particular importance in the context of undifferentiated embryonic stem cells (UD ES cells). I therefore wanted to look at the binding profile of MOF (recently published by Li and colleagues, 2012) and its interaction partners in ES cells and by contrast, in Mouse Embryonic Fibroblasts (MEFs). In addition, I wanted to examine the effect on UD ES cells, and upon differentiation, of changes in the level of MOF protein using overexpression.

6.2 MOF over-expression in ES cells

As in previous studies, I first opted to look at MOF complexes by adding a flag tag to enable a simple pull down unaffected by poor antibody specificity. I first tried an approach which had previously been successfully used for the over-expression of Ring1B (Eskeland et al., 2010) in ES cells, utilising the pTLC plasmid (Fig 6.1A). I selected for ES colonies stably containing the construct, which results in expression of dsRed fluorescent protein and the gene for puromycin resistance. To determine whether the population was pure, I used FACS to check levels of dsRed and subsequently introduced Cre, under control of a CAG promoter. pTLC contains two LoxP sites; excision at these sites should result in loss of dsRed and puromycin resistance, and gain of expression for Flag-MOF and GFP. GFP (and therefore Flag-MOF) expressing cells can then be FACS sorted prior to use in downstream experiments.

The initial transfection and stable cell line generation was successful, all selected colonies expressed dsRed as detected by FACS to a similar extent (an example is shown in Fig. 6.1B). However, after addition of Cre, only a very small proportion of the population was GFP positive. Around 25% remained dsRed positive, but surprisingly, around 70% showed loss of dsRed, without an associated gain of GFP. The low number of GFP positive cells was sorted, but a subsequent analysis 96 hours later showed that the majority of these had lost expression of GFP (the population going from 100% GFP positive to 6.7 % GFP positive) (Fig. 6.1C). This suggests that MOF over-expression is detrimental to ES cells and made the construct impractical for experiments which required a large number of cells.

As a different approach for MOF expression was needed, I then put flag-MOF into a construct allowing for direct expression in transfected cells (Fig. 6.2). I first tested the construct for expression in the easily transfected Human Embryonic Kidney (HEK) 293T cell line. This showed strong expression of flag-MOF, but not in control transfection with the empty vector, or in untransfected controls. By contrast, the transient expression of the same construct in ES cells showed that no Flag could be detected by western blot. This confirms that high levels of MOF are selected against in ES cells.

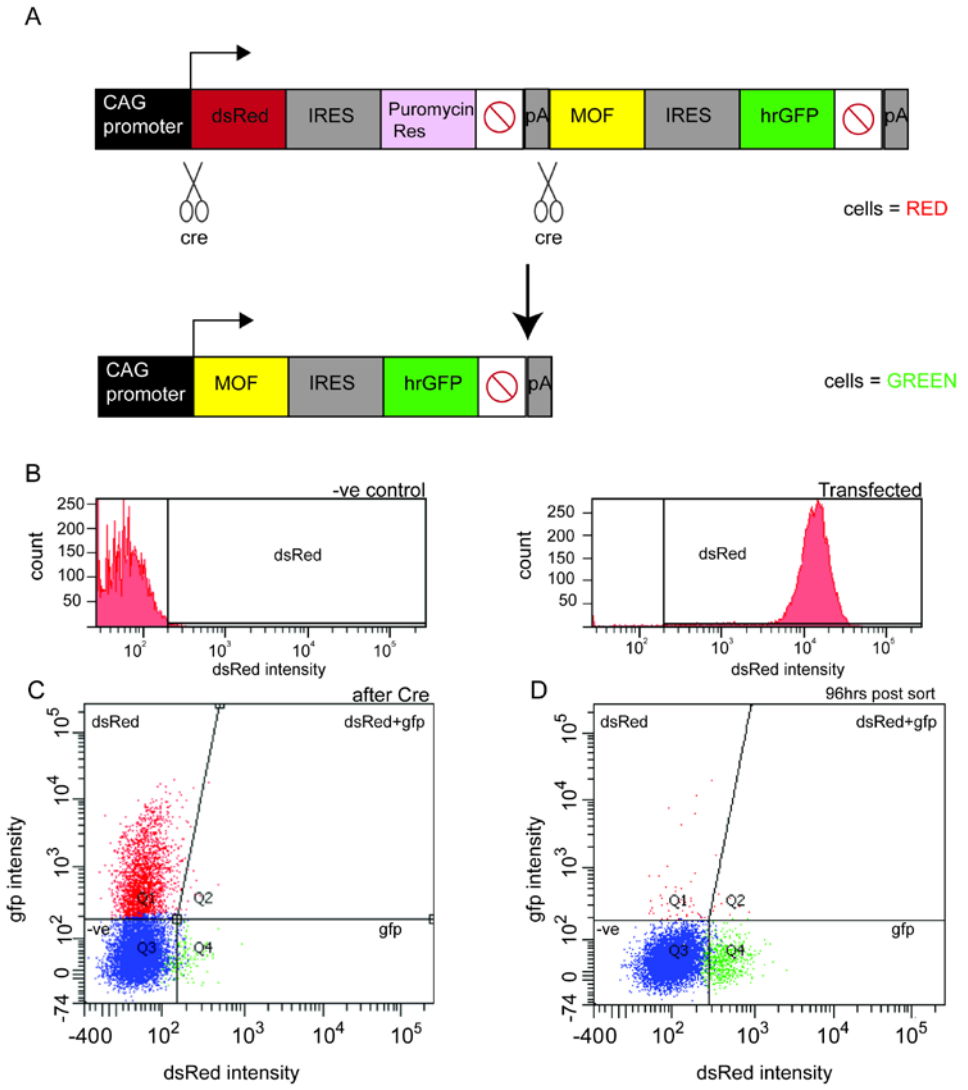


Figure 6.1 *Overexpression of Flag-MOF in ES cells using pTLC strategy*

A: Illustration of system for Flag-MOF overexpression. Schematic of pTLC plasmid as initially inserted. LoxP sites are represented by scissors. B: FACS analysis for untransfected control (*left*), and stable, puromycin resistant colonies (*right*). Cells exposed to a 561nm laser for excitation of RFP. x-axis shows fluorescence intensity. C: Cre-loxed stable puromycin resistant cells, 5 days after Cre transfection, split harshly and allowed to form colonies. Cells exposed to a 561nm laser and a 525nm laser for excitation of RFP and GFP respectively. Results then plotted according to intensity; x-axis shows fluorescence intensity of GFP, y-axis shows fluorescence intensity for RFP. D: As C, for Cre-loxed GFP sorted cells, 96 hours post-sort for GFP.

Figure 6.2 *Overexpression of Flag-MOF in ES cells using Zeocin resistance strategy*

A: Schematic of pCAGxn plasmid. B: Immunoblot for anti-FLAG, of cells transiently transfected (72 hours post-transfection) with flag-p52 (flag western blot positive control), flag-MOF (fl-MOF), empty vector (e.v.) or non-transfected control. Position of molecular weight markers is indicated on the right (kD). Cell type is indicated above. C: As B, for zeocin resistant stable ES cell lines, and transiently transfected flag-p52 in 293T cells (flag western blot positive control).

6.3 Preliminary data for interaction partners of endogenous MOF

Given the difficulties involved in overexpressing Flag-MOF, I decided to focus instead on isolating the complex of MOF using the endogenous protein. Work by colleagues in the lab showed that after running nuclear extract of ES cells on a gel filtration column, MOF separates into a fraction corresponding to approximately 700kDa in size; interestingly, MLL is also found in this fraction (Ragnhild Eskeland, unpublished, Fig 6.3A).

The interaction between MOF and MLL has been shown previously (Dou et al., 2005), but was not seen in subsequent work (Mendjan et al., 2006; Cai et al., 2009). In addition, all previous work on mammalian MOF interactions has been undertaken in human cells (293T and HeLa) using overexpression. To investigate this further, using an unbiased approach which did not rely on a large number of high-quality antibodies, I used the SILAC technique (Fig 6.3B). Briefly, cells are labelled with either heavy or medium lysine/arginine; essential amino acids which the cell cannot produce. They are then lysed and used in a traditional immunoprecipitation. Pull down from IgG and proteins of interest are then pooled and washed, boiled, and separated by SDS-PAGE. Ten gel slices per fraction were cut and digested in-gel with trypsin. The purified peptides were then separated (Ultimate U3000, trap-enriched nanoflow LC-system, Dionex), and identified (LTQ Orbitrap XL, Thermoscientific, via nano ES ion source, Proxeon Biosystems). Quantification (MaxQuant, based on 2D centroid of isotope clusters within each SILAC pair) can distinguish between the samples, to give a ratio of protein of interest to IgG. Background proteins would be expected to have a ratio of 1:1 and can therefore be disregarded. SDS-PAGE, LC-MS/MS and quantification were carried out by Dundee Cell Proteomics.

I performed these experiments on MEFs and ES cells (Table 6.1). In MEFs, the result falls in line with data previously seen in the literature, as MOF is found in a complex with the MSL complex homologues. In ES cells, however the MSL proteins were not found. A considerable amount of further work is needed to verify this result; for example antibodies for the MSL proteins must be acquired and optimised, and reciprocal IPs performed in both cell types. Given the important role of MOF in ES cells it would be interesting if complexes differ between ES cells and other cell types. I did not find MOF in complex with MLL in either cell type.

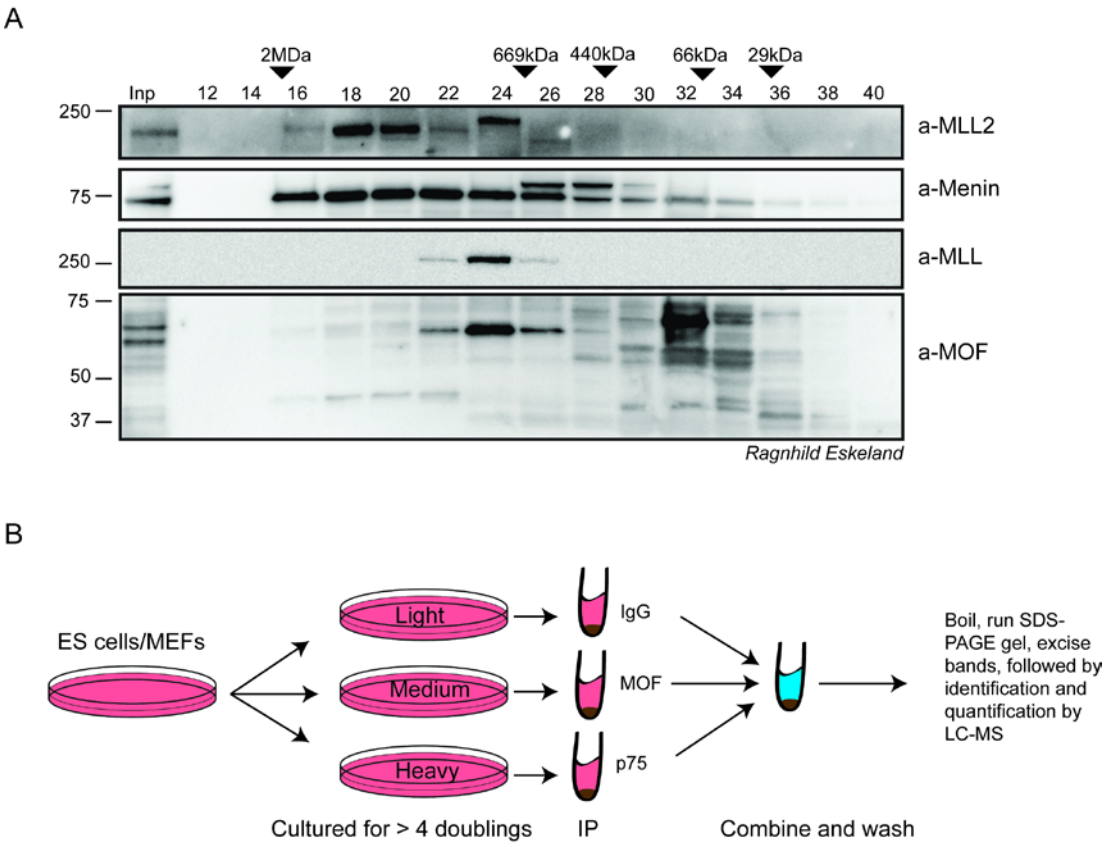


Figure 6.3 *Determining interaction partners for MOF in ES cells*

A: Immunoblot of a SDS-polyacrylamide gel for aliquots of fractions (input = I and numbered column fractions 12-40) derived from ES cell nuclear extract on a Superose 6 gel-filtration column (Ragnhild Eskeland, unpublished). Arrows indicate approximate complex sizes. Numbers above blot indicate fraction numbers (0.5mL) from start to finish. B: Work flow for labelling cells with heavy isotope containing amino acids. Cells were seeded into DMEM containing “light” (R0K0), “medium” (R6K4), or “heavy” (R10K8) amino acids (Dundee Cell Products). After at least four cell doublings, the cells are collected, lysed, and used in conventional IP for IgG, MOF, or p75. The beads are then pooled and washed, and are then processed for LC-MS/MS by Dundee Cell Proteomics. Identical pairs of peptides with different isotope compositions can be distinguished by the mass spectrometer based on mass differences, and the relative ratio of abundance can be calculated.

Table 6.1 Results of SILAC experiment in MEF and ES cells.

MEF				
Name	**MW (kD)**	**Description**	**R6K4/R0K0 - MOF/IgG**	**Peptides**
Clint1	*67.73*	clathrin interactor 1	***52.001***	*17*
Msl1	67.319	Isoform 1 of Male-specific lethal 1 homolog	**39.618**	15
Cltc	*191.98*	*Clathrin, heavy polypeptide*	***27.552***	93
Ppp1r9b	89.519	Isoform 1 of Neurabin-2	**26.828**	25
Ap2a1	107.66	Isoform A of AP-2 complex subunit alpha-1	**22.1**	34
Kat8	**52.573**	**Isoform 1 of Probable histone acetyltransferase MYST1**	**21.532**	8
Ap2b1	105.72	Isoform 2 of AP-2 complex subunit beta	**19.69**	34
Msl3	60.291	Isoform 1 of Male-specific lethal 3 homolog	**16.863**	2
Ap2a2	104.02	adaptor protein complex AP-2, alpha 2 subunit	**11.893**	25
Msl2	62.537	Male-specific lethal 2 homolog	**10.357**	4
Bmp2k	126.18	Isoform 1 of BMP-2-inducible protein kinase	**10.139**	3
Reps1	86.518	Isoform 1 of RalBP1-associated Eps domain-containing protein 1	**8.7978**	4
Actn4	*104.98*	*Alpha-actinin-4*	***8.7417***	*67*
Ap2m1	49.654	AP-2 complex subunit mu	**8.2163**	5
Spna2	*285.34*	*spectrin alpha 2*	***6.9795***	*66*
C330027C09Rik	102.11	Protein CIP2A	**6.3761**	2
Ppp1ca	37.54	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	**6.3231**	9
Spnb2	*274.22*	*Isoform 1 of Spectrin beta chain, brain 1*	***6.2025***	*42*
Pml	98.241	Isoform 1 of Probable transcription factor PML	**5.6959**	5
Col5a2	145.02	Collagen alpha-2(V) chain	**5.3838**	3
Col5a1	183.67	Collagen alpha-1(V) chain	**5.3265**	9

Lima1	84.059	LIM domain and actin binding 1 isoform a	**5.1746**	43
Cenpj	153.05	Centromere protein J	**4.39**	2
Actn1	*103.07*	*Alpha-actinin-1*	***4.3391***	*61*
Ppp1cb	37.186	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	**3.8085**	11
Svil	243.16	Supervillin	**2.6957**	3
Efhd2	26.8	EF hand domain containing 2	**2.6809**	11
Cttn	61.249	Cortactin, isoform CRA_e	**2.369**	26
Utrn	392.7	utrophin	**2.069**	3
Rab30	23.058	Ras-related protein Rab-30	**1.8759**	2
Iqgap1	188.75	Ras GTPase-activating-like protein IQGAP1	**1.8611**	12
Flnb	277.75	Filamin-B	**1.8565**	51
Cd44	85.837	CD44 antigen isoform a	**1.8313**	5
Plec	*534.18*	*plectin isoform 12alpha*	***1.8079***	*349*
Hspa5	72.421	78 kDa glucose-regulated protein	**1.7475**	24
Flna	*281.22*	*Filamin, alpha*	***1.7101***	*109*
Coro1b	53.912	Coronin-1B	**1.6833**	8
Plec-1a	*517.31*	*Isoform PLEC-1A of Plectin-1*	***1.6811***	*345*
Mfge8	51.268	Isoform 1 of Lactadherin	**1.5989**	5
Lrrfip2	47.147	Isoform 1 of Leucine-rich repeat flightless-interacting protein 2	**1.5866**	10
Hspa9	*73.527*	*Stress-70 protein, mitochondrial*	***1.5806***	*12*
Rpl19	*23.466*	*60S ribosomal protein L19*	***1.5507***	*2*
Dbn1	*72.544*	*Isoform E2 of Drebrin*	***1.5479***	*41*
Mme	85.701	Neprilysin	**1.4878**	6
Ppp1r12a	114.99	Isoform 1 of Protein phosphatase 1 regulatory subunit 12A	**1.48**	28
Hspa8	70.87	Heat shock cognate 71 kDa protein	**1.401**	31
G6pd2	59.125	Glucose-6-phosphate 1-dehydrogenase 2	**1.3867**	1
Flii	144.8	Protein flightless-1 homolog	**1.3312**	31

Myo18a	*233.98*	*Isoform 6 of Myosin-XVIIIa*	***1.2829***	*32*
Vim	*53.687*	*Vimentin*	***1.2265***	*54*
Pik3c2a	190.76	Isoform 1 of Phosphatidylinositol-4-phosphate 3-kinase C2 domain-containing alpha polypeptide	**1.2037**	2
Nes	207.12	Isoform 1 of Nestin	**1.1324**	53
Coro1c	*53.12*	*Coronin-1C*	***1.0771***	*27*
Arpc5l	*16.98*	*Actin-related protein 2/3 complex subunit 5-like protein*	***1.0714***	*4*
Twf1	40.079	Twinfilin-1	**1.0594**	12
Rbm14	69.448	Isoform 1 of RNA-binding protein 14	**1.0437**	5
Flnc	292.34	Isoform 1 of Filamin-C	**1.042**	18
Cd109	161.66	CD109 antigen	**1.0224**	5
Arpc4	*19.667*	*Actin-related protein 2/3 complex subunit 4*	***1.0128***	*7*

ES

Name	MW (kD)	Description	MOF/IgG	Peptides
Cltc	*191.98*	*clathrin, heavy polypeptide (Hc)*	***8.3419***	*9*
Clint1	69.758	clathrin interactor 1	**5.5724**	6
Kat8	**52.573**	**K(lysine) acetyltransferase 8**	**5.1921**	**7**
H2afv	*13.608*	*H2A histone family, member V*	***2.7119***	*4*
Pdcd11	207.78	programmed cell death 11	**1.7509**	4
Wdr3	105.77	WD repeat-containing protein 3	**1.2309**	4

Results are ordered according to ratio of MOF/IgG, for proteins above 1:1 background level. Italics indicate frequent contaminants of SILAC experiments using Dynabeads (Trinkle-Mulcahy et al., 2008). MOF is highlighted in red. Green highlighting indicate proteins shared between MEF and ES cells.

6.4 MOF chromodomain shows weak interactions with a variety of histone modifications

The presence of a chromodomain in MOF is conserved from yeast to human. Though the domain is very highly conserved in vertebrates it has much lower conservation between flies and mammals (Fig 6.4A) (Sievers et al., 2011). Chromodomains are conserved protein domains, usually involved in chromatin remodelling (Yap and Zhou, 2011). Some are known to recognise histone modifications; for example, the chromodomain of Drosophila HP1 binds to methylated H3K9, and this binding is necessary for HP1 dependent gene silencing. (Bannister et al., 2001) The Drosophila Polycomb (Pc) protein similarly binds methylated lysine, but has a preference for H3K27 (Fischle et al., 2003).

In Drosophila, the chromobarrel domain of MOF has been well studied, and is extremely important for the enzymatic activity. Disruption of the domain leads to global loss of H4K16ac in both male and female flies, though MOF is still bound to chromatin in this instance (Conrad et al., 2012). The chromodomain of MOF in mammals is less well studied, though the lack of conservation between flies and mammals implies that there may be some difference in function.

To investigate this further, I cloned the chromodomain of mouse MOF (the domain is identical in humans) with the addition of a GST tag. I expressed the protein in bacteria, then purified the protein for use in a MODified histone peptide array. There was no single histone modification which was clearly bound by the chromodomain of MOF. The top hits (Fig6.3D) were H3K36ac, H3K36me3, and H4R17me2a, though the top two modifications are found only once on the MODified peptide array (Fig 6.3C), whilst H4R17me2a is found on 8 spots on the array (giving a more robust estimation of specificity).

By comparison to the analysis of the highly specific H4K16ac antibody (Fig 3.1C), or of specific histone modification binding domain from the literature (Pradeepa et al., 2012), the specificity of binding to both these marks is extremely low, and may be purely background. Further work (for example, peptide pull downs using these modifications) would be necessary to confirm whether the interaction takes place *in vitro* by a different experimental method, and also *in vivo*.

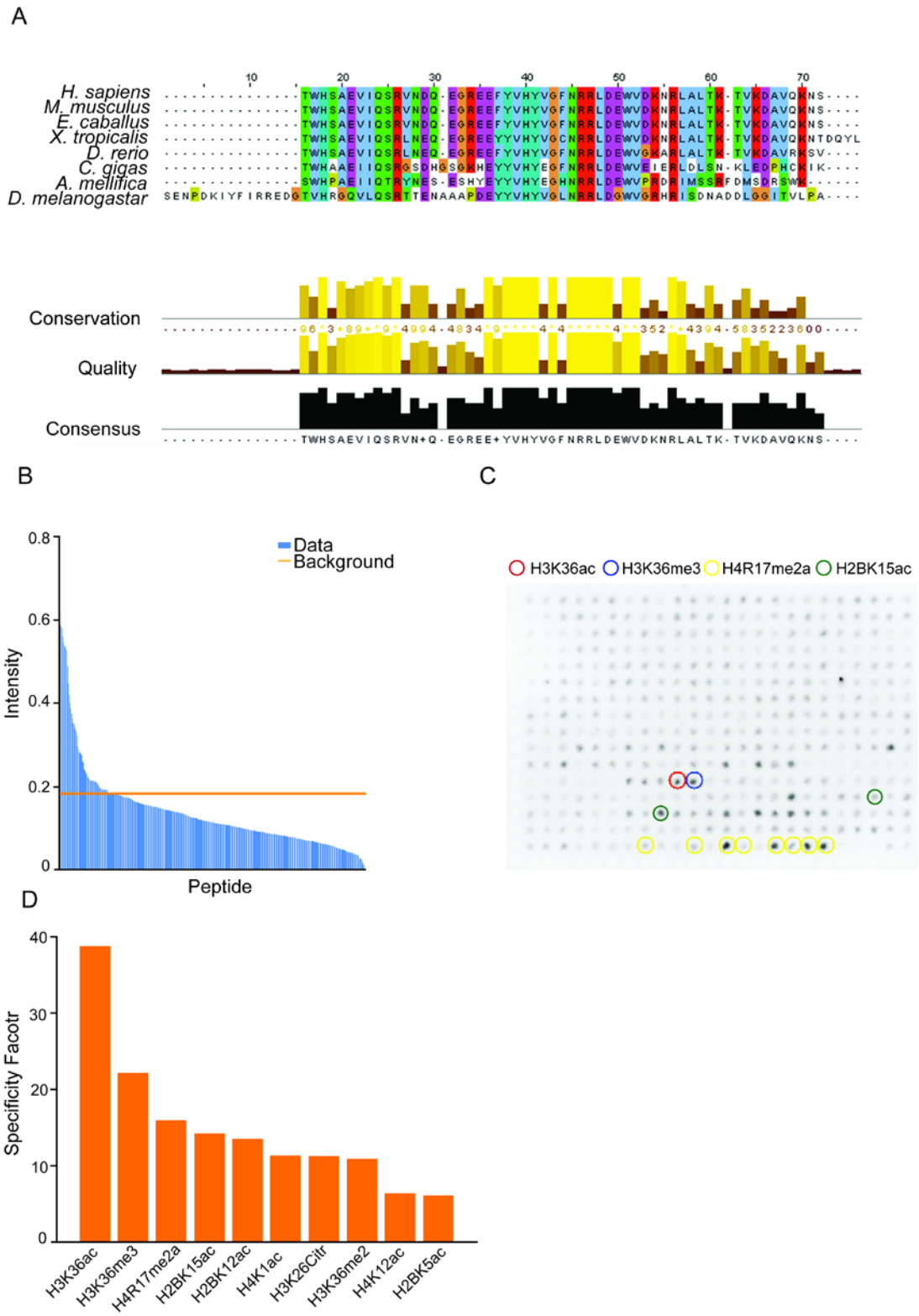

A
H. sapiens
M. musculus
E. caballus
X. tropicalis
D. rerio
C. gigas
A. mellifica
D. melanogastar
Conservation
Quality
Consensus
B
Data
Background
Intensity
Peptide
C
H3K36ac
H3K36me3
H4R17me2a
H2BK15ac
D
Specificity Facotr
H3K36ac
H3K36me3
H4R17me2a
H2BK15ac
H2BK12ac
H4K1ac
H3K26Citr
H3K36me2
H4K12ac
H2BK5ac

Figure 6.4 *MOF chromodomain has weak interaction with multiple histone modifications* (Previous page)

A: Sequence alignment (ClustalOmega) of chromodomains detected by SMART (Letunic et al., 2011)from full length MOF sequences from vertebrate (human, mouse, horse, frog, zebrafish), and invertebrate (honeybee, oyster, drosophila) species. B: Signal intensities for all spots on MODified Histone Peptide array. Orange line represents threshold for signal intensity designated as background. C: Example of a-GST antibody signal after incubation of GST-MOF chromodomain on MOFified Histone Peptide array. Spots containing top modification hits are circled as described above. D: Specificity factors calculated by Array Analyse software (Activ motif) given background designated in (B) for top 10 hits (n=2).

6.5 Summary of MOF interactions

From these results, multiple future directions present themselves. The initial overexpression was unsuccessful, but the endogenous immunoprecipitation showed promising preliminary results which can be built on. The chromodomain of Drosophila MOF has been shown to bind RNA, and this is necessary for its function. The chromodomain of mammalian MOF, however, though highly conserved between vertebrates, is diverged from its homolog in Drosophila, and has not been shown to have the ability to bind RNA. Neither has it been examined for ability to bind methylated amino groups, as many chromodomains can (Bannister et al., 2001; Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002). My results showed a very low specificity binding to both acetyl and methyl H3K36, along with a number of other acetylations and methylations of Lysine (and one Arginine residue). Further work is needed to determine whether this is biologically significant.

I had previously intended to observe the effects of a shRNA mediated knock down of MOF in ES cells and upon differentiation. A MOF knock out experiment was performed in ES cell by Li et al (2012), whose data showed that MOF is vital for ES cell renewal and correct gene expression, emphasising its particular relevance for pluripotency.

MOF is expressed in a wide variety of cell types to a similar extent (Thomas et al., 2007), but seems to be of particular importance for ES cells and pluripotency. As shown in previous chapters, H4K16ac may mark enhancers only in ES cells. Therefore, determining the complex (or complexes) or which MOF is a part in ES cells could be an important clue as to its mechanism in this cell type.

Chapter 7: Discussion and Future Directions

Acetylation of H4K16 has been widely studied *in vitro* for its effects on chromatin compaction, and in Drosophila for its role in the dosage compensation complex, but its genome wide profile in mammalian cells has been limited to only a few studies. In spite of its importance in chromatin structure, H4K16ac was not one of the histone acetylations included as part of the ENCODE (Consortium, 2012) or Roadmap (http://www.roadmapepigenomics.org/complete_epigenomes/) projects to locate regulatory elements of the genome. Though H4K16ac and MOF are known to be of particular importance to mammalian development, and knock out of MOF leads to loss of self-renewal in pluripotent stem cells, until this study, there had been no genome wide profile of H4K16ac in ES cells, or during differentiation.

Therefore, the aim of this work was to generate a genome wide profile of H4K16 acetylation (H4K16ac) in mouse Embryonic Stem cells (ES cells), and to determine whether changes in H4K16ac over large domains could be used to predict changes in chromatin compaction and gene expression. A further aim was to study the Histone Acetyl Transferase (HAT), MOF, which is specifically responsible for H4K16 acetylation, by looking at its complex or complexes in ES cells.

I have shown that H4K16ac is increased on the promoters of active genes, and that the profile changes in a manner broadly correlated with expression upon differentiation. Surprisingly, in ES cells, H4K16ac is also present on a subset of genes connected with differentiation. Secondly, I found that H4K16ac and MOF were located on active enhancers, sometimes independently of H3K27ac, which is also known to mark active enhancers. Unlike H3K27ac, however, acetylation of H4K16ac on enhancers is potentially ES cell specific.

Thirdly, loss of H4K16ac upon differentiation over a large domain was not found to correlate with a change chromatin compaction when analysed by FISH.

Finally, it was not possible to overexpress MOF in ES cells, and so further work is needed to elucidate the role of the MOF containing protein complexes in the mechanism of H4K16ac targeting. However, the chromodomain of mammalian MOF was shown to have weak specificity for several lysine modifications of histone tails.

7.1 H4K16ac and global transcription

The profile of H4K16ac has been examined globally in $CD4^+$T cells (Wang et al., 2008), and, more recently, in immature and mature erythrocytes (Wong et al., 2011), and mouse day E14.1 mouse embryo brains (Badeaux et al., 2012). The correlation of H4K16ac with transcriptional level was not tested in the study of mouse embryo brains, but in both developing erythrocytes and $CD4^+$T cells, H4K16ac presence was highly correlated with transcriptional activation of genes.

My data shows that this is also true in mouse ES cells; those genes upregulated in undifferentiated (UD) ES cells have higher levels of H4K16ac in UD cells than genes upregulated in differentiated cells, and vice versa. Similarly, genes which move from low expression in UD ES cells to high expression in differentiated cells also gain H4K16ac upon differentiation. Unsurprisingly given the correlation with increased expression, I also found that H4K16ac correlates well with presence of RNA polymerase II on promoters in UD ES cells.

The effect of MOF and H4K16ac on transcription is not only correlative. It is known that siRNA mediated knockdown of MOF leads to knockdown of a selection of genes in HeLa cells (Dou et al., 2005). In Drosophila, twofold upregulation of genes on the male X chromosome depends on MOF (Hilfiker et al., 1997). *In vitro* (and *in vivo* in yeast), targeting H4K16ac to the promoter of a gene can relieve repression (Akhtar and Becker, 2000).

How H4K16ac and MOF achieve activation/derepression of transcription is still not precisely known, but it has been linked to both the processes of transcriptional initiation and transcriptional elongation.

H4K16ac and MOF are found around the 5' end of active genes on the autosomes in both Drosophila and mammals. This profile supports a role in transcriptional initiation. Further evidence for this is found in the fact that the MOF-containing NSL complex is required for recruitment of RNA Polymerase II (Pol II) and the pre-initiation complex at housekeeping genes in Drosophila (though levels of H4K16ac at these promoters was not examined in this study) (Lam et al., 2012). In addition, the MOF-containing MSL complex has been suggested to upregulate the genes of the male X chromosome by enhancing transcriptional initiation via Pol II binding at these promoters (Conrad et al., 2012). This study also suggested that the MSL complex may be involved in transcriptional elongation, since in the

genes of the male X chromosome, Pol II levels on gene bodies are increased compared to the female X chromosome. Transcriptional elongation is another mechanism which H4K16ac has been suggested to positively influence in Drosophila, leading to upregulation of transcription. Knockdown of components of the MSL complex leads to loss of Pol II throughout the body of the gene (Larschan et al., 2011).

My data shows that, like in $CD4^{+}$T cells, H4K16ac in ES cells is present broadly over the promoters of active genes (Chapter 3, Fig 3.10), but does extend into the body of the genes somewhat. In Drosophila, there is evidence for segregation of the mechanisms by which MOF can enhance transcription depending on the function of the genes. On the male-X linked promoters, transcriptional elongation has a larger role, evidenced by a broader profile of MOF and H4K16ac on these genes compared with autosomal genes where they are limited to the 5' end and therefore may be more linked with transcriptional initiation (Kind et al., 2008).

In mammals, the picture is more complicated, as MOF binding in ES cells has been split into two classes based on the binding profile (either 5'/promoter binding, or binding in the body of the gene), but as yet no functional relationship has been elucidated between MOF binding profile and gene type (Li et al., 2012). It would be interesting to classify genes based on the presence of H4K16ac in a similar way, to determine whether the genes which emerge as marked on the 5' with H4K16ac versus genes marked through the gene body correlate with those so bound by MOF.

7.2 H4K16ac, pluripotency and development

H4K16ac has a global correlation with active transcription, as expected from previous data, but it is also, specifically, important for maintenance of pluripotency in a manner which seems to be independent of its activity in differentiated cells. Li et al (2012) showed that MOF was intrinsic to the pluripotency network, but that overexpression of Nanog could partially rescue the $MOF^{-/-}$ phenotype, implying that once Nanog expression has been achieved, MOF is at least somewhat dispensable.

Consistent with that study's report of a reduction in H4K16ac upon retinoic acid differentiation, the level of H4K16ac I detected by ChIP-sequencing was generally lower in the differentiated samples than the undifferentiated (Chapter 4). Although H4K16ac is reduced on the promoters of pluripotency genes in $MOF^{-/-}$ ES cells, it is not clear whether

their reduction in expression is due to hypoacetylation at H4K16. MOF is also known to acetylate p53 when in the context of the NSL complex, and this acetylation leads to upregulation of a selection of p53 target genes (Li et al., 2009). Nanog is silenced by p53 upon retinoic acid differentiation (Lin et al., 2004), but it is not clear whether MOF is found within a complex which renders it capable of accepting K120 of p53 as a substrate in ES cells, or whether this acetylation would prevent p53 silencing Nanog in undifferentiated ES cells (Fig 7.1A). Further work is therefore needed to separate the histone acetylation activity of MOF from the p53 acetylation activity to determine whether the H4K16ac I find on the pluripotency genes in ES cells is functionally important to their expression.

In Drosophila, MOF is found on many of the active promoters, but not all (Kind et al., 2008), and the MOF containing NSL complex binds to a large number of housekeeping genes, but only activates a defined subset (Feller et al., 2011). In mammals, it has been noted that MOF binds to differentiation associated genes in ES cells, even though these genes are not expressed in that cell type (Li et al., 2012). The authors suggest that MOF has a priming activity on genes "poised" for later activation; similarly, I have shown that peaks of H4K16ac are found over the promoters of genes associated with differentiation (Chapter 4, Fig 4.9). This is consistent with evidence from *S. cerevisae*, in which inducible genes are acetylated even in the non-inducible state (Roh et al., 2004), notably, for the PHO5 gene, transcription in the induced state is dependent on prior acetylation (Nourani et al., 2004).

Interestingly, H4K16ac is excluded from the early Hox genes in undifferentiated ES cells (Chapter 4, Fig 4.4), which are classic examples of poised developmental genes, bivalently marked by H3K27me3 and H3K4me3 (Bernstein et al., 2006). MOF was also shown to have no significant enrichment on bivalent domains (Li et al., 2012). This is in contrast with H3K9ac and H3K14ac which are both found alongside H3K27me3 and H3K4me3 at bivalent promoters (Karmodiya et al., 2012). This, and my data supports the idea that histone acetylation can 'prime' genes for subsequent activation, and presents the possibility of another subset of poised developmental genes, separate from known bivalent promoters, which are marked by at least H4K16ac and/or MOF.

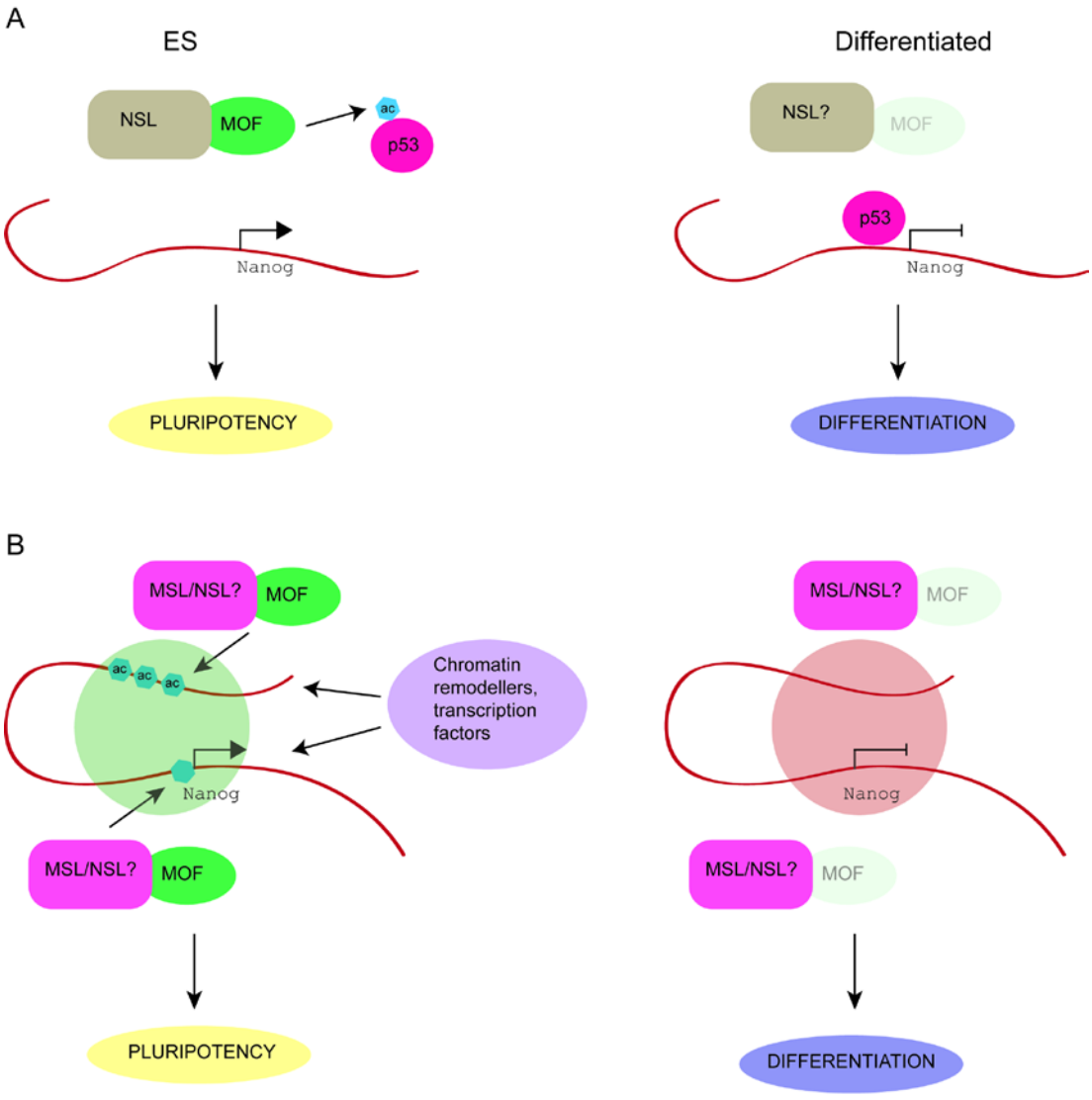


Figure 7.1: *Potential models for action of MOF/H4K16ac on Nanog transcription*

A: MOF, in the context of the NSL complex acetylates p53 on K120, which prevents p53 from silencing its target gene, Nanog. Nanog is expressed and pluripotency is maintained. Upon differentiation, MOF levels are reduced (the NSL complex may still assemble), and unacetylated p53 is free to silence Nanog, leading to differentiation. B: MOF, in the context of either the MSL or NSL complex acetylates histones on H4K16 at the promoter and regulatory region of Nanog, generating an open chromatin environment and leading to recruitment of transcription factors and chromatin remodellers, which then leads to transcription of Nanog and maintenance of pluripotency. In the absence of MOF, the chromatin environment is non-conducive to TF binding, and Nanog is silenced, leading to differentiation.

7.3 H4K16ac and Long Range Control

One of the most important findings of this study was the fact that H4K16ac is present on active enhancers in ES cells. This was unexpected, since a previous study in $CD4^+$T cells analysed regulatory regions for histone modifications, and H4K16ac was not significantly enriched on enhancers in that cell type (Wang et al., 2008). However, previous studies have shown that H3 acetylation generally is enriched on regulatory regions (Roh et al., 2005; Wang et al., 2008). In addition, H3K9ac and H3K14ac were recently noted to mark active enhancers in ES cells (Karmodiya et al., 2012); neither of these modifications was found to be enriched on regulatory regions in the earlier study by Wang et al. Potentially, this is simply due to a low level of sequencing depth in the early experiment, but it could also be due to cell type specificity. On examination of a well-known limb regulatory region in a limb cell line, I was unable to detect any H4K16ac (Chapter 4, Fig 4.14), whilst a modest amount of H3K27ac and H3K4me1 was present on the region in this cell type.

Interestingly, I noticed that though there is a good correlation between H4K16ac and H3K27ac on active enhancers, there is a subset of enhancers marked by one of the two modifications (together with H3K4me1), but not both. This is reminiscent of the 'classes' of enhancers which were defined by Zentner et al (2011) who used H3K27ac, H3K27me3, H3K36me3 and H3K4me1 to subclassify active enhancers. It is possible that the post translational modification profile of individual enhancers are more variable than previously thought; though to confirm this it would be necessary to generate genome wide ChIP-seq datasets for all possible histone modifications in a variety of cell types. The fact that H3K27ac, H3K14ac, H3K9ac and H4K16ac have been found on regulatory regions in ES cells suggests the possibility that histone acetylation is a general feature of active enhancers in this cell type. Knockout and knockdown of a number of HATs has been shown to have an adverse effect on regulation of differentiation, which points to a more general role of histone acetylation, but MOF is the only HAT which targets the core pluripotency factors directly; the others function downstream – for example, depletion of Nanog causes reduction of Tip60-p400 binding to target promoters (Fazzio et al., 2008), but depletion of Gcn5, p300, or Tip60 does not lead to reduction in expression of the core pluripotency factors in ES cells (Lin et al., 2007; Fazzio et al., 2008; Zhong and Jin, 2009) as depletion of MOF does. Further work with $MOF^{-/-}$ ES cells would be necessary to determine whether ablation of H4K16ac has an effect on other histone acetylation modifications at regulatory regions.

What function could H4K16ac perform on enhancers, and is its presence there necessary? This is a question which is asked of all the histone modifications found on enhancers. In the example of virally induced activation of the IFN-β enhancer, a group of induced transcription factors binds to a nucleosome free region, and recruits the HATs Gcn5 and CBP. Acetylation induces binding of the SWI/SNF complex, leading to nucleosome remodelling, which is necessary for induction of IFN-β transcription (Agalioti et al., 2000). This is, however, an individual case, and further work is needed to determine whether histone acetylation has a general function on active enhancers genome wide.

7.4 H4K16ac and chromatin compaction

In vitro, H4K16 acetylation status directly affects the chromatin compaction state; H4K16ac of nucleosome arrays leads to their decompaction, and its presence prevents the compaction which would otherwise be induced by increasing salt concentration. In this regard, it is one of the few histone modifications to be directly correlated with chromatin compaction state.

In vivo, however, I was unable to use H4K16ac state to detect changes in chromatin compaction at a level that could be detected by FISH (Chapter 5). This suggests that FISH may measure a different level of chromatin structure than that measured in many in vitro assays.

Firstly, H4K16ac has been shown to be very important for cation induced intramolecular association of nucleosome arrays (Fig 7.2). Acetylation of H4K16 abolishes array folding (Shogren-Knaak et al., 2006; Robinson et al., 2008), and to a much greater extent than the charge abolishing H4K16→Q mutation, or acetylation of H4K5, H4K8 and H4K12 (Allahverdi et al., 2010; Liu et al., 2011). This is likely due to the role of H4K16 in the interaction between the H4 tail of one nucleosome and H2A/H2B of the following nucleosome (Fig 1.7). By contrast, H4K16ac has a minimal effect on cation induced inter-array self-association (Fig 7.2); acetylation of H4K5, K8, K12 and K16 or mutation of those lysine residues to uncharged glutamine has a much greater effect than acetylation or mutation of K16 alone. This supports the idea that nucleosome array self-association occurs due to a non-specific electrostatic mechanism (Allahverdi et al., 2010; Liu et al., 2011). Potentially, this second mechanism of chromatin compaction can be detected by FISH, whilst the first cannot. The acetylation state of the other H4 lysine residues in UD ES and differentiated cells is unknown.

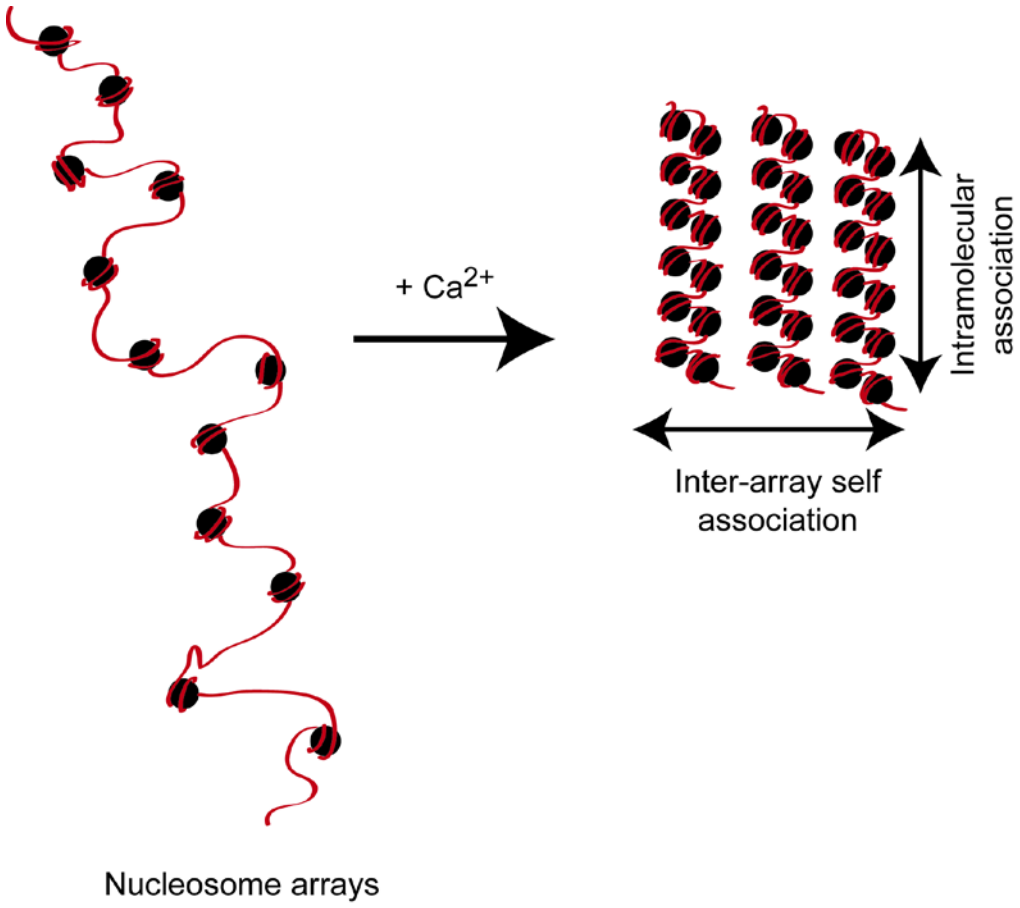


Figure 7.2: *Illustration of cation induced nucleosome array folding.*

Upon addition of a cation, such as Ca^{2+}, nucleosome arrays condense, both within the array itself ("Intramolecular association"), which can be measured by analytical ultracentrifugation, and between arrays ("Inter-array self-association"), which can be measured by precipitation assays. H4K16ac is important for intramolecular association, but is less important for inter-array self-association.

Alternatively, my failure to detect a change in chromatin compaction state around the Nanog and Sox2 loci upon differentiation may be because deacetylation of H4K16ac is only one step in the process of chromatin compaction; indeed, chromatin compaction *in vitro* requires both hypoacetylaton of H4K16 and removal of the H1 linker histone. The presence of the H1 histone has not been examined in this work. It would be interesting to increase the length of cellular exposure to retinoic acid to determine whether any of the loci undergo compaction at a later stage in differentiation.

Finally, the effect on chromatin compaction of a *gain* of H4K16ac over a large locus was not examined in this work, due to a difficulty in locating large regions which gain H4K16ac upon differentiation (as opposed to small, localised gains over promoters of expressed genes), potentially due to an overall loss of H4K16ac upon retinoic acid differentiation (Li et al., 2012). Therefore, to look at chromatin compaction after a gain of H4K16ac, it may be necessary to use another system of differentiation where this does not occur.

7.5 MOF

My preliminary investigation into MOF function showed that overexpression of MOF in ES cells was potentially detrimental to cell survival, as ES cells quickly silenced flag-MOF overexpression, and even under antibiotic selection, expression was extremely low/undetectable by western blot. This implies that dosage of MOF in ES cells may be important. Previous studies have overexpressed MOF in HeLa cells (Dou et al., 2005; Mendjan et al., 2006; Cai et al., 2009), indicating that this sensitivity may be cell type specific. This would be in line with the specific role of MOF in ES cells.

Since I was unable to overexpress flag-MOF in ES cells, isolating the complex in ES cells became more complicated. To perform an unbiased screen of MOF interaction partners, I used an antibody against endogenous MOF to perform an IP in conjunction with SILAC mass spectrometry in Mouse Embryonic Fibroblasts (MEFs) and ES cells. In MEFs, as expected from previous studies which used overexpressed tagged proteins, MOF was partnered with MSL1, 2, and 3, though I did not detect interaction with any members of the NSL complex. In ES cells, I was unable to isolate MOF in conjunction with any of the MSL or NSL proteins, which may be due to technical problems. Further work is needed to determine whether the complex differs between MEFs and ES cells. Size separation of protein complexes found in ES cell nuclei confirm that MOF is found in a complex of size

~700kDa in ES cells (Fig 6.3A). This experiment could be repeated for MEF nuclei to look for differences in complex size.

In addition to looking at the protein interactions of MOF, I also investigated the similarity of mammalian MOF chromodomain with the domain in fish, frogs, and invertebrates. I also examined the binding specificity of its chromodomain for modified histones, which could potentially explain how MOF and H4K16ac are targeted to specific loci. The results showed that the chromodomain binds weakly to a variety of lysine or arginine residues. The strongest specificity was to H3K36ac, followed by H3K36me3 and H4R17me2a. There is so far no data to connect H3K36ac to MOF binding or H4K16ac, with the exception that both are found on active promoters, and show higher levels on inducible genes upon stimulation of lymphocytes (Wang et al., 2008; Lim et al., 2009).

Interestingly, however, H3K36me3 has been shown to have an antagonistic cross talk with H4K16ac. Reduction in levels of H3K36me3 (but not H3K36me2) leads to increase of H4K16ac across the gene body of examined genes (Bell et al., 2007). Potentially, this could mean that binding of MOF chromodomain to H3K36me3 leads to a reduction in MOF HAT activity. Binding specificity and other functions of the MOF chromodomain have not been examined in mammals before. However, in Drosophila it is required for MOF binding to the dosage compensation linked non-coding RNA roX, and for global H4K16ac, via an interaction with DNA which influences the HAT activity of MOF. The chromodomain of MOF is extremely well conserved in vertebrates, but there is significantly more divergence between mammals and flies. This indicates that there may be differences in mechanism and/or binding capabilities. Further work is needed to determine whether MOF binds to DNA *in vivo* in mammalian cells, and whether the weak binding to modified lysine/arginine residues is physiologically significant.

7.6 Future Directions

The work presented in this thesis provides new insight into the role of H4K16ac during pluripotency and differentiation. Whilst examination of data in three cell types suggests that H4K16ac presence on enhancers is ES cell specific, this is by no means a comprehensive analysis, and more cell types must be examined before this conclusion can be drawn.

To show that H4K16ac presence on ES cell enhancers has a role in gene regulation, it would be necessary to show that induced hypoacetylation of H4K16ac at these enhancers would

negatively regulate gene expression. Targeted recruitment of a SIR family deacetylase to one of these enhancers via zinc fingers or TALEN fusion proteins might be an interesting avenue for further investigation. However, whilst SirT2 shows a preference for H4K16ac *in vivo*, it has not been tested fully for specificity to loci of interest (Vaquero et al., 2006). Since targeting SirT2 to pluripotency promoters or enhancers would not affect the ability of MOF to acetylate p53, this method could provide a way of decoupling the effect of MOF acetyltransferase activity on histones and on p53.

In addition, it will be interesting to determine whether hypoacetylation of H4K16ac has a different downstream effect than hypoacetylation of H3K27ac, or other acetylations which are known to mark enhancers.

Building on the SILAC results presented in Chapter 4 would help be to determine whether H4K16ac presence on enhancers is dependent on a particular complex. If this is the case, the complex could be targeted (for example by RNAi), and the downstream effects on transcription analysed. Finally, the problem of mechanism could be approached from another angle, by investigating whether H4K16ac is necessary for protein binding at enhancers or promoters. This could be performed using sequential re-ChIP for H4K16ac, then H3K4me1 or H3K4me3. Candidate proteins could be identified by immunoblot or mass spectrometry, then analysed for presence of a bromodomain, indicating potential to bind to H4K16ac.

Bibliography

Journals

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000). Ordered Recruitment of Chromatin Modifying and General Transcription Factors to the IFN-β Promoter. Cell *103*, 667–678.

Akhtar, A. (2001). The histone H4 acetyltransferase MOF uses a C2HC zinc finger for substrate recognition. EMBO Reports *2*, 113–118.

Akhtar, A., and Becker, P.B. (2000). Activation of Transcription through Histone H4 Acetylation by MOF, an Acetyltransferase Essential for Dosage Compensation in Drosophila. Molecular Cell *5*, 367–375.

Akhtar, A., Zink, D., and Becker, P.B. (2000). Chromodomains are protein–RNA interaction modules. Nature *407*, 405–409.

Alexandrow, M.G., and Hamlin, J.L. (2005). Chromatin decondensation in S-phase involves recruitment of Cdk2 by Cdc45 and histone H1 phosphorylation. J Cell Biol *168*, 875–886.

Allahverdi, A., Yang, R., Korolev, N., Fan, Y., Davey, C.A., Liu, C.-F., and Nordenskiöld, L. (2010). The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. Nucleic Acids Res.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. PNAS *51*, 786–794.

Anderson, E., Peluso, S., Lettice, L.A., and Hill, R.E. (2012). Human limb abnormalities caused by disruption of hedgehog signaling. Trends in Genetics *28*, 364–373.

Andy Bannister, T.K. (2012). Guide to epigenetic marks | Abcam.

Attikum, H. van, and Gasser, S.M. (2005). The histone code at DNA breaks: a guide to repair? Nature Reviews Molecular Cell Biology *6*, 757–765.

Axton, R., Hanson, I., Danes, S., Sellar, G., Heyningen, V. van, and Prosser, J. (1997). The incidence of PAX6 mutation in patients with simple aniridia: an evaluation of mutation detection in 12 cases. J Med Genet *34*, 279–286.

Badeaux, A.I., Yang, Y., Cardenas, K., Vemulapalli, V., Chen, K., Kusewitt, D., Richie, E., Li, W., and Bedford, M.T. (2012). Loss of the methyl lysine effector protein PHF20 impacts the expression of genes regulated by the lysine acetyltransferase MOF. J. Biol. Chem. *287*, 429–437.

Bain, G., Kitchens, D., Yao, M., Huettner, J.E., and Gottlieb, D.I. (1995). Embryonic stem cells express neuronal properties in vitro. Dev. Biol. *168*, 342–357.

Baker, B.S., Gorman, M., and Marin, I. (1994). Dosage Compensation in Drosophila. Annual Review of Genetics *28*, 491–521.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Bannister, A.J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. , Published Online: 26 December 1996; | Doi:10.1038/384641a0 *384*, 641–643.

Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial Distribution of Di- and Tri-methyl Lysine 36 of Histone H3 at Active Genes. J. Biol. Chem. *280*, 17732–17736.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007a). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell *129*, 823–837.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007b). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell *129*, 823–837.

Baylin, S.B., and Schuebel, K.E. (2007). Genomic biology: The epigenomic era opens. Nature *448*, 548–549.

Bell, O., Schwaiger, M., Oakeley, E.J., Lienert, F., Beisel, C., Stadler, M.B., and Schübeler, D. (2010). Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. Nature Structural & Molecular Biology *17*, 894–900.

Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N.D., Schwaiger, M., MacAlpine, D.M., Zilbermann, F., Van Leeuwen, F., Bell, S.P., Imhof, A., et al. (2007). Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. EMBO J. *26*, 4974–4984.

Belmont, A.S., Dietzel, S., Nye, A.C., Strukov, Y.G., and Tumbar, T. (1999). Large-scale chromatin structure and function. Current Opinion in Cell Biology *11*, 307–311.

Benetti, R., García-Cao, M., and Blasco, M.A. (2007). Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. Nature Genetics *39*, 243–250.

Benko, S., Fantes, J.A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C.T., et al. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nature Genetics *41*, 359–364.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Billon, N., Jolicoeur, C., Ying, Q.L., Smith, A., and Raff, M. (2002). Normal timing of oligodendrocyte development from genetically engineered, lineage-selectable mouse ES cells. J Cell Sci *115*, 3657–3665.

Bird, A.W., Yu, D.Y., Pray-Grant, M.G., Qiu, Q., Harmon, K.E., Megee, P.C., Grant, P.A., Smith, M.M., and Christman, M.F. (2002). Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. Nature *419*, 411–415.

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

Bone, J.R., Lavender, J., Richman, R., Palmer, M.J., Turner, B.M., and Kuroda, M.I. (1994). Acetylated histone H4 on the male X chromosome is associated with dosage compensation in Drosophila. Genes Dev. *8*, 96–104.

Borycki, A.G., Brunk, B., Tajbakhsh, S., Buckingham, M., Chiang, C., and Emerson, C.P., Jr (1999). Sonic hedgehog controls epaxial muscle determination through Myf5 activation. Development *126*, 4053–4063.

Brookes, E., De Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157–170.

Brooks, C.L., and Gu, W. (2011). The impact of acetylation and deacetylation on the p53 pathway. Protein Cell *2*, 456–462.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation. Cell *84*, 843–851.

Buck, M.J., and Lieb, J.D. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics *83*, 349–360.

Bulger, M., and Groudine, M. (2011). Functional and Mechanistic Diversity of Distal Transcription Enhancers. Cell *144*, 327–339.

Cai, Y., Jin, J., Swanson, S.K., Cole, M.D., Choi, S.H., Florens, L., Washburn, M.P., Conaway, J.W., and Conaway, R.C. (2009). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. J Biol Chem.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002a). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039–1043.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002b). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science *298*, 1039–1043.

Chambeyron, S., and Bickmore, W. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription.

Collins, N., Poot, R.A., Kukimoto, I., García-Jiménez, C., Dellaire, G., and Varga-Weisz, P.D. (2002). An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. Nat. Genet. *32*, 627–632.

Conrad, T., and Akhtar, A. (2012). Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nature Reviews Genetics *13*, 123–134.

Conrad, T., Cavalli, F.M.G., Holz, H., Hallacli, E., Kind, J., Ilik, I., Vaquerizas, J.M., Luscombe, N.M., and Akhtar, A. (2012a). The MOF Chromobarrel Domain Controls Genome-wide H4K16 Acetylation and Spreading of the MSL Complex. Developmental Cell *22*, 610–624.

Conrad, T., Cavalli, F.M.G., Vaquerizas, J.M., Luscombe, N.M., and Akhtar, A. (2012b). Drosophila Dosage Compensation Involves Enhanced Pol II Recruitment to Male X-Linked Promoters. Science.

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

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., et al. (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res. *16*, 123–131.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U.S.A. *107*, 21931–21936.

Dang, W., Steffen, K.K., Perry, R., Dorsey, J.A., Johnson, F.B., Shilatifard, A., Kaeberlein, M., Kennedy, B.K., and Berger, S.L. (2009). Histone H4 lysine 16 acetylation regulates cellular lifespan. Nature *459*, 802–807.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 Å Resolution. Journal of Molecular Biology *319*, 1097–1113.

Davis, A.P., and Capecchi, M.R. (1996). A mutational analysis of the 5’ HoxD genes: dissection of genetic interactions during limb development in the mouse. Development *122*, 1175–1185.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. Journal of Biological Chemistry *285*, 26114–26120.

Diaz, A., Park, K., Lim, D.A., and Song, J.S. (2012). Normalization, bias correction, and peak calling for ChIP-seq. Stat Appl Genet Mol Biol *11*,.

Dingwall, C., Lomonossoff, G.P., and Laskey, R.A. (1981). High sequence specificity of micrococcal nuclease. Nucleic Acids Res. *9*, 2659–2673.

Dion, M.F., Altschuler, S.J., Wu, L.F., and Rando, O.J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. PNAS *102*, 5501–5506.

Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T.J. (2003). Chromatin Fiber Folding: Requirement for the Histone H4 N-terminal Tail. Journal of Molecular Biology *327*, 85–96.

Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L., and Roeder, R.G. (2005). Physical Association and Coordinate Function of the H3 K4 Methyltransferase MLL1 and the H4 K16 Acetyltransferase MOF. Cell *121*, 873–885.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Egelhofer, T.A., Minoda, A., Klugman, S., Lee, K., Kolasinska-Zwierz, P., Alekseyenko, A.A., Cheung, M.-S., Day, D.S., Gadel, S., Gorchakov, A.A., et al. (2011). An assessment of histone-modification antibody quality. Nature Structural & Molecular Biology *18*, 91–93.

Elgin, S.C.R. (1981). DNAase I-hypersensitive sites of chromatin. Cell *27*, 413–415.

Emmott, E., Rodgers, M.A., Macdonald, A., McCrory, S., Ajuh, P., and Hiscox, J.A. (2010). Quantitative Proteomics Using Stable Isotope Labeling with Amino Acids in Cell Culture Reveals Changes in the Cytoplasmic, Nuclear, and Nucleolar Proteomes in Vero Cells Infected with the Coronavirus Infectious Bronchitis Virus. Mol Cell Proteomics *9*, 1920–1936.

Eriksson, M., Brown, W.T., Gordon, L.B., Glynn, M.W., Singer, J., Scott, L., Erdos, M.R., Robbins, C.M., Moses, T.Y., Berglund, P., et al. (2003). Recurrent de novo point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. Nature *423*, 293–298.

Eskeland, R., Freyer, E., Leeb, M., Wutz, A., and Bickmore, W.A. (2010a). Histone acetylation and the maintenance of chromatin compaction by Polycomb repressive complexes. Cold Spring Harb. Symp. Quant. Biol. *75*, 71–78.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., et al. (2010b). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol. Cell *38*, 452–464.

Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., Jones, S., Bickmore, W., Fukushima, Y., Mannens, M., et al. (1995). Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum. Mol. Genet. *4*, 415–422.

Fazzio, T., Huff, J., and Panning, B. (2008). An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity. Cell *134*, 162–174.

Feller, C., Prestel, M., Hartmann, H., Straub, T., Söding, J., and Becker, P.B. (2011). The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. Nucleic Acids Research.

Felsenfeld, G., Boyes, J., Chung, J., Clark, D., and Studitsky, V. (1996). Chromatin structure and gene expression. PNAS *93*, 9384–9388.

Feng, B., Ng, J.-H., Heng, J.-C.D., and Ng, H.-H. (2009). Molecules that Promote or Enhance Reprogramming of Somatic Cells to Induced Pluripotent Stem Cells. Cell Stem Cell *4*, 301–312.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.

Fraga, M.F., and Esteller, M. (2007). Epigenetics and aging: the targets and the marks. Trends in Genetics *23*, 413–418.

Furniss, D., Lettice, L.A., Taylor, I.B., Critchley, P.S., Giele, H., Hill, R.E., and Wilkie, A.O.M. (2008). A variant in the sonic hedgehog regulatory sequence (ZRS) is associated

with triphalangeal thumb and deregulates expression in the developing limb. Human Molecular Genetics *17*, 2417–2423.

Gelbart, M.E., Larschan, E., Peng, S., Park, P.J., and Kuroda, M.I. (2009). Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. Nat. Struct. Mol. Biol.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin Architecture of the Human Genome: Gene-Rich Domains Are Enriched in Open Chromatin Fibers. Cell *118*, 555–566.

Gilmour, D.S., and Lis, J.T. (1985). In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. Molecular and Cellular Biology *5*, 2009.

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

Gordon, F., Luger, K., and Hansen, J.C. (2005). The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays. J. Biol. Chem. *280*, 33701–33706.

Grewal, S.I.S., and Elgin, S.C.R. (2007). Transcription and RNA interference in the formation of heterochromatin. Nature *447*, 399–406.

Gross, D.S., and Garrard, W.T. (1988). Nuclease Hypersensitive Sites in Chromatin. Annual Review of Biochemistry *57*, 159–197.

Gu, W., and Roeder, R.G. (1997). Activation of p53 Sequence-Specific DNA Binding by Acetylation of the p53 C-Terminal Domain. Cell *90*, 595–606.

Gupta, A., Sharma, G.G., Young, C.S.H., Agarwal, M., Smith, E.R., Paull, T.T., Lucchesi, J.C., Khanna, K.K., Ludwig, T., and Pandita, T.K. (2005a). Involvement of human MOF in ATM function. Mol. Cell. Biol *25*, 5292–5305.

Gupta, A., Sharma, G.G., Young, C.S.H., Agarwal, M., Smith, E.R., Paull, T.T., Lucchesi, J.C., Khanna, K.K., Ludwig, T., and Pandita, T.K. (2005b). Involvement of human MOF in ATM function. Mol. Cell. Biol *25*, 5292–5305.

Hebbes, T.R., Clayton, A.L., Thorne, A.W., and Crane-Robinson, C. (1994). Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. The EMBO Journal *13*, 1823.

Hebbes, T.R., Thorne, A.W., and Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J *7*, 1395–1402.

Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S.M., and Grunstein, M. (1995). Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. Cell *80*, 583–592.

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., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

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., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311–318.

Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A., and Lucchesi, J.C. (1997). mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila. The EMBO Journal *16*, 2054–2060.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. U.S.A. *106*, 9362–9367.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and 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*, 305–314.

Hooper, M., Hardy, K., Handyside, A., Hunter, S., and Monk, M. (1987). HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature *326*, 292–295.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., et al. (2012). InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res *40*, D306–D312.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS Lett. *583*, 1713–1720.

Illingworth, R.S., Botting, C.H., Grimes, G.R., Bickmore, W.A., and Eskeland, R. (2012). PRC1 and PRC2 are not required for targeting of H2A.Z to developmental genes in embryonic stem cells. PLoS ONE *7*, e34848.

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

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. Science *295*, 2080–2083.

Jeppesen, P., and Turner, B.M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell *74*, 281–289.

Jessell, T.M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nature Reviews Genetics *1*, 20–29.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nature Reviews Genetics *10*, 161–172.

Jiang, J., Chan, Y.-S., Loh, Y.-H., Cai, J., Tong, G.-Q., Lim, C.-A., Robson, P., Zhong, S., and Ng, H.-H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat. Cell Biol. *10*, 353–360.

Johnson, L.M., Kayne, P.S., Kahn, E.S., and Grunstein, M. (1990). Genetic evidence for an interaction between SIR3 and histone H4 in the repression of the silent mating loci in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. *87*, 6286–6290.

Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics *13*, 424.

Kennedy, B.K., Gotta, M., Sinclair, D.A., Mills, K., McNabb, D.S., Murthy, M., Pak, S.M., Laroche, T., Gasser, S.M., and Guarente, L. (1997). Redistribution of Silencing Proteins from Telomeres to the Nucleolus Is Associated with Extension of Life Span in S. cerevisiae. Cell *89*, 381–391.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, and D. (2002). The Human Genome Browser at UCSC. Genome Res. *12*, 996–1006.

Keohane, A.M., O'neill, L.P., Belyaev, N.D., Lavender, J.S., and Turner, B.M. (1996). X-Inactivation and Histone H4 Acetylation in Embryonic Stem Cells. Developmental Biology *180*, 618–630.

Kim, J., and Lauderdale, J.D. (2006). Analysis of Pax6 expression using a BAC transgene reveals the presence of a paired-less isoform of Pax6 in the eye and olfactory bulb. Developmental Biology *292*, 486–505.

Kimura, A., Umehara, T., and Horikoshi, M. (2002). Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. Nature Genetics *32*, 370–377.

Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P., and Akhtar, A. (2008). Genome-wide Analysis Reveals MOF as a Key Regulator of Dosage Compensation and Gene Expression in Drosophila. Cell *133*, 813–828.

Kleinjan, D.A., Seawright, A., Elgar, G., and Van Heyningen, V. (2002). Characterization of a novel gene adjacent to PAX6, revealing synteny conservation with functional significance. Mamm. Genome *13*, 102–107.

Kleinjan, D.A., Seawright, A., Schedl, A., Quinlan, R.A., Danes, S., and Van Heyningen, V. (2001). Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. Hum. Mol. Genet. *10*, 2049–2059.

Kmita, M., and Duboule, D. (2003). Organizing Axes in Time and Space; 25 Years of Colinear Tinkering. Science *301*, 331–333.

Knezetic, J.A., and Luse, D.S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell *45*, 95–104.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaöz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., et al. (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res. *17*, 691–707.

Kondilis-Mangum, H.D., and Wade, P.A. (2012). Epigenetics and the adaptive immune response. Molecular Aspects of Medicine.

Koolen, D.A., Kramer, J.M., Neveling, K., Nillesen, W.M., Moore-Barton, H.L., Elmslie, F.V., Toutain, A., Amiel, J., Malan, V., Tsai, A.C.-H., et al. (2012). Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. Nature Genetics *44*, 639–641.

Kornberg, R.D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. Science *184*, 868–871.

Kornberg, R.D., and Thomas, J.O. (1974). Chromatin Structure: Oligomers of the Histones. Science *184*, 865–868.

Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Mol. Cell *44*, 410–423.

Krishnan, V., Chow, M.Z.Y., Wang, Z., Zhang, L., Liu, B., Liu, X., and Zhou, Z. (2011). Histone H4 lysine 16 hypoacetylation is associated with defective DNA repair and premature senescence in Zmpste24-deficient mice. Proc. Natl. Acad. Sci. U.S.A *108*, 12325–12330.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., et al. (2003). Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Mol. Cell. Biol. *23*, 4207–4218.

Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004). Different Ezh2-Containing Complexes Target Methylation of Histone H1 or Nucleosomal Histone H3. Molecular Cell *14*, 183–193.

Lam, K.C., Mühlpfordt, F., Vaquerizas, J.M., Raja, S.J., Holz, H., Luscombe, N.M., Manke, T., and Akhtar, A. (2012). The NSL complex regulates housekeeping genes in Drosophila. PLoS Genet. *8*, e1002736.

Lange, M., Kaynak, B., Forster, U.B., Tönjes, M., Fischer, J.J., Grimm, C., Schlesinger, J., Just, S., Dunkel, I., Krueger, T., et al. (2008). Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. Genes Dev. *22*, 2370–2384.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Larschan, E., Bishop, E.P., Kharchenko, P.V., Core, L.J., Lis, J.T., Park, P.J., and Kuroda, M.I. (2011). X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature *471*, 115–118.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73–84.

Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., Beer, P. de, Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and Graaff, E. de (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.

Lettice, L.A., Hill, A.E., Devenney, P.S., and Hill, R.E. (2008). Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. Hum. Mol. Genet. *17*, 978–985.

Lettice, L.A., Horikoshi, T., Heaney, S.J.H., Baren, M.J. van, Linde, H.C. van der, Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. PNAS *99*, 7548–7553.

Lettice, L.A., Williamson, I., Wiltshire, J.H., Peluso, S., Devenney, P.S., Hill, A.E., Essafi, A., Hagman, J., Mort, R., Grimes, G., et al. (2012). Opposing Functions of the ETS Factor Family Define Shh Spatial Expression in Limb Buds and Underlie Polydactyly. Developmental Cell *22*, 459–467.

Letunic, I., Doerks, T., and Bork, P. (2011). SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Research *40*, D302–D305.

Li, B., Carey, M., and Workman, J.L. (2007). The Role of Chromatin during Transcription. Cell *128*, 707–719.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009a). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, X., Li, L., Pandey, R., Byun, J.S., Gardner, K., Qin, Z., and Dou, Y. (2012). The Histone Acetyltransferase MOF Is a Key Regulator of the Embryonic Stem Cell Core Transcriptional Network. Cell Stem Cell *11*, 163–178.

Li, X., Wu, L., Corsa, C.A.S., Kunkel, S., and Dou, Y. (2009b). Two Mammalian MOF Complexes Regulate Transcription Activation by Distinct Mechanisms. Mol. Cell *36*, 290–301.

Liber, D., Domaschenz, R., Holmqvist, P.-H., Mazzarella, L., Georgiou, A., Leleu, M., Fisher, A.G., Labosky, P.A., and Dillon, N. (2010). Epigenetic Priming of a Pre-B Cell-Specific Enhancer through Binding of Sox2 and Foxd3 at the ESC Stage. Cell Stem Cell *7*, 114–126.

Lim, P.S., Hardy, K., Bunting, K.L., Ma, L., Peng, K., Chen, X., and Shannon, M.F. (2009). Defining the chromatin signature of inducible genes in T cells. Genome Biol. *10*, R107.

Lin, T., Chao, C., Saito, S., Mazur, S.J., Murphy, M.E., Appella, E., and Xu, Y. (2004). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. Nature Cell Biology *7*, 165–171.

Lin, W., Srajer, G., Evrard, Y.A., Phan, H.M., Furuta, Y., and Dent, S.Y.R. (2007). Developmental potential of Gcn5(-/-) embryonic stem cells in vivo and in vitro. Dev. Dyn. *236*, 1547–1557.

Liu, E.T., Pott, S., and Huss, M. (2010). Q&A: ChIP-seq technologies and the study of gene regulation. BMC Biology *8*, 56.

Liu, Y., Lu, C., Yang, Y., Fan, Y., Yang, R., Liu, C.-F., Korolev, N., and Nordenskiöld, L. (2011). Influence of histone tails and H4 tail acetylations on nucleosome-nucleosome interactions. J. Mol. Biol. *414*, 749–764.

Lombardi, P.M., Cole, K.E., Dowling, D.P., and Christianson, D.W. (2011). Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. Curr. Opin. Struct. Biol. *21*, 735–743.

Longo, V.D., and Kennedy, B.K. (2006). Sirtuins in aging and age-related disease. Cell *126*, 257–268.

Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell *49*, 203–210.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8A resolution. Nature *389*, 251–260.

Maas, S.A., Suzuki, T., and Fallon, J.F. (2011). Identification of spontaneous mutations within the long-range limb-specific Sonic hedgehog enhancer (ZRS) that alter Sonic hedgehog expression in the chicken limb mutants oligozeugodactyly and silkie breed. Dev. Dyn. *240*, 1212–1222.

Mack, G.S. (2010). To selectivity and beyond. Nature Biotechnology *28*, 1259–1266.

Malone, C.S., Omori, S.A., and Wall, R. (1997). Silencer elements controlling the B29 (Igbeta) promoter are neither promoter- nor cell-type-specific. Proc. Natl. Acad. Sci. U.S.A. *94*, 12314–12319.

Mann, R.K., and Grunstein, M. (1992). Histone H3 N-terminal mutations allow hyperactivation of the yeast GAL1 gene in vivo. EMBO J *11*, 3297–3306.

McCool, K.W., Xu, X., Singer, D.B., Murdoch, F.E., and Fritsch, M.K. (2007). The Role of Histone Acetylation in Regulating Early Gene Expression Patterns during Early Embryonic Stem Cell Differentiation. J. Biol. Chem. *282*, 6696–6706.

McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D., and Felsenfeld, G. (1981). A 200 base pair region at the 5′ end of the chicken adult β-globin gene is accessible to nuclease digestion. Cell *27*, 45–55.

Mendjan, S., and Akhtar, A. (2007). The right dose for every sex. Chromosoma *116*, 95–106.

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., et al. (2006). Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. Mol. Cell *21*, 811–823.

Mersfelder, E.L., and Parthun, M.R. (2006). The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. Nucl. Acids Res. *34*, 2653–2662.

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

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., De Laat, W., Spitz, F., and Duboule, D. (2011). A Regulatory Archipelago Controls Hox Genes Transcription in Digits. Cell *147*, 1132–1145.

Morales, V., Straub, T., Neumann, M.F., Mengus, G., Akhtar, A., and Becker, P.B. (2004). Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. EMBO J *23*, 2258–2268.

Morey, C., Kress, C., and Bickmore, W.A. (2009). Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. Genome Res. *19*, 1184–1194.

Mortimer, R.K., and Johnston, J.R. (1959). Life Span of Individual Yeast Cells. , Published Online: 20 June 1959; | Doi:10.1038/1831751a0 *183*, 1751–1752.

Nakano, T., Windrem, M., Zappavigna, V., and Goldman, S.A. (2005). Identification of a conserved 125 base-pair Hb9 enhancer that specifies gene expression to spinal motor neurons. Developmental Biology *283*, 474–485.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.

Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A., and Loraine, A.E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics *25*, 2730–2731.

Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature *416*, 103–107.

Nourani, A., Utley, R.T., Allard, S., and Côté, J. (2004). Recruitment of the NuA4 complex poises the PHO5 promoter for chromatin remodeling and activation. The EMBO Journal *23*, 2597–2607.

O'Connor, M.D., Kardel, M.D., Iosfina, I., Youssef, D., Lu, M., Li, M.M., Vercauteren, S., Nagy, A., and Eaves, C.J. (2008). Alkaline phosphatase-positive colony formation is a sensitive, specific, and quantitative indicator of undifferentiated human embryonic stem cells. Stem Cells *26*, 1109–1116.

O'Neill, L.P., Keohane, A.M., Lavender, J.S., McCabe, V., Heard, E., Avner, P., Brockdorff, N., and Turner, B.M. (1999). A developmental switch in H4 acetylation upstream of Xist plays a role in X chromosome inactivation. EMBO J. *18*, 2897–2907.

O'Neill, L.P., and Turner, B.M. (1995). Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. EMBO J *14*, 3946–3957.

O'Neill, L.P., and Turner, B.M. (2003). Immunoprecipitation of native chromatin: NChIP. Methods *31*, 76–82.

O'Neill, T.E., Roberge, M., and Bradbury, E.M. (1992). Nucleosome arrays inhibit both initiation and elongation of transcripts by bacteriophage T7 RNA polymerase. J. Mol. Biol. *223*, 67–78.

Ong, C.-T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet *12*, 283–293.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet *10*, 669–680.

Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N., et al. (2010). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucl. Acids Res. *38*, 4958–4969.

Peichel, C.L., Prabhakaran, B., and Vogt, T.F. (1997). The mouse Ulnaless mutation deregulates posterior HoxD gene expression and alters appendicular patterning. Development *124*, 3481–3492.

Pfister, S., Rea, S., Taipale, M., Mendrzyk, F., Straub, B., Ittrich, C., Thuerigen, O., Sinn, H.P., Akhtar, A., and Lichter, P. (2008). The histone acetyltransferase hMOF is frequently downregulated in primary breast carcinoma and medulloblastoma and constitutes a biomarker for clinical outcome in medulloblastoma. Int. J. Cancer *122*, 1207–1213.

Pogo, B.G., Allfrey, V.G., and Mirsky, A.E. (1966). RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. PNAS *55*, 805–812.

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W.A. (2012). Psip1/Ledgf p52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing. PLoS Genet *8*, e1002717.

Prestel, M., Feller, C., Straub, T., Mitlöhner, H., and Becker, P.B. (2010). The activation potential of MOF is constrained for dosage compensation. Mol. Cell *38*, 815–826.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Raja, S.J., Charapitsa, I., Conrad, T., Vaquerizas, J.M., Gebhardt, P., Holz, H., Kadlec, J., Fraterman, S., Luscombe, N.M., and Akhtar, A. (2010). The nonspecific lethal complex is a transcriptional regulator in Drosophila. Mol. Cell *38*, 827–841.

Rea, S., Xouri, G., and Akhtar, A. (2007). Males absent on the first (MOF): from flies to humans. Oncogene *26*, 5385–5394.

Riddle, R.D., Johnson, R.L., Laufer, E., and Tabin, C. (1993). Sonic hedgehog mediates the polarizing activity of the ZPA. Cell *75*, 1401–1416.

Robinson, P.J.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D. (2008a). 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. Journal of Molecular Biology *381*, 816–825.

Robinson, P.J.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D. (2008b). 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. Journal of Molecular Biology *381*, 816–825.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J. Biol. Chem. *273*, 5858–5868.

Roh, T., Ngau, W.C., Cui, K., Landsman, D., and Zhao, K. (2004). High-resolution genome-wide mapping of histone modifications. Nature Biotechnology *22*, 1013–1016.

Roh, T.-Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev. *19*, 542–552.

Roh, T.-Y., Wei, G., Farrell, C.M., and Zhao, K. (2007). Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Research *17*, 74–81.

Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone Acetyltransferases. Annual Review of Biochemistry *70*, 81–120.

Sabo, P.J., Humbert, R., Hawrylycz, M., Wallace, J.C., Dorschner, M.O., McArthur, M., and Stamatoyannopoulos, J.A. (2004). Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. PNAS *101*, 4537–4542.

Sachs, R.K., Van den Engh, G., Trask, B., Yokota, H., and Hearst, J.E. (1995). A random-walk/giant-loop model for interphase chromosomes. Proc. Natl. Acad. Sci. U.S.A. *92*, 2710–2714.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development *132*, 797–803.

Sauve, A.A., Wolberger, C., Schramm, V.L., and Boeke, J.D. (2006). The Biochemistry of Sirtuins. Annual Review of Biochemistry *75*, 435–465.

Schneider, R., Bannister, A.J., Weise, C., and Kouzarides, T. (2004). Direct binding of INHAT to H3 tails disrupted by modifications. J. Biol. Chem. *279*, 23859–23862.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell *132*, 887–898.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of Site-Specific Histone Acetylation and Deacetylation. Annual Review of Biochemistry *76*, 75–100.

Sharma, G.G., So, S., Gupta, A., Kumar, R., Cayrou, C., Avvakumov, N., Bhadra, U., Pandita, R.K., Porteus, M.H., Chen, D.J., et al. (2010). MOF and histone H4 acetylation at

lysine 16 are critical for DNA damage response and double-strand break repair. Mol. Cell. Biol. *30*, 3582–3595.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012a). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012b). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. Cell *119*, 941–953.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic Remodeling of Individual Nucleosomes Across a Eukaryotic Genome in Response to Transcriptional Perturbation. PLoS Biol *6*, e65.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. Science *311*, 844–847.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology *7*,.

Smith, E.R., Cayrou, C., Huang, R., Lane, W.S., Côté, J., and Lucchesi, J.C. (2005). A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. Mol. Cell. Biol *25*, 9175–9188.

Sobel, R.E., Cook, R.G., Perry, C.A., Annunziato, A.T., and Allis, C.D. (1995). Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. PNAS *92*, 1237–1241.

Solomon, M.J., Larsen, P.L., and Varshavsky, A. (1988). Mapping proteinDNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. Cell *53*, 937–947.

Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell *113*, 405–417.

Spitz, F., Gonzalez, F., Peichel, C., Vogt, T.F., Duboule, D., and Zákány, J. (2001). Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. Genes Dev. *15*, 2209–2214.

Spitz, F., Herkenne, C., Morris, M.A., and Duboule, D. (2005). Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. Nat. Genet. *37*, 889–893.

Straub, T., and Becker, P.B. (2007). Dosage compensation: the beginning and end of generalization. Nature Reviews Genetics *8*, 47–57.

Sykes, S.M., Mellert, H.S., Holbert, M.A., Li, K., Marmorstein, R., Lane, W.S., and McMahon, S.B. (2006). Acetylation of the p53 DNA-Binding Domain Regulates Apoptosis Induction. Molecular Cell *24*, 841–851.

Szutorisz, H., Canzonetta, C., Georgiou, A., Chow, C.-M., Tora, L., and Dillon, N. (2005). Formation of an Active Tissue-Specific Chromatin Domain Initiated by Epigenetic Marking at the Embryonic Stem Cell Stage. Mol. Cell. Biol. *25*, 1804–1820.

Taipale, M., Rea, S., Richter, K., Vilar, A., Lichter, P., Imhof, A., and Akhtar, A. (2005). hMOF Histone Acetyltransferase Is Required for Histone H4 Lysine 16 Acetylation in Mammalian Cells. Mol. Cell. Biol. *25*, 6798–6810.

Tamburini, B.A., and Tyler, J.K. (2005). Localized Histone Acetylation and Deacetylation Triggered by the Homologous Recombination Pathway of Double-Strand DNA Repair. Mol. Cell. Biol. *25*, 4903–4913.

Tashiro, K., Teissier, A., Kobayashi, N., Nakanishi, A., Sasaki, T., Yan, K., Tarabykin, V., Vigier, L., Sumiyama, K., Hirakawa, M., et al. (2011). A mammalian conserved element derived from SINE displays enhancer properties recapitulating Satb2 expression in early-born callosal projection neurons. PLoS ONE *6*, e28497.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A Mammalian Histone Deacetylase Related to the Yeast Transcriptional Regulator Rpd3p. Science *272*, 408–411.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell *60*, 909–920.

The UniProt Consortium (2011). Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Research *40*, D71–D75.

Thomas, T., Dixon, M.P., Kueh, A.J., and Voss, A.K. (2008). Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture. Mol. Cell. Biol *28*, 5093–5105.

Thomas, T., Loveland, K.L., and Voss, A.K. (2007). The genes coding for the MYST family histone acetyltransferases, Tip60 and Mof, are expressed at high levels during sperm development. Gene Expression Patterns *7*, 657–665.

Thorne, A.W., Kmiciek, D., Mitchelson, K., Sautiere, P., and Crane-Robinson, C. (1990). Patterns of histone acetylation. European Journal Of Biochemistry / FEBS *193*, 701–713.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tóth, K.F., Knoch, T.A., Wachsmuth, M., Frank-Stöhr, M., Stöhr, M., Bacher, C.P., Müller, G., and Rippe, K. (2004). Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin. J. Cell. Sci. *117*, 4277–4287.

Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urcia, R., Boisvert, F.-M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H., et al. (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. The Journal of Cell Biology *183*, 223–239.

Tsai, W.-W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.-Y., Shi, X., Schwarzer, D., Plunkett, W., et al. (2010). TRIM24 links a non-canonical histone signature to breast cancer. Nature *468*, 927–932.

Tse, C., Sera, T., Wolffe, A.P., and Hansen, J.C. (1998). Disruption of Higher-Order Folding by Core Histone Acetylation Dramatically Enhances Transcription of Nucleosomal Arrays by RNA Polymerase III. Mol Cell Biol. *18*, 4629–4638.

Turner, B. (2001). ChIP with Native Chromatin: Advantages and Problems Relative to Methods Using Cross-Linked Material.

Turner, B.M., Birley, A.J., and Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. Cell *69*, 375–384.

Turner, B.M., O'Neill, L.P., and Allan, I.M. (1989). Histone H4 acetylation in human cells Frequency of acetylation at different sites defined by immunolabeling with site-specific antibodies. FEBS Letters *253*, 141–145.

Vaquero, A., Scher, M.B., Lee, D.H., Sutton, A., Cheng, H.-L., Alt, F.W., Serrano, L., Sternglanz, R., and Reinberg, D. (2006). SirT2 is a histone deacetylase with preference for histone H4 Lys 16 during mitosis. Genes Dev. *20*, 1256–1261.

Vaquero, A., Sternglanz, R., and Reinberg, D. (2007). NAD+-dependent deacetylation of H4 lysine 16 by class III HDACs. Oncogene *26*, 5505–5520.

Vega, V.B., Cheung, E., Palanisamy, N., and Sung, W.-K. (2009). Inherent Signals in Sequencing-Based Chromatin-ImmunoPrecipitation Control Libraries. PLoS ONE *4*, e5241.

Vettese-Dadey, M., Grant, P.A., Hebbes, T.R., Crane- Robinson, C., Allis, C.D., and Workman, J.L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. EMBO J. *15*, 2508–2518.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell *138*, 1019–1031.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897–903.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science *193*, 848–856.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *advance online publication*,.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory

factor maintains the developmental potential of embryonic stem cells. , Published Online: 15 December 1988; | Doi:10.1038/336684a0 *336*, 684–687.

Williamson, I., Eskeland, R., Lettice, L.A., Hill, A.E., Boyle, S., Grimes, G.R., Hill, R.E., and Bickmore, W.A. (2012). Anterior-posterior differences in HoxD chromatin topology in limb development. Development *139*, 3157–3167.

Williamson, P., and Felsenfeld, G. (1978). Transcription of histone-covered T7 DNA by Escherichia coli RNA polymerase. Biochemistry *17*, 5695–5705.

Wong, L.-Y., Recht, J., and Laurent, B.C. (2006). Chromatin remodeling and repair of DNA double-strand breaks. J Mol Hist *37*, 261–269.

Wong, P., Hattangadi, S.M., Cheng, A.W., Frampton, G.M., Young, R.A., and Lodish, H.F. (2011). Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes. Blood *118*, e128–138.

Woo, Y.H., and Li, W.-H. (2012). Evolutionary conservation of histone modifications in mammals. Mol. Biol. Evol. *29*, 1757–1767.

Woodcock, C.L., and Dimitrov, S. (2001). Higher-order structure of chromatin and chromosomes. Current Opinion in Genetics & Development *11*, 130–135.

Wu, L., Zee, B.M., Wang, Y., Garcia, B.A., and Dou, Y. (2011). The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. Mol. Cell *43*, 132–144.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86–90.

Yap, K.L., and Zhou, M.-M. (2011). Structure and Mechanisms of Lysine Methylation Recognition by the Chromodomain in Gene Transcription. Biochemistry *50*, 1966–1980.

Yokota, H., Singer, M.J., Van den Engh, G.J., and Trask, B.J. (1997). Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. Chromosome Res. *5*, 157–166.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952 –1958.

Zee, B.M., Levin, R.S., Xu, B., LeRoy, G., Wingreen, N.S., and Garcia, B.A. (2010). In Vivo Residue-specific Histone Methylation Dynamics. J. Biol. Chem. *285*, 3341–3350.

Zeller, R., López-Ríos, J., and Zuniga, A. (2009). Vertebrate limb bud development: moving towards integrative analysis of organogenesis. Nature Reviews Genetics *10*, 845–858.

Zeng, L., and Zhou, M.-M. (2002). Bromodomain: an acetyl-lysine binding domain. FEBS Letters *513*, 124–128.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. *21*, 1273–1283.

Zhang, K., Williams, K.E., Huang, L., Yau, P., Siino, J.S., Bradbury, E.M., Jones, P.R., Minch, M.J., and Burlingame, A.L. (2002). Histone acetylation and deacetylation: identification of acetylation and methylation sites of HeLa histone H4 by mass spectrometry. Mol. Cell Proteomics *1*, 500–508.

Zhang, X., Yang, Z., Khan, S.I., Horton, J.R., Tamaru, H., Selker, E.U., and Cheng, X. (2003). Structural Basis for the Product Specificity of Histone Lysine Methyltransferases. Molecular Cell *12*, 177.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008a). Model-based Analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev. *15*, 2343–2360.

Zhang, Y., Shin, H., Song, J., Lei, Y., and Liu, X.S. (2008b). Identifying Positioned Nucleosomes with Epigenetic Marks in Human from ChIP-Seq. BMC Genomics *9*, 537.

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.-K., Shahab, A., Kuznetsov, V.A., et al. (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell *1*, 286–298.

Zhong, X., and Jin, Y. (2009). Critical Roles of Coactivator p300 in Mouse Embryonic Stem Cell Differentiation and Nanog Expression. J. Biol. Chem. *284*, 9168–9175.

Zhou, B.O., Wang, S.-S., Zhang, Y., Fu, X.-H., Dang, W., Lenzmeier, B.A., and Zhou, J.-Q. (2011). Histone H4 Lysine 12 Acetylation Regulates Telomeric Heterochromatin Plasticity in Saccharomyces cerevisiae. PLoS Genet *7*, e1001272.

Zhu, L., Gazin, C., Lawson, N., Pages, H., Lin, S., Lapointe, D., and Green, M. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237.

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. Cell *138*, 1122–1136.

Zollino, M., Orteschi, D., Murdolo, M., Lattante, S., Battaglia, D., Stefanini, C., Mercuri, E., Chiurazzi, P., Neri, G., and Marangi, G. (2012). Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. Nature Genetics *44*, 636–638.

Websites

Andrews, Simon. "Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data." *FastQC A Quality Control Tool for High Throughput Sequence Data*. Babraham Bioinformatics, 9 Sept. 2011. Web. 28 Nov. 2012. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

BACPAC Resources Centre. "BAC Clones Distribution Center - BACPAC Resources Center." *BAC Clones Distribution Center - BACPAC Resources Center*. Children's Hospital Oakland Research Institure, n.d. Web. 28 Nov. 2012. <http://bacpac.chori.org/>.

Gentleman, Robert, and Ross Ihaka. *The R Project for Statistical Computing*. Computer software. *The R Project for Statistical Computing*. Vers. 2.14.0. Comprehensive R Archive Network, 31 Oct. 2011. Web. 28 Nov. 2012. <http://www.r-project.org/>.

Kouzarides, Tony, and Andy Bannister. "Guide to Epigenetic Marks." *Guide to Epigenetic Marks*. Abcam, n.d. Web. 28 Nov. 2012. <http://www.abcam.com/index.html?pageconfig=resource&rid=11924>.

NCBI. "Gene Expression Omnibus (GEO)." *NCBI*. U.S. National Library of Medicine, n.d. Web. 28 Nov. 2012. <http://www.ncbi.nlm.nih.gov/geo/>.

Roadmap. "Roadmap Epigenomics Project." *Roadmap - Epigenomes*. Roadmap Epigenomics Project, n.d. Web. 29 Nov. 2012. <http://www.roadmapepigenomics.org/complete_epigenomes/>.

Saldanha, Alok. *Java TreeView Homepage*. Computer software. *Java TreeView Homepage*. Vers. 1.1.6r2. SouceForge.net, 30 Aug. 2011. Web. 28 Nov. 2012. <http://jtreeview.sourceforge.net/>.

Shin, Hyunjin, and Tao Liu. *Cis-regulatory Element Annotation System*. Computer software. *CEAS - Cis Regulatory Element Annotation*. Vers. 1.0.2. N.p., 11 May 2009. Web. 28 Nov. 2012. <http://liulab.dfci.harvard.edu/CEAS/index.html>.

Turner, Bryan. "ChIP with Native Chromatin: Advantages and Problems Relative to Methods Using Cross-Linked Material." *Mapping Protein/DNA Interactions by Cross-Linking [Internet]*. Paris: Institut National De La Santé Et De La Recherche Médicale, 2001. N. pag. *National Center for Biotechnology Information*. U.S. National Library of Medicine, 2001. Web. 28 Nov. 2012. <http://www.ncbi.nlm.nih.gov.ezproxy.webfeat.lib.ed.ac.uk/books/NBK7099/>.

Wall, Larry. *Perl*. Computer software. *The Perl Programming Language*. Vers. 5. N.p., 2002. Web. 16 Sept. 2010. <http://www.perl.org/>.

WCB Edinburgh. *Galaxy*. Computer software. *Galaxy*. WCB Edinburgh, n.d. Web. 28 Nov. 2012. <http://bifx-core.bio.ed.ac.uk/>.

WTSI, and EBI. "Ensembl Genome Browser." *Archive! Ensembl Version 61*. EBI, Feb. 2011. Web. 28 Mar. 2011. <http://feb2011.archive.ensembl.org/index.html>.

Appendix1: Loci covered by NimbleGen tiling arrays

Table Appendix 1.1: Genomic Regions covered by custom Hox Nimblegen tiling array. Coordinates used from UCSC genome browser, Mouse NCBI37, July 2007, mm9 build.

Locus	Base pairs	Coordinates
Aard	8426	chr15:51,869,406-51,877,831
Actb	5414	chr5:143,663,892-143,669,305
Adssl1	31878	chr12:113,852,987-113,884,864
Ankhd1	6329	chr18:36,715,293-36,721,621
Asl1	6000	chr10:86,953,114-86,959,113
Aurkb	9030	chr11:68,856,737-68,865,766
Brachyury	150001	chr17:8,554,691-8,704,691
Brd3	41055	chr2:27,296,495:27,337,549
Cdx2	12700	chr5:148,110,571-148,123,270
Chchd7	10373	chr4:3,862,488-3,872,860
Chmp4b	56637	chr2:154,473,323-154,529,959
c-myc	15063	chr15:61,811,875-61,826,937
Crabp1	12453	chr9:54,610,540-54,622,992
Ctnnal1	81381	chr4:56,810,244-56,891,624
Ddx6	48314	chr9:44,404,928-44,453,241
Dlk1	136468	chr12:110,693,533-110,830,000
Dnmt3b	16799	chr2:153,470,702-153,487,500
Emilin2	88185	chr17:71,586,819-71,675,003
Eno1	8210	chr4:149,609,995-149,618,204
Fgf10	115623	chr13:119,484,571-119,600,193
Fus	21830	chr7:135,107,354-135,129,183
Gata1	17112	chrX:7,532,108-7,549,219
Gata4	13904	chr14:63,855,878-63,869,781
Gnas	69619	chr2:174,105,382-174,175,000
Grem1	990000	chr2:112,862,000-113,856,000
Gsc	50001	chr12:105,687,160-105,737,160
Gypa	25112	chr8:83,013,758-83,038,869
H19	137587	chr7:149,730,414-149,868,000
Hand2	900000	chr8:59,050,000-59,951,000
Hba	237547	chr11:32,087,454-32,325,000
Hbb	197639	chr7:110,861,362-111,059,000
Hebp1	46046	chr6:135,079,862-135,125,907
Hemgn	15293	chr4:46,404,312-46,419,604
Hhex	40001	chr19:37,489,346-37,529,346
Hnrpa1	8303	chr15:103,068,649-103,076,951
HoxA	818585	chr6:51,781,416-52,600,000
HoxB	755001	chr11:95,845,224-96,600,224

HoxC	702398	chr15:102,400,000-103,102,397
HoxD	668990	chr2:74,241,352-74,910,341
Jarid1b	8076	chr1:136,453,744-136,461,819
Kcnq1	394001	chr7:150,256,000-150,650,000
Klf4	13200	chr4:55,536,078-55,549,277
Lin28	27001	chr4:133,553,000-133,580,000
Ly6a	6777	chr15:74,823,425-74,830,201
Ly6c1	5724	chr15:74,874,491-74,880,214
Magea3	8801	chrX:151,379,200-151,388,000
Mapk1ip1l	37167	chr14:47,911,795-47,948,961
Mest	225001	chr6:30,685,000-30,910,000
mirlet7-d	14164	chr13:48,630,830-48,644,993
Mlkl	39156	chr8:113,829,174-113,868,329
Myf5	6262	chr10:106,919,672-106,925,933
Nanog	10001	chr6:122,656,000-122,666,000
Nestin	28077	chr3:87,765,656-87,793,732
Nkx2-9	6000	chr12:57,709,722-57,715,721
Npm3	5490	chr19:45,820,805-45,826,294
Nrn1	13244	chr13:36,815,287-36,828,530
Olig2	7038	chr16:91,223,137-91,230,174
Pax6	35000	chr2:105,505,001-105,540,000
Pcdha	1012030	chr18:36,996,988-38,009,017
Pdlim2	20183	chr14:70,560,661-70,580,843
Pou5f1	9401	chr17:35,640,600-35,650,000
Ppm1b	5838	chr17:85,352,959-85,358,796
Ppp1cc	5664	chr5:122,604,605-122,610,268
Pramel4	15290	chr4:143,646,480-143,661,769
Prf1	9716	chr10:60,757,991-60,767,706
Psip1	36768	chr4:83,097,979-83,134,746
Rpl19	6164	chr11:97,886,000-97,892,163
Shh	1221001	chr5:28,489,000-29,710,000
Shmt2	11919	chr10:126,950,487-126,962,405
Sif1	13637	chr16:18,262,780-18,276,416
Slc7a3	9198	chrX:98,273,946-98,283,143
Snx17	12596	chr5:31,492,177-31,504,772
Sox2	23606	chr3:34,540,712-34,564,317
Srebf2	86538	chr15:81,963,276-82,049,813
Tcfap2a	27278	chr13:40,807,558-40,834,835
Tex19.1	8770	chr11:121,004,657-121,013,426
Tle1	4134	chr4:71,858,971-71,863,104
Tnik	6790	chr3:28,159,250-28,166,039
Xist	34292	chrX:100,649,996-100,684,287
Xlr4a	12245	chrX:70,317,643-70,329,887
Zbtb2	31535	chr10:5,953,177-5,984,711

Table Appendix 1.2: Genomic Regions covered by custom Limb Regulatory region Nimblegen tiling array. Coordinates used from UCSC genome browser, Mouse NCBI37, July 2007, mm9 build.

Locus	Base Pairs (Mbp)	Coordinates
ZRS	1.66	chr5:28,398,000-30,058,000
HoxD	1.76	chr2:73,733,000-75,498,000
Gremlin	0.99	chr2:112,862,000-113,856,000
Gli1	0.053	chr10:126,746,000-126,799,000
Gli2	0.99	chr1:120,420,000-121,408,000
Gli3	1.39	chr13:14,720,000-16,118,000
dHAND(Hand2)	0.9	chr8:59,050,000-59,951,000
Lmx1b	0.48	chr2:33,291,000-33,770,000
Tbx3/Tbx5	0.43	chr5:119,158,000-120,596,000
Sal1	1.53	chr8:91,251,000-92,780,000
Fgf10	1.55	chr13:118,650,000-120,200,000
HoxA	1.08	chr6:51,525,000-52,600,000
Etv4/Meox1	0.33	chr11:101,520,000-101,850,000
Etv5	0.26	chr16:22,245,000-22,500,000
Ets1	2.65	chr9:32,212,00-34,860,000
Alx4	0.425	chr2:93,250,000-93,675,000
Fgfr2	0.965	chr7:136,730,000-137,695,000
Sulf1	1.37	chr1:11,535,000-12,900,000
Ptch1	0.6	chr13:63,400,000-64,000,000
Twist1	1.24	chr12:34,000,000-35,240,000
Bapx1(Nkx3-2)	2.18	chr5:40,100,000-42,280,000