CHARACTERISTICS OF RANDOM MONOALLELIC GENE EXPRESSION DURING EMBRYONIC STEM CELL DIFFERENTIATION

A Dissertation Presented to
The Watson School of Biological Sciences
at Cold Spring Harbor Laboratory

in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

by

Mélanie Anne Eckersley-Maslin

Cold Spring Harbor Laboratory
October 2013

Acknowledgements

There are many people whom I would like to thank for their help and support, and without whom this thesis would not have been possible.

First and foremost, I would like to thank my supervisor David. He has been everything I was hoping for in a mentor and more, with continuous support and enthusiasm for my project, even when I had all but given up hope. I greatly appreciate the freedom he gave me to explore my own ideas, and inevitably learn from my own mistakes. His passion for science and dedication to his research group are qualities that I aspire to achieve if I ever have the chance to have a group of my own.

Next I would like to thank all of the past and present members of the Spector laboratory for all the lively discussions and helpful experimental advice. I would like to acknowledge in particular Megan who taught me all she knew about embryonic stem cell biology, and who was a invaluable friend in time of need; and Jan who has been a wonderful collaborator in the lab and who helped shaped my PhD journey, making me both a stronger person and better scientist.

This project would not have been possible without an amazing collaboration with David Thybert, John Marioni and Paul Flicek at EMBL-EBI, Cambridge, UK who provided all the necessary bioinformatic support. David has been a great collaborator who has taught me how computational scientists approach biological questions, and I appreciate all of the long Skype conversations in which we debated, clarified and discussed all aspects of the project from the statistics used to the biological implications of monoallelic expression.

Additionally, I would like to acknowledge all of the many individuals at Cold Spring Harbor Laboratory who have provided great advice, suggestions, reagents and experimental help. In particular, Jörg Drenkow and Tom Gingeras for help regarding RNA-sequencing library prepara-

tion, Emily Hodges and Greg Hannon for help with the DNA methylation studies, and members of Chris Vakoc's laboratory for help with the chromatin aspects of the project. Furthermore, I would like to thank the James building for a wonderful community to work in, and the past and present graduate students for a great support network.

I would also like to thank my thesis committee members: Greg Hannon my academic mentor, Tom Gingeras my thesis chair, as well as Chris Vakoc and Rob Martienssen, for all of their time and advice on my project and for pushing me at the right moment to do the genome-wide screen. I really enjoyed and appreciated the discussions we had during my committee meetings. I would also like to thank Marisa Bartolomei for agreeing to be my external examiner.

Many thanks are also due to the caring past and present staff of the Watson School of Biological Sciences, for making sure my PhD experience ran as smoothly as possible, and for making me feel like home immediately upon my arrival from Australia, and to the CSHL library and staff for providing a friendly and quiet environment to write. I would also like to thank my funding sources, the George A. and Marjorie H. Anderson Fellowship and the Genentech Fellowship, without which this journey would not have been possible.

Last but not least, I would like to thank my amazing friends, both at CSHL as well as those scattered all over the world, with whom I have shared many unforgettable experiences and countless laughs. You have all helped make the last five years an incredible and enjoyable experience. And finally to my family, who, although far away in time and space, have always supported me in chasing my dreams, and who have never stopped believing in me or in what I can achieve.

Abstract

Gene expression is not only tightly regulated during differentiation, forming the basis of lineage specification and cellular identity, but also flexible and stochastic, contributing to cellular heterogeneity within the organism. Monoallelic gene expression refers to the transcription of a gene from one of its two homologous alleles, and could potentially contribute to cellular diversity, although it has not been studied in a developmental context, nor characterised at the molecular level. An allele-specific RNA sequencing screen was performed in clonal populations of hybrid mouse embryonic stem cells (ESCs) and neural progenitor cells (NPCs) and identified 67 and 376 inheritable autosomal random monoallelically expressed genes respectively, a 5.6-fold increase upon differentiation. While DNA methylation and nuclear positioning did not play a role in maintaining monoallelic expression, specific histone modifications were differentially enriched between the two alleles and are thus likely involved in preserving the monoallelic state. Interestingly, transcript levels of 8% of the monoallelically expressed genes remained similar between monoallelic and biallelic NPC clones. These results support a model in which random monoallelic expression occurs stochastically during differentiation, and for some genes is compensated for by the cell to maintain the required transcriptional output of these genes. Therefore, monoallelic expression exemplifies the stochastic and plastic nature of gene expression in single cells during development.

Contents

List of Abbreviations

BER	Base Excision Repair
ChIP	Chromatin Immunoprecipitation
CNS	Central Nervous System
DMR	Differentially Methylated Region
DNMT	DNA Methyltransferase
ESC	Embryonic Stem Cell
FDR	False Discovery Rate
FISH	Fluorescence *In Situ* Hybridisation
GERP	Genomic Evolutionary Rate Profiling
ICR	Imprinting Control Region
Il	Interleukin
iPSC	induced Pluripotent Stem Cell
LAD	Lamin Associated Domain
LIF	Leukemia Inhibitory Factor
Lmna	Lamin A
MEF	Mouse Embryonic Fibroblast

NAD	Nucleolar Associated Domain
ncRNA	non-coding RNA
NK	Natural Killer cell
NP95	Nuclear Protein 95
NPC	Neural Progenitor Cell
NRPK	Normalised Reads Per Kilobase
NSC	Neural Stem Cell
OR	Olfactory Receptor
Pcdh	Protocadherin
PCR	Polymerase Chain Reaction
PGC	Primordial Germ Cell
PRC2	Polycomb Repressive Complex 2
Q-PCR	Quantitative Polymerase Chain Reaction
RA	Retinoic Acid
RNP	Ribonucleoprotein
RT	Reverse Transcriptase
SEM	Standard Error of the Mean
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
TAD	Topologically Associated Domain
XIC	X-Inactivation Center

List of Figures

List of Tables

Chapter 1

Introduction: gene expression regulation and monoallelic expression in diploid mammalian cells

1.1 Layers of gene regulation in mammalian cells: from nucleotide to nucleus

Gene expression is the process through which the information contained in the linear DNA sequence is transcribed into RNA. Critical to all forms of life, gene expression must be tightly regulated throughout the lifetime of an organism to maintain homeostasis, yet flexible enough to allow for responses to stimuli, ranging from developmental cell fate decisions to responses to the extracellular environment. The molecular processes by which transcription takes place in eukaryotes is well characterised biochemically: assembly of a pre-initiation complex at the promotor leads to transcription factor mediated recruitment of RNA polymerase, initiation, elongation and termination of RNA synthesis. However, what is of particular interest is how the process of gene expression is regulated such that a highly similar genomic sequence can give rise to hundreds of different cellular phenotypes, and ultimately the complexity of a multicellular organism, such as a mouse or man. The DNA itself, contains motifs which are essential for the proper recruitment of proteins and associated complexes to the DNA, however extra layers of

gene regulation exist on top of the linear double-stranded DNA sequence. These range from methylation of the cytosine nucleotide, to the composition and modifications of histone proteins which with DNA form the basic structural unit of chromatin, the nucleosome, to how the nucleosomes are further packaged into chromatin, and finally the interplay between different gene regions with each other and within the three dimensional nuclear space. In this way gene regulation in mammalian cells occurs in multiple layers, from the nucleotide to the nucleus.

1.1.1 DNA methylation

DNA methylation is perhaps one of the best studied and well characterised epigenetic modifications. Essential for mammalian development, DNA methylation is present throughout the plant, animal, fungal and prokaryotic kingdoms (Feng et al., 2010), with some notable exceptions including *Drosophila* and *C. Elegans* (Deaton and Bird, 2011). Cytosine methylation was put forward as a potential inheritable modification affecting gene regulation and cellular differentiation as early as the mid-1970s (Holliday and Pugh, 1975; Riggs, 1975). In mammals, DNA methylation occurs primarily at the symmetrical CpG dinucleotide (Riggs, 1975), although non-CpG methylation, primarily at CpA and to some extent CpT, can occur, particularly in ESCs in which up to 25% of DNA methylation can be non-CpG (Lister et al., 2009; Ramsahoye et al., 2000), as well as in the brain (Xie et al., 2012). The mammalian genome is generally CpG depleted: the human genome contains approximately 28 million CpGs, of which 10% occur in clusters known as CpG islands (Deaton and Bird, 2011). These islands are on average 1kb long, containing approximately 1 CpG per 10 base pairs, and are often unmethylated (Deaton and Bird, 2011). CpG islands are enriched over the transcription start sites of house keeping and developmental genes (Larsen et al., 1992), although half of CpG islands, known as orphan CpG islands, do not occur at annotated promoters, and may be more dynamically regulated through development (Illingworth et al., 2010; Smith and Meissner, 2013). While 60-80% of CpGs are methylated in somatic cells, CpG islands tend to be resistant to DNA methylation (Smith and Meissner, 2013). In contrast to the hypermethylated genome of somatic cells, primordial germ cells (PGCs) and blastocyst-stage embryos are hypomethylated (Seisenberger et al., 2012), a feature which is thought to be linked to their pluripotency.

DNA methylation is inherited stably across cell divisions through the action of the DNA

methylation maintenance machinery. The two key proteins involved are Nuclear Protein 95 (NP95, also known as UHRF1), which recognises specifically hemimethylated DNA (Bostick et al., 2007; Sharif et al., 2007); and DNA MethylTransferase 1 (DNMT1), which catalyses the transfer of a methyl group to the unmethylated cytosine from its substrate S-adenosyl methionine (SAM) (Bestor and Ingram, 1983; Gruenbaum et al., 1982). NP95 and DNMT1 form a complex with Proliferating Cell Nuclear Antigen (PCNA) (Chuang et al., 1997; Sharif et al., 2007), targeting the DNA methylation maintenance machinery to replicating DNA in S-phase, although PCNA is not necessary for DNMT1 loading (Spada et al., 2007). Interestingly, NP95 requires di- or tri-methylation of histone H3 lysine 9 (H3K9me2/3) to bind in S-phase (Rothbart et al., 2012), and DNMT1 binding is inhibited by trimethylation of histone H3 lysine 4 (H3K4me3) (Ooi et al., 2007; Otani et al., 2009), coupling DNA methylation to the chromatin state of the underlying DNA. Due to the low affinity and poor catalytic activity of DNMT1 at unmethylated DNA (Jeltsch, 2006; Song et al., 2011), *de novo* methylation is catalysed by DNA MethylTransferase 3a and 3b (DNMT3a, DNMT3b) enzymes (Okano et al., 1998), which are often found in complexes with epigenetic repressors (Smith and Meissner, 2013). Thus promoters which are targeted by histone deacetylases, H3K9 methyltransferases, including G9a, and chromatin remodelers such as LSH, undergo heterochromatin formation and subsequent DNA methylation to ensure long term silencing (Ayyanathan et al., 2003; Lehnertz et al., 2003). In this way, DNA methylation is thought to act secondary to gene inactivation as a lock to further stabilise the silent state (Smith and Meissner, 2013).

The mechanisms by which DNA demethylation takes place are still under investigation and likely depend on both sequence context as well as the stage of development (Seisenberger et al., 2013; Smith and Meissner, 2013). DNA demethylation can be passive and/or active (figure 1.1 on page 18). DNA methylation can be lost passively with cell divisions if the maintenance DNA methyltransferase Dnmt1 is not recruited to the DNA. This dilution of DNA methylation marks requires several cell divisions and consequently has slow kinetics. There are many proposed mechanisms by which active demethylation may occur. One is through the Base Excision Repair (BER) pathway, initiated by DNA glycosylases (Jost, 1993; Jost et al., 1995) which cleaves the bond between 5-methylcytosine base and the deoxyribose. Subsequent removal of the deoxyribose, filling of the gap by DNA polymerase and repair by DNA ligase, results in an

unmethylated cytosine in place of the methylated cytosine (Zhu, 2009). A slight modification of this pathway involves prior conversion of methylated cytosine to thymine by the Activation Induced Deaminase Aid, causing a mismatch in the DNA sequence which is recognised by the DNA glycosylases Tgd and Mbd4 (Cortellino et al., 2011), and subsequently excised via BER pathway as above. Alternatively, 5-methylcytosine can be oxidised by members of the ten-eleven-translocation (Tet) family of enzymes to 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009). 5hmC is either then passively lost during DNA replication, or further oxidised by the Tet enzymes to 5-formylcytosine and 5-carboxylcytosine which are then excised through the BER pathway (Branco et al., 2011).

While DNA methylation levels remain relatively constant in adult somatic tissues, there are two stages of development in which widespread waves of DNA demethylation occurs (figure 1.2 on page 20): during germ cell development, and in the pre-implantation embryo (reviewed in Seisenberger et al. 2013; Smith et al. 2013). Primordial Germ Cell (PGC) development starts at E7.25 in the epiblast, in which following an initial increase in DNA methylation levels similar to that of the surrounding somatic cells, the PGCs undergo a wave of demethylation, reestablishing their developmental potency (Guibert et al., 2012; Hajkova et al., 2002). Some DNA elements, such as imprinted Differentially Methylated Regions (DMRs), and the most active retrotransposons, such as the Intracisternal A particles (IAPs), are resistant to this wave of demethylation (Guibert et al., 2012; Lane et al., 2003; Seisenberger et al., 2012). Classically, it was believed that the demethylation occurs during a short time window from E11.5 to E13.5 at a similar time as imprint erasure (Hajkova et al., 2002), through the BER or TET pathways, however recently it has been shown that demethylation may commence as early as E8.5 and therefore passive demethylation is also a feasible mechanism by which DNA demethylation may occur (Seisenberger et al., 2012). Following the wave of demethylation, *de novo* DNA methylation commences around E14.5 to E16.5 through the action of Dnmt3a and Dnmt3b, providing germ cell fate specification and establishing sex-specific imprints (Davis et al., 1999, 2000; Kato et al., 2007; Ueda et al., 2000). The final level of DNA methylation in gametes differs between the sexes, with 85% methylation levels in sperm, compared to 30% in oocytes (Popp et al., 2010; Smallwood et al., 2011; Smith et al., 2012).

The second wave of demethylation occurs in the zygote following fertilisation. Here the

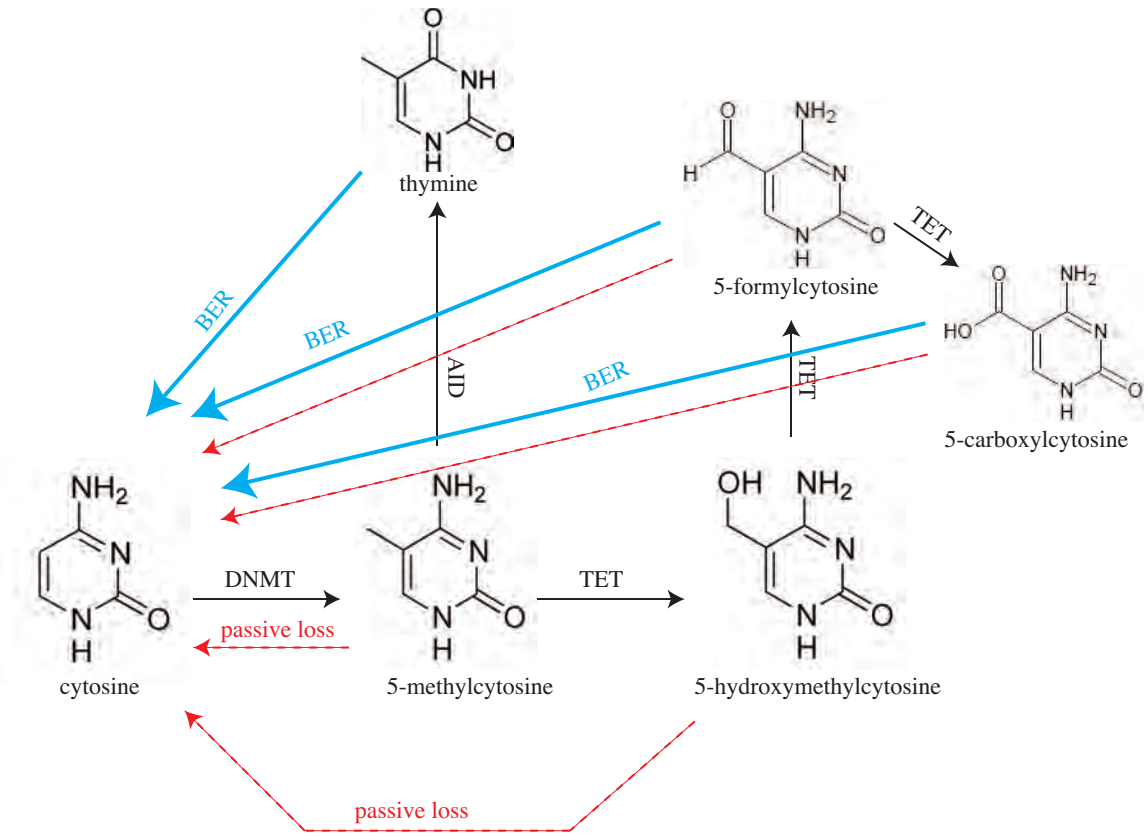


Figure 1.1: Mechanisms for removing cytosine methylation.
Cytosine is converted to 5-methylcytosine by DNA methyltransferases (DNMT). Removal of 5-methylcytosine can either occur passively (red dotted lines) or actively through either an AID-induced thymine intermediate or TET-induced 5-hydroxymethylcytosine, 5-formylcytosine or 5-carboxylcytosine intermediates, through either the Base Excision Repair (BER, blue line) pathway, or passively (red dotted lines).

two parental genomes show dramatically different demethylation kinetics: the paternal genome undergoes a rapid global demethylation which is largely completed prior to the first cell division (Santos et al., 2002; Wossidlo et al., 2010). The maternal genome, on the other hand, undergoes passive demethylation over several cell cycles, partially aided by the exclusion of Dnmt1 from the nucleus (Howell et al., 2001) and protection from the initial wave of demethylation that occurs at the paternal genome by the maternal factor Stella (Nakamura et al., 2007; Payer et al., 2003). At the blastocyst stage, the first lineage specifications occur. The outer trophectoderm cells remain largely unmethylated, where as the inner cell mass, which gives rise to the embryo proper and from which ESCs are derived, is once again largely hypermethylated (Seisenberger et al., 2013). Interestingly, ESCs cultured conventionally in serum are hypermethylated, while those grown in the two kinase inhibitor (2i) medium, believed to reflect the ground state of pluripotency (Ying et al., 2008), are hypomethylated (Habibi et al., 2013), consistent with 2i grown ESCs representing a more primal stage of development.

1.1.2 Histone modifications and variants.

Within the cell, DNA is wrapped around an octomer of histone proteins forming a complex known as a nucleosome. While this allows packaging of the 2 meters or so of DNA into a nucleus approximately 10μm in diameter, it also reduces the accessibility to DNA binding proteins such as transcription factors. To circumvent this, histone modifications, variants and nucleosome positioning alter the accessibility of the chromatin and consequently influence gene expression. The histone octomer is comprised of two of each of the core histone proteins: H3, H4, H2A and H2B. The proteins are encoded by large gene arrays and are highly expressed during S-phase, coupling chromatin assembly with DNA replication. The core histones can be replaced by non-canonical histone variants and/or undergo post-translational modifications including methylation, acetylation, phosphorylation and ubiquitylation. These variants and modifications correlate with different genetic elements and transcriptional activity, however whether they directly influence gene expression or if they are a secondary correlative effect is highly debated (Henikoff and Shilatifard, 2011; Rando, 2012).

The core histone proteins contain a characteristic structural motif known as a histone fold, containing 3 alpha-helices and 2 loops, and a long unstructured N-terminal tail. The tail, com-

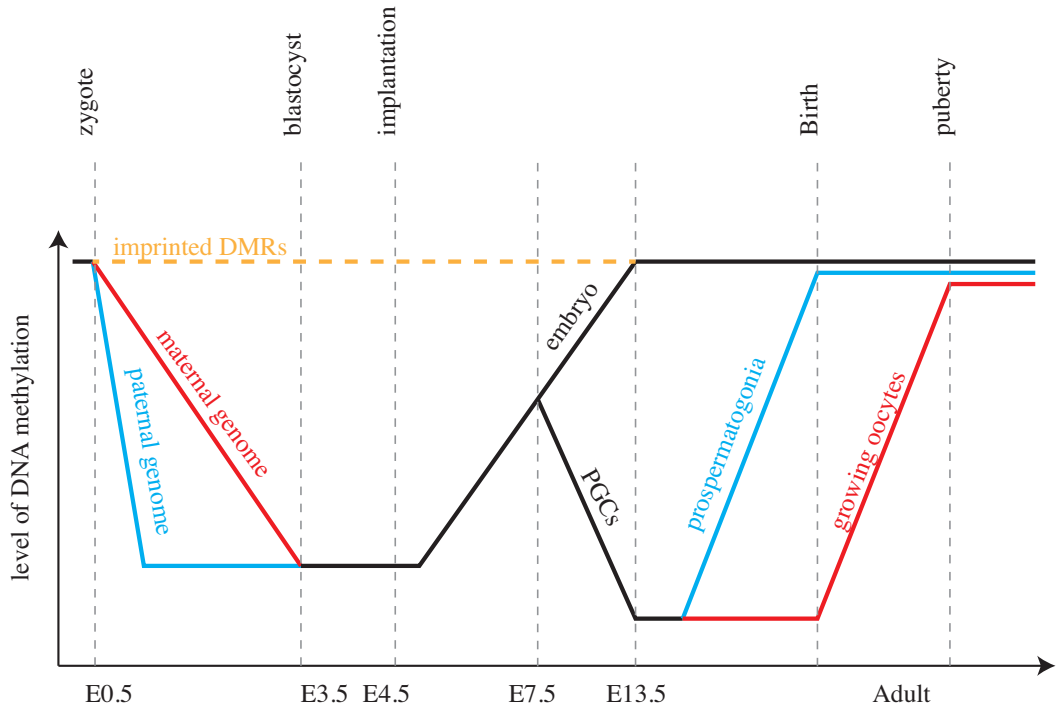


Figure 1.2: DNA methylation dynamics throughout development.
Upon fertilisation, the paternal genome (blue) undergoes rapid active demethylation, where as the maternal genome (red) undergoes passive demethylation. Imprinted Differentially Methylated Regions (DMR, orange) are resistant to this initial first wave of demethylation. Upon implantation and lineage specification, the embryonic genome becomes hypermethylated. At E7.5, Primordial Germ Cells (PGCs) undergo a second wave of demethylation. Remethylation of prospermatogonia (blue) occurs earlier in development than the growing oocytes (red). Adult somatic cells are hypermethylated. Modified from (Auclair and Weber, 2012).

prising up to 30% of the mass of the histone, protrudes out of the nucleosome (Luger et al., 1997), and is where the majority of the known post-translational modifications occur. Histone acetylation was the first modification discovered (Phillips, 1963) and shortly thereafter it was proposed that methylation of lysine residues in the histone protein could modify RNA synthesis (Allfrey and Mirsky, 1964; Murray, 1964). The modification of histone tails have been proposed to alter gene transcription in two ways. First, modifications may alter the biochemical properties of the nucleosome. For example, neutralisation of the highly positively charged histones through acetylation or phosphorylation could result in a decrease of electrostatic interactions between the histones and the negatively charged DNA, resulting in a decompacted and consequently more accessible state. Indeed point mutations at lysine residues in histone tails in budding yeast revealed that modifications of specific amino acid residues were mostly redundant, with the exception of histone H3 lysine 14, and thus likely act by increasing overall negative charge of the histones (Dion et al., 2005). Histone acetylation is therefore associated with transcription, DNA replication and DNA repair: all processes which require accessibility to the DNA (Zentner and Henikoff, 2013). However not all modifications, alter the overall charge of the histone. Some modifications, including phosphorylation, provide protein binding sites, directing the enzymatic activity of protein complexes to specific regions of the genome. In 2000, the first histone methyltransferase SUV39H1 was discovered (Rea et al., 2000), which lead to the discovery of additional lysine methyltransferases through homology searches of the catalytic SET domain. Many histone binding proteins contain different domains that recognise specific histone modifications including bromodomains (acetylated lysines), chromodomains (methylated histones); PHD domains (methylated lysines) and Tudor domains. It is also important to note that many of the histone modifying enzymes also have non-histone substrates, which can confound experiments designed to show the causality between histone modifications and gene expression regulation.

With the advent of ChIP-chip and ChIP-seq technologies, distribution maps of various histone modifications have been performed genome-wide across a multitude of cell types, exemplified by the efforts of the ENCODE consortium (ENCODE Project Consortium et al., 2012). This has led to the hypothesis of the histone code, in which a unique biological outcome could be specified by a particular combinations of histone modifications (Strahl and Allis, 2000). Although there is a clear correlation between certain genomic features and combina-

tions of histone modifications, the causality between the two has not been conclusively shown and is highly contested (Barth and Imhof, 2010; Henikoff and Shilatifard, 2011). Nonetheless, chromatin can be classified into anywhere between 6 and 51 distinct states (reviewed in (Bickmore and van Steensel, 2013)), based on the abundance of certain histone modifications, which themselves can be in turn correlated with biochemical properties of the chromatin including histone turnover, DNase I hypersensitivity and binding of transcription factors and chromatin remodelers. For example, active gene promotors are characterised by an abundance of histone H3 lysine 4 mono- di- and tri-methylation (H3K4me1/2/3) flanking the TSS and extending to a certain extent into the gene body (figure 1.3A on page 24). Throughout the gene body of active genes, the abundance of histone H3 lysine 79 methylation (H3K79me1/2/3), H3K9me1, H4K20me1, H2BK5me1 and H3K27me1 decreases, where as there is an accumulation of H3 lysine 36 trimethylation (H3K36me3) which peaks at the 3' end of the gene (figure 1.3A). Active promoters are also characterised by an increase in histone acetylation, including H3 lysine 9 acetylation (H3K9ac), which may act to help open the chromatin, providing accessibility to the transcriptional machinery. The transcriptional start site itself, is devoid of nucleosomes, and this nucleosome free region is often instead occupied by RNA polymerase. In contrast, inactive promoters are marked by methylation of histone H3 lysine 9 and/or 27 (H3K9me3, H3K27me3), and H4 lysine 20 (H4K20me3) which extends into the gene body (figure 1.3B). Of the two major inactive-associated modifications, H3K9me3 tends to be associated with constitutive heterochromatin, where as H3K27me3 tends to be more dynamic and associated with facultative heterochromatin. Interestingly, some promoters in ESCs contain both active H3K4me3 and inactive H3K27me3 marks. These regions, known as bivalent domains, are enriched over developmental gene promoters (Bernstein et al., 2006; Mikkelsen et al., 2007).

Repetitive elements and pericentromeric heterochromatin are marked by H3K9me2/3 and H4K20me3, where as subtelomeric regions contain H3K27me2/3, H3K9me2/3, and telomeres contain H3K4me3 and monomethylation of lysine 5 of H2B (H2BK5me1). Enhancer elements can be identified by a signature containing the p300 co-activator protein, acetylation of H3 lysine 27 (H3K27ac) and H3K4me1. There are over 100 different histone modifications described (Zentner and Henikoff, 2013), including ADP ribosylation, glycosylation ubiquitylation and sumoylation, which have half lives ranging from minutes to days (Barth and Imhof, 2010). The exact

roles for many of these modifications remain unclear, and the interplay, or histone crosstalk, between different marks will need to be determined. For example, binding of Heterochromatin Protein 1 (HP1) to H3K9me3 is inhibited by phosphorylation of serine 10 (H3S10ph) on the same histone tail (Fischle et al., 2005; Hirota et al., 2005), and ubiquitylation of lysine 123 of histone H2B (H2BK123Ub) promotes H3K4me3 and H3K79me3 deposition during transcription (Briggs et al., 2002). In this way, the exact functions of various histone modifications are likely to be complex and context dependent.

In addition to the extensive post-translational modifications of histones, nucleosomes have an added layer of complexity in that three of the four core histones have separately encoded histone variants. Unlike the core histones, the histone variants are typically expressed throughout the cell cycle and thus can be incorporated into chromatin in a replication-independent manner. There are several variants of histone H3: CENP-A is a key protein at the centromeres involved in centromere structure and stability. Histone H3.3, which differs from the core histones H3.1 and H3.2 by only a few amino acids, is a characteristic component of active genes and promotors due to its incorporation as a replacement histone following passage of RNA polymerase II. In this way, H3.3 levels can act as a read out of the dynamics of chromatin (Skene and Henikoff, 2013). Histone H2A also contains several variants. H2A.Z is perfectly positioned at the +1 and -1 nucleosomes flanking the transcriptional start site (Barski et al., 2007; Luk et al., 2010). However, its exact functions are unclear as H2A.Z is also associated with heterochromatin, transcriptional activation and repression, transcriptional elongation and chromosome segregation at mitosis (Skene and Henikoff, 2013). In contrast, the roles for H2A.X are better defined. Following DNA damage, H2A.X is incorporated and rapidly phosphorylated at it's C-terminal generating γH2A.X, and facilitating assembly of the DNA repair machinery at the site of DNA damage. Finally, macroH2A, a larger H2A variant, is found enriched on the inactive X-chromosome in female cells, as well as at regulatory elements of pluripotency genes upon differentiation (Pasque et al., 2012).

1.1.3 Nuclear structure and organisation

In addition to molecular factors, such as DNA methylation, histone modifications and variants, the regulation of gene expression can also be influenced by the organisation of the genome within

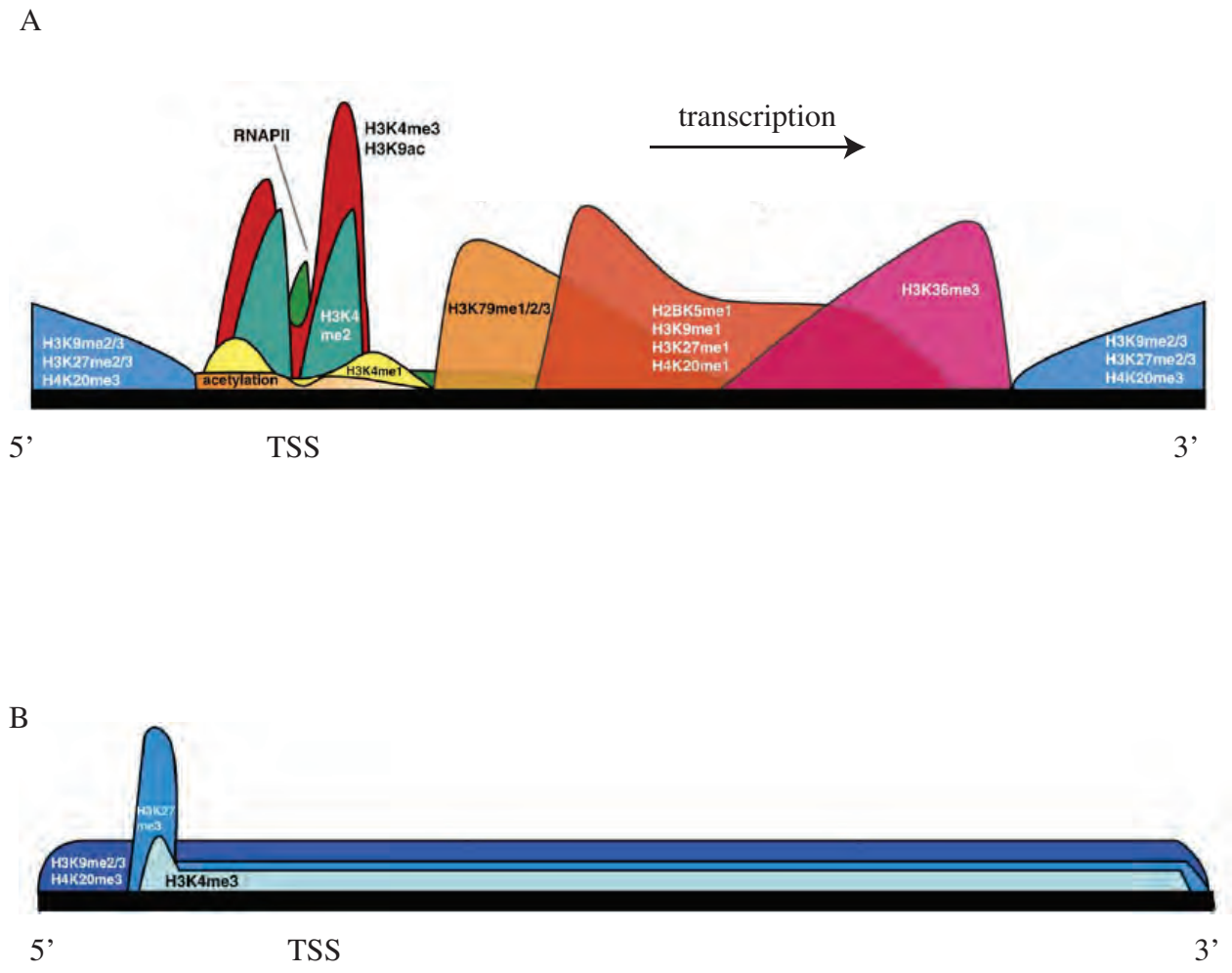


Figure 1.3: Typical distribution of histone modifications across active and inactive genes. Distribution of histone modifications across active (A) and inactive (B) genes. Modified from (Barth and Imhof, 2010).

the three dimensional nuclear space. Nucleosomes, containing 147bp of DNA and the histone octomer, are separated from each other by 10-80bp of DNA and associated linker histone H1. This gives rise to a 10nm fibre, resembling "beads on a string" (Olins and Olins, 1974) (figure 1.4 on the next page). The 10nm fibre has long been proposed to fold into a 30nm fibre (reviewed in (Grigoryev and Woodcock, 2012; Hübner et al., 2013)). However, while the 30nm fibre is readily observed *in vitro*, evidence for its existence *in vivo* is limited (Scheffer et al., 2011; Woodcock, 1994), suggesting that chromatin organisation above the 10nm fibre may not be the prevailing structural unit in living cells (Eltsov et al., 2008; Fussner et al., 2012; Nishino et al., 2012). Instead, nucleosomes are proposed to be arranged *in vivo* into disordered interdigitated globules (Eltsov et al., 2008; Nishino et al., 2012) which permit a more flexible and dynamic organisation of the genome than would be possible through a static ordered structures.

The development of chromosomal conformational capture (3-C) technologies has greatly improved our view of how chromatin is structured within the nuclear space (de Laat and Dekker, 2012). These methodologies involve formaldehyde-based cross-linking of associating regions of DNA within the nucleus, enzymatic digestion to reduce these interactors to the minimal regions, and ligation of the two interactors, followed by either PCR amplification or high-throughput sequencing. There are many variations of 3C (reviewed in (de Wit and de Laat, 2012)) which either investigate a specific region of interest (3C, 4C), or take a genome-wide view of all interactions (5C, Hi-C). This has enabled microscopy-based observations made over many decades at individual loci through DNA Fluorescence *In Situ* Hybridisation (FISH), to be generalised genome-wide. What is now apparent is that chromatin is structured into large fractal globules (Lieberman-Aiden et al., 2009), containing between 200kb to 1Mb of DNA, known as Topologically Associated Domains (TADs) (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). Regions of DNA that are linearly separated from each other by large distances physically associate with each other at a higher frequency than regions between TADs (figure 1.4). The globular structures maximise packing of chromatin, while enabling any region of interest to unfold and loop out of the globule to interact with other genomic regions or with protein complexes or nuclear domains (Lieberman-Aiden et al., 2009). The observation that while inter-TAD interactions change, the TADs themselves are stable and remain unchanged upon differentiation (Dixon et al., 2012; Nora et al., 2012), suggests that TADs may be defined by

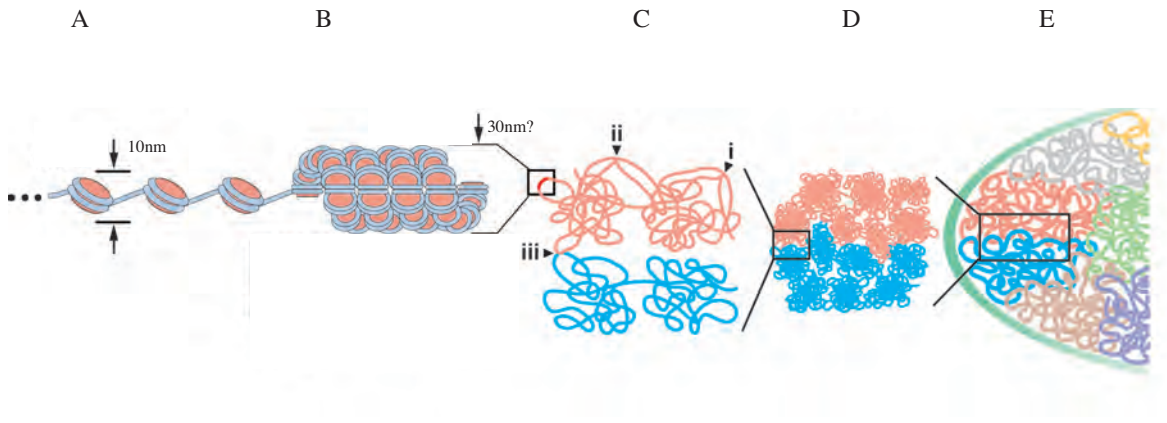


Figure 1.4: Organisation of chromatin in the mammalian nucleus.
(A) Chromatin is resolved as a 10 nm 'beads on a string' fiber consisting of nucleosomes. (B) Chromatin may form a 30 nm fiber or exist as a polymer melt, which is further organised into self-assembling fractal globules. (C) Chromatin fibers interact (i) within a fractal globule (frequent), (ii) between fractal globules of the same chromosome territory (rare), or (iii) between adjacent chromosome territories (very rare). (D) These fractal globules form chromosome territories, and fractal globules from adjacent chromosome territories can interdigitate. (E) Chromosomes are organized in chromosome territories. Modified from Hübner, Eckersley-Maslin and Spector, 2013.

genomic features rather than the transcriptional program of the cell. However, it is unclear what defines the sharp TAD boundaries: while they are enriched for CTCF binding, housekeeping gene promotors, transfer RNA genes and short interspersed element (SINE) retrotransposons (Dixon et al., 2012), these elements are not exclusive to TAD boundaries, suggesting that other factors are involved.

Interestingly, but perhaps not surprisingly, TADs overlap with other features of chromatin including DNA replication timing domains and Lamin Associated Domains (LADs, discussed below). The replication of DNA in S-phase is temporally segregated with large clusters of replication origins, several hundred kilobases to many megabases in size, firing nearly synchronously at specific stages of S-phase (reviewed in (Rhind and Gilbert, 2013)). For approximately half of the genome, the replication timing is dependent on cell type and usually linked to the transcriptional status of the genes in that region (Hansen et al., 2010; Ryba et al., 2010; Schwaiger et al., 2009). Early replicating domains are generally enriched for actively transcribing genes, where as heterochromatic and silent regions of the genome replicate towards the end of S-phase (Dimitrova and Gilbert, 1999; Hiratani et al., 2008). Interestingly some regions of the genome replicate asynchronously, and tend to be associated with monoallelically expressed genes (Donley et al., 2013; Kitsberg et al., 1993; Mostoslavsky et al., 2001). The boundaries of replication timing domains overlap with TAD boundaries, although the converse is not always true, as replication timing domains can contain several TADs (Dixon et al., 2012; Ryba et al., 2010). In this way, the DNA replication timing profile is linked to the underlying chromatin structure and state, providing a potential mechanism through which the transcriptional status of the region can be inherited, through differential incorporation of chromatin components at different stages of S-phase (Hiratani and Gilbert, 2009; McNairn and Gilbert, 2003).

Beyond TADs, the chromatin is further organised such that each chromosome occupies a distinct region of the nucleus, termed chromosome territories (figure 1.4). The concept of chromosome territories was first proposed over 100 years ago, by Carl Rabl in 1885 and Theodor Boveri in 1909 (reviewed in (Cremer and Cremer, 2006)), and in 1977 the first evidence for the existence of chromosome territories was published (Stack et al., 1977). In 1982, further evidence for chromosome territories was provided by experiments which followed the distribution of chromatin irradiated by UV in interphase through the cell cycle (Cremer et al., 1982).

Chromosome territories can be observed using FISH with whole chromosome composite painting probes (Guan et al., 1993), and are now supported by the genome-wide high resolution maps provided by the 3C-based technologies (Lieberman-Aiden et al., 2009). Chromosome paint combined with exon paint depicts a picture where the exonic sequences predominantly localise to the surface of the chromosome territory (Boyle et al., 2011), consistent with genes looping out of their chromosome territory and potentially interacting with regions of other chromosomes. Whether the genes move prior to activation, or if they require movement to specific regions to be transcribed has been a topic of immense debate (Hübner et al., 2013). The observed localisation of RNA polymerase II into foci, termed transcription factories (Jackson et al., 1993), in which numerous transcriptionally active genes localise, often from different chromosomes, set forth the hypothesis that genes move to these factories in order to be transcribed (Papantonis and Cook, 2010). However, recently high resolution live cell imaging of RNA polymerase II revealed that RNA polymerase II clusters form transiently with a life time of approximately 5 seconds (Cisse et al., 2013), arguing against stable transcription factories that genes must relocate to, and instead supporting a model whereby clusters of RNA polymerase II form at highly transcribed loci, which tend to be positioned outside of their chromosome territories and thus nearby other transcribed regions of the genome.

In higher eukaryotes, the interphase nucleus is a highly ordered yet dynamic structure, containing numerous specialised compartments (figure 1.5 on page 30), including the nuclear lamina, nuclear pores, nucleoli, nuclear speckles, Cajal bodies, PML bodies and paraspeckles (reviewed in (Mao et al., 2011; Spector, 2006; Zhao et al., 2009)). One domain in particular, the nuclear lamina, has been extensively studied in relation to its role in transcriptional regulation. The nuclear lamina is a meshwork of type V intermediate filament proteins called lamins, which together with various associated proteins lines the inner nuclear membrane and interacts with the nearby chromatin. Early electron microscopy studies first revealed a layer of heterochromatin at the periphery of interphase nuclei, as well as surrounding the nucleoli (reviewed in (Moses, 1956)). This perinuclear heterochromatin tends to replicate in late S-phase (O'Keefe et al., 1992). The recent development of DamID technology (Greil et al., 2006), which provides genome-wide DNA-nuclear lamina interaction maps, has revealed that genes that locate to the periphery are often, but not always, transcriptionally inactive, supporting previous observations

made at single loci by FISH based approaches (reviewed in (Takizawa et al., 2008b)). Mammalian genomes contain approximately 1,100-1,400 Lamin Associated Domains (LADs), ranging in size from 50kb to 10Mb (Peric-Hupkes and van Steensel, 2011). LADs contain approximately 2 fold less genes than inter-LADs, and typically lack RNA polymerase II and histone modifications associated with active transcription (Guelen et al., 2008; Peric-Hupkes et al., 2010; Pickersgill et al., 2006). Consistent with microscopy based studies (Kosak, 2002; Williams, 2006), changes in gene transcription during differentiation correlates with changes in the nuclear localisation of genes either towards or away from the nuclear lamina (Peric-Hupkes et al., 2010). Interestingly, ESCs, which are reported to have a distinctive flexible and dynamic nuclear structure compared to differentiated cells (Fisher and Fisher, 2011; Meshorer and Misteli, 2006), show a clear LAD structure and organisation, which differs in content by only 10% from differentiated cell types (Peric-Hupkes et al., 2010). Perhaps somewhat surprising at first, is the observation that LADs overlap with domains of repressive chromatin that interact with the nucleolus, so called Nucleolar Associated Domains (NADs) (Bickmore and van Steensel, 2013; Németh et al., 2010; van Koningsbruggen et al., 2010). The conundrum as to how chromatin domains, although similarly repressive, can be associated both with the nuclear lamina and the nucleolus, was solved recently by live-cell imaging of LADs in single cells (Kind et al., 2013). Not only did this study reveal that only approximately 30% of LADs were associated with the nuclear periphery in a given single cell, but also following mitosis there was not strict inheritance but stochastic shuffling of these LADs, such that only a subset remain at the nuclear periphery and others relocate to the nucleolar periphery (Kind et al., 2013). This demonstrates the heterogeneity of gene expression regulation in single cells, which is often masked by high-throughput genome wide analyses performed in populations of cells.

In addition to the nuclear lamina, the nucleus contains many other structures most of which are less understood. The nucleolus contains the clusters of repetitive ribosomal RNA genes which are highly transcribed by RNA polymerase I, accounting for up to 90% of RNA in the cell (McStay and Grummt, 2008). As mentioned above, the nucleolus is surrounded by transcriptionally inactive perinucleolar chromatin, which is enriched for satellite DNA as well as silent ribosomal DNA clusters (Sullivan et al., 2001). The Cajal body contains RNAs and proteins involved in the assembly and modification of small nuclear ribonucleoproteins (snRNPs) and

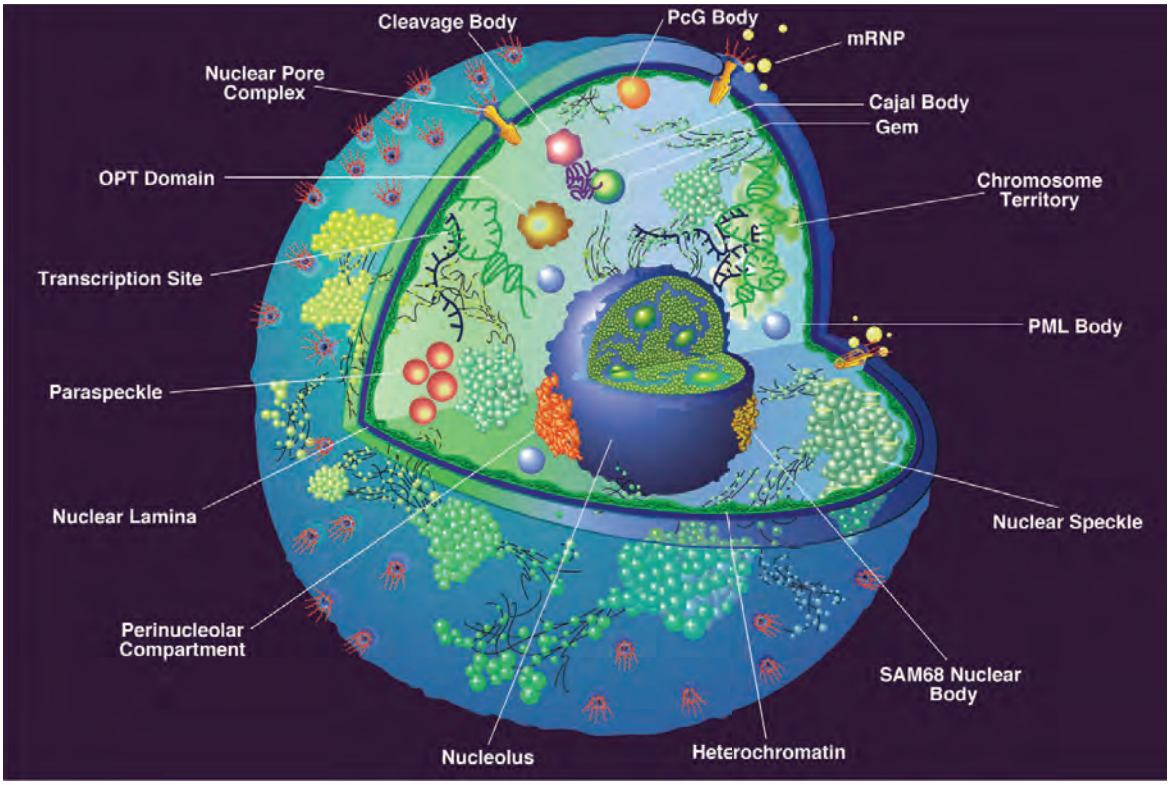


Figure 1.5: Structure of the interphase nucleus in higher eukaryotes.
Graphical representation of the 3-dimensional nucleus showing the different nuclear compartments and bodies. Figure from (Spector, 2001).

small nucleolar ribonucleoproteins (snoRNPs) (Nizami et al., 2010). The nucleus also contains between 30 to 50 nuclear speckles, also known as interchromatin granule clusters (IGCs), which are enriched with splicing factors including snRNPs and SR proteins (Lamond and Spector, 2003). Identified as a satellite to nuclear speckles, the paraspeckle is less abundant and its function poorly understood, although it is involved in retention of some A-to-I hyper-edited mRNAs (Mao et al., 2011). Polycomb bodies contain polycomb group proteins, and are associated with gene repression (Kerppola, 2009), where as ProMyelocytic Leukemia (PML) bodies are associated with gene rich and transcriptionally active chromatin, including the major histocompatibility complex and *p53* gene locus (Bernardi and Pandolfi, 2007). Correlations between gene positioning and their associations with these, as well as potentially other as of yet unidentified nuclear bodies, are abundant, however the main question remains: is nuclear positioning a cause or consequence of gene transcription?

When considering gene positioning relative to TADs, LADs, chromosome territories, or other nuclear structures, an important consideration is whether changes in gene positioning are a cause or consequence of changes in gene expression (figure 1.6 on the next page). For many examples, such as Hox gene movement from a H3K27me3 repressive domain to a permissive H3K4me3 domain upon gene activation (Noordermeer et al., 2011), it is unknown whether the structural changes that accompany gene activation are necessary for transcription to occur or if they are a secondary event stabilising the gene expression program in the cell. There are many more examples of gene movement correlating with gene expression changes (reviewed in (Hübner et al., 2013)) performed in static cells looking at before and after snapshots in time. What is required are live-cell systems, in which the activation and/or repression of endogenous alleles can be followed in single cells over time, with respect to their chromosome territories and/or nuclear structures. These types of experiments will be critical to extend our knowledge of nuclear organisation in transcriptional regulation.

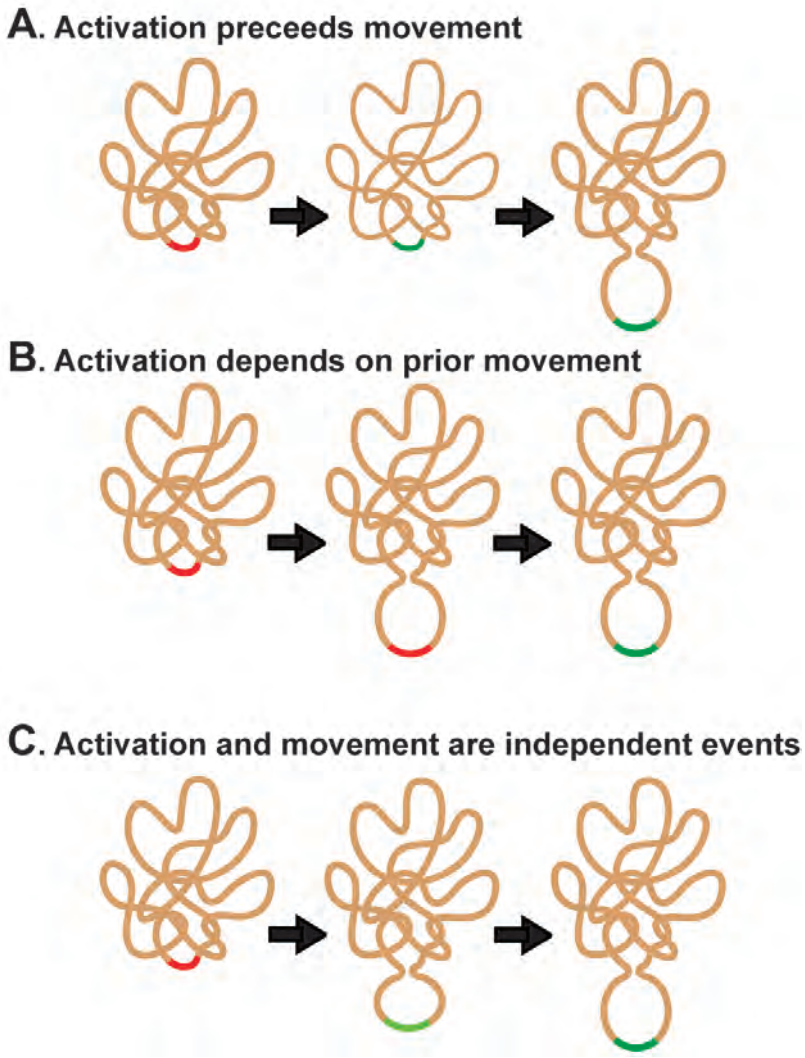


Figure 1.6: Gene movement relative to transcriptional activation.
(a) Transcriptional activation of a gene may precede its movement within the nucleus. (b) An inactive gene may get activated subsequent to its movement to a site that is favorable to transcriptional activation. (c) Transcriptional activation and gene movement may be independent of each other. Red, inactive gene; green, active gene. From Hübner, Eckersley-Maslin and Spector, 2013.

1.2 Classic examples of monoallelic gene expression in diploid cells

The majority of gene expression in diploid cells is carried out through expression of both alleles of each gene. However, several interesting cases of monoallelic expression, in which there is transcription from only one allele, have been documented over the last 60 years (figure 1.7 on the following page). Well characterized and extensively studied examples include X-chromosome inactivation, in which female mammalian cells randomly silence the vast majority of genes on one of the two X chromosomes to compensate for the increased dosage of X-linked genes (Guidi et al., 2004; Schulz and Heard, 2013), and genomic imprinting in which DNA methylation marks in the gametes regulate the expression of genes from either the maternal or paternal allele in the offspring (Bartolomei and Ferguson-Smith, 2011; McAnally and Yampolsky, 2010). Interestingly, random monoallelic expression can also occur on autosomes independently of the parental origin or underlying genotype (Chess, 2012b; Guo and Birchler, 1994). This has been classically studied for large gene families in which monoallelic expression generates receptor diversity and cellular identity. For example, the immune system utilizes monoallelic expression to ensure each B-cell expresses a single uniquely rearranged immunoglobulin receptor (Pernis et al., 1965). Similarly, in the nervous system, neurons express olfactory receptors in a monogenic and monoallelic manner to provide cell-identity and aid in neural connectivity (Chess et al., 1994).

1.2.1 X-inactivation

The first case reported, and one of the best understood examples, of monoallelic expression in mammalian cells is X-chromosome inactivation, an essential process which equalises the dosage of genes between the autosomes and sex chromosomes within an individual, as well as the dosage of X-linked genes between the sexes. It was first eluded to in 1949 (figure 1.7), when Murray Barr observed the presence of a dark subnuclear structure located adjacent to the nucleolus, specifically within the nucleus of female, but not male, cat neurons (Barr and Bertram, 1949). This 'nucleolar satellite', which was later renamed to the Barr body in honour of it's discoverer, was sufficient to determine the sex of the cell, and was later discovered to contain a single condensed X-chromosome (Ohno et al., 1959). Subsequently, Mary Lyon made the

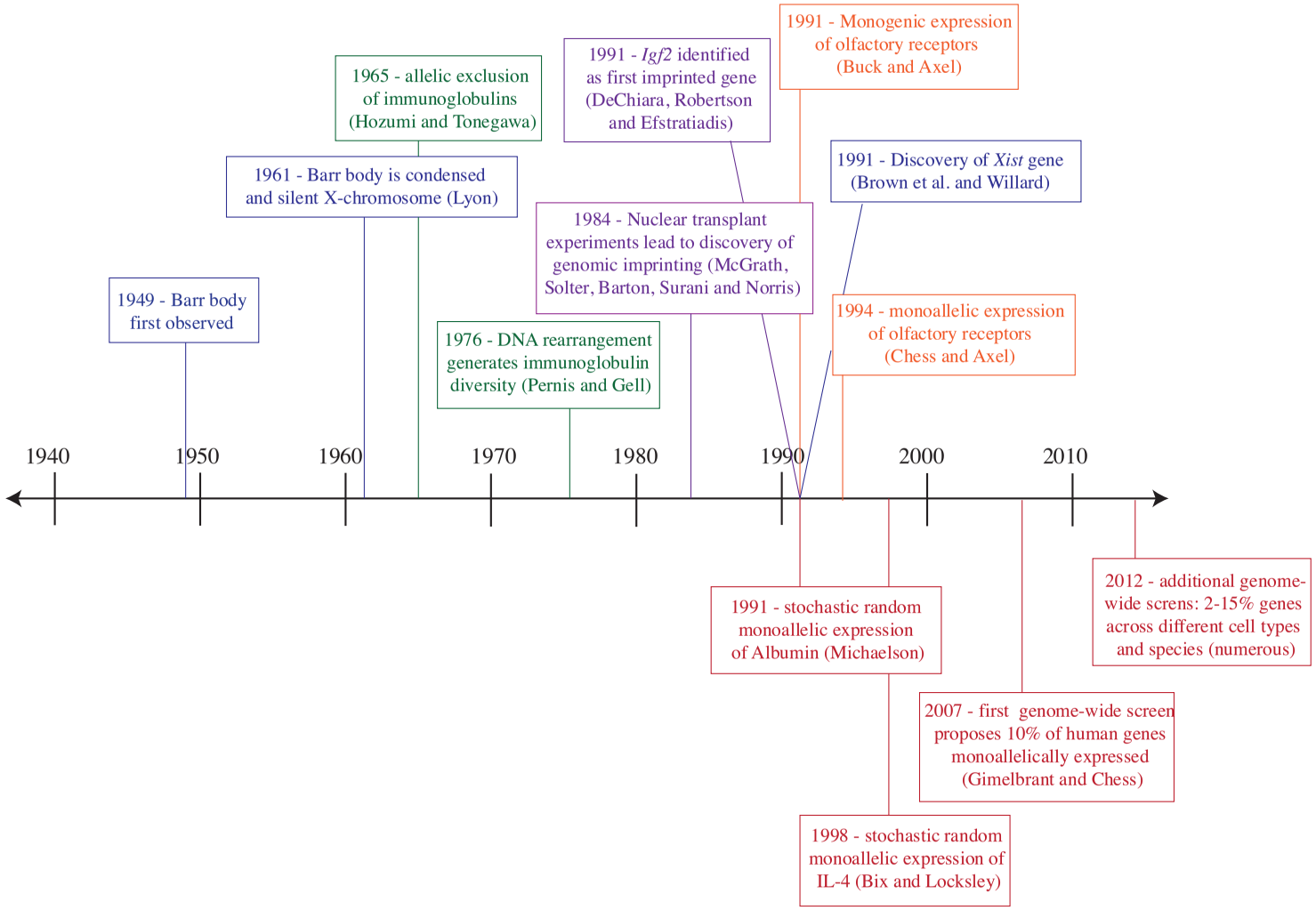


Figure 1.7: Time line depicting major initial discoveries of monoallelic expression.
Major discoveries made during the past 60 years towards our understanding of monoallelic expression. Blue denotes discoveries towards X-inactivation, green discoveries towards immunoglobulin allelic-exclusion, purple discoveries towards genomic imprinting, orange discoveries towards olfactory receptors and red towards random autosomal monoallelic expression.

discovery that the Barr body was in fact an entire condensed and silent X chromosome of either maternal or paternal origin (Lyon, 1961) (figure 1.7). Since this discovery over 50 years ago, there has been extensive investigation into the molecular mechanisms and functional consequences underlying X-chromosome inactivation. Different species have evolved distinct methods for dealing with the problem of sex chromosome dosage. In *Drosophila*, the expression level from the two X chromosomes in female cells is normal, and instead the single X-chromosome in male cells is up-regulated two-fold to match autosomal levels (Conrad and Akhtar, 2011). Similarly, the worm *Caenorhabditis elegans*, in contrast, increases expression levels two fold from the single X chromosome in male cells, where as the two X chromosomes in hermaphrodites remain expressed at normal levels (Meyer, 2010). In mammals, expression from the X-chromosome is increased two-fold in both males and females, however one of the two X-chromosomes is silenced in female cells, equalizing the dosage (Schulz and Heard, 2013). Marsupials show imprinted X-inactivation in which the paternally derived X-chromosome is always silenced (Sharman, 1971), where as in eutherian mammals the final choice of which X-chromosome to inactivate is randomly determined, except for in the placental tissue which remains paternally imprinted.

During development, X-chromosome inactivation in eutherian mammals undergoes cycles of silencing and reactivation (reviewed in Schulz and Heard, 2013). In the early embryo, at the 2-4 cell stage, the paternal X chromosome is inactivated. This persists until the blastocyst when cell lineages begin to be specified. In the trophectoderm, which gives rise to the placental tissue, the paternal X chromosome continues to remain inactive and does so throughout the development of this tissue. It remains unclear, however, whether paternal imprinting of the X-chromosome in placental tissue is limited to mice or if it also occurs in humans. In contrast, in the inner cell mass the paternal X chromosome is reactivated and *de novo* X-chromosome inactivation takes place, this time in a random manner so that either the paternal or the maternal X-chromosome is inactivated. Once established, the choice of inactive X-chromosome is maintained such that adult female mammals are mosaics.

The molecular mechanism of X-chromosome inactivation in eutherian mammals is centered around a DNA element on the X-chromosome termed the X Inactivation Center (XIC), which is necessary and sufficient to regulate inactivation (Brown et al., 1991). The XIC is involved in all stages of X-chromosome inactivation, including counting the number of X-chromosomes

in the cell, choice of which X-chromosome to inactivate, and initiation of silencing. The XIC contains a large non-coding RNA (ncRNA), called *X-Inactivation Specific Transcript* (*Xist*) which is expressed exclusively from the inactive X-chromosome. Deletion of this 17-20kb transcript results in the failure to silence the chromosome *in cis* (Marahrens et al., 1997; Penny et al., 1996), and ectopic placement on autosomes leads to silencing (Jiang et al., 2013; Lee et al., 1996; Migeon et al., 1999; Wutz et al., 2002), indicating that a central role for *Xist* in X-chromosome inactivation. Furthermore, placement of *Xist* on one of three copies of chromosome 21 in induced pluripotent stem (iPS) cells from Down's Syndrome patients, results in the complete inactivation of the corresponding chromosome 21 and restoration of normal diploid active chromosome dosage (Jiang et al., 2013), indicating that *Xist* ncRNA itself is sufficient to induce dosage compensation, and that the molecular mechanisms of X-inactivation can be applied to any chromosome. *Xist* acts by binding and recruiting Polycomb Repressive Complex 2 (PRC2) (Zhao et al., 2008), the protein complex involved in H3K27me3 mediated gene silencing (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). This occurs initially at a nucleation center, at which YY1 binding exclusively on the future inactive X chromosome bridges PRC2, *Xist* and the inactive X chromatin, allowing the spreading of inactive chromatin *in cis* along the entire X-chromosome (Pollex and Heard, 2012). Interestingly, not all genes on the X-chromosome are silenced. Approximately 3% of X-linked genes in mice and 15% of X-linked genes in human escape silencing. The molecular mechanisms for X-inactivation escape remain unclear. While *Xist* has a central role in X-chromosome inactivation, many other ncRNAs and protein complexes, including the antisense antagonistic ncRNA *Tsix* (Lee et al., 1999), have been identified and are implicated in different stages of X-chromosome inactivation. Furthermore, the nuclear organisation of the X-chromosomes appears to play a role, with the two X-chromosomes coming together in the nucleus at the onset of inactivation, potentially as a counting mechanism in the cell (Bacher et al., 2006; Masui et al., 2011). Interestingly, the *Xist* ncRNA is unique to eutherian mammals and evolved only in the last 150 million years from a protein coding gene Lnx3 (Duret et al., 2006), indicating that *Xist* mediated X-inactivation is restricted to the eutherian mammals. Marsupials do not have a corresponding gene, although they do have a different 27kb ncRNA called *Rsx* which also coats the inactive X-chromosome *in cis* (Grant et al., 2012). It remains to be determined if *Rsx* acts in a similar manner to *Xist* or

if an alternative mechanism for X-chromosome inactivation is involved. Nonetheless, ncRNAs appear to play a central role in both random and imprinted X-inactivation, a feature which is also shared by another class of monoallelically expressed genes: genomic imprinted genes.

1.2.2 Imprinting

Genomic imprinting refers to the exclusive expression of genes from either the maternal or paternal genome. The discovery of genomic imprinting in mammals came about following nuclear transplantation experiments in the early 1980s (Barton et al., 1984; McGrath and Solter, 1984; Surani et al., 1984), in which diploid androgenetic (containing two paternal pronuclei) and gynogenetic (containing two maternal pronuclei) embryos failed to develop, indicating that both the maternal and paternal genomes were required for embryonic development (figure 1.7). By studying different Robersonian translocations, the genomic regions responsible for these effects were mapped to distinct chromosomal regions on chromosomes, which are now known to contain clusters of imprinted genes (reviewed in (Cattanach, 1986; Solter, 1988)). There are approximately 150 known imprinted genes in mouse, and less in humans, although this may be because human development has not been as extensively studied. More than 80% of the known imprinted genes fall into 16 genomic clusters (Barlow, 2011). Each cluster spans from 80 to 4,000kb DNA (Barlow, 2011) and contains anywhere between 3 and 12 or more imprinted genes, including usually one non-coding RNA. With few exceptions, within each cluster the imprinted protein coding genes are mostly expressed from the same parental chromosome, where as the imprinted ncRNA is expressed from the alternate parental chromosome (Barlow, 2011). Many clusters also contain an Imprint Control Region (ICR), some of which also function as Differentially Methylated Regions (DMRs), although these two features are not necessarily exclusive (Lee and Bartolomei, 2013). A large proportion of imprinted genes encode for proteins involved in growth and metabolism (Lee and Bartolomei, 2013). This, along with the higher incidence of imprinting observed in placental tissue, suggests a role for imprinting in balancing the growth requirements of the fetus versus the energy demands on the mother. Supporting this, monotremes lack genomic imprinting and marsupials have fewer imprinted genes (Renfree et al., 2013), indicating a link between the evolution of imprinting and placentation and/or viviparity. Some genes are imprinted in a tissue-specific manner, most commonly in the placenta. One

report suggested that thousands of genes may be imprinted in the brain (Gregg et al., 2010), however these claims have been refuted based on technical and analytical errors in the original study (DeVeale et al., 2012).

There are two main models describing the molecular mechanisms regarding genomic imprinting: the insulator model; and the ncRNA model (figure 1.8 on the next page, reviewed in Lee and Bartolomei, 2013). The insulator model is evolutionarily older and conserved amongst the marsupials, and is exemplified by the *H19/Igf2* imprinted cluster (figure 1.8A), containing *Igf2*, the first imprinted gene discovered (DeChiara et al., 1991), and *H19*, the first long noncoding RNA identified (Brannan et al., 1990). In this cluster, allele-specific methylation at the ICR serves as an insulator element which is necessary for imprinting at the locus (Thorvaldsen et al., 1998). When unmethylated on the maternal allele, the ICR is bound by CTCF, preventing upstream enhancers from interacting with the downstream *Igf2* gene, rendering it inactive (Bell et al., 2010; Hark et al., 2000). In contrast, the paternal ICR is methylated, preventing CTCF from binding and thus allowing the enhancer elements to activate transcription of *Igf2* (Bell et al., 2010; Hark et al., 2000). At the maternal allele, the enhancers loop to the *H19* ncRNA resulting in its transcription, where as at the paternal allele, the methylation of the ICR spreads to silence the *H19* ncRNA (Lee and Bartolomei, 2013).

In contrast to the insulator model, the ncRNA model is more widely used model in eutherian mammals (Lee and Bartolomei, 2013), and was first described at the *Igf2r/Airn* locus (figure 1.8B). In this model, the DMR spans the promoter of the 108kb *Airn* ncRNA within the gene body of *Igf2r* (Lyle et al., 2000). When unmethylated on the paternal allele, *Airn* is transcribed, leading to silencing of the nearby *Slc22a2* and *Slc22a3* genes *in cis* (Pauler et al., 2012). Transcription of *Airn*, rather than the transcript itself, is required to silence *Igf2r* through preventing RNA polymerase II assembly on the overlapping *Igf2r* promoter (Latos et al., 2012). In contrast, at the maternal allele, the DMR is methylated, blocking transcription of *Airn*, allowing instead the expression of *Slc22a2*, *Slc22a3* and *Igf2r* genes (Pauler et al., 2012). The *Kcnq1* and *Snrpn* imprinted domains are also proposed to function through a ncRNA model. The insulator and ncRNA models are mutually exclusive and may both be involved in regulating imprinted gene clusters (Lee and Bartolomei, 2013).

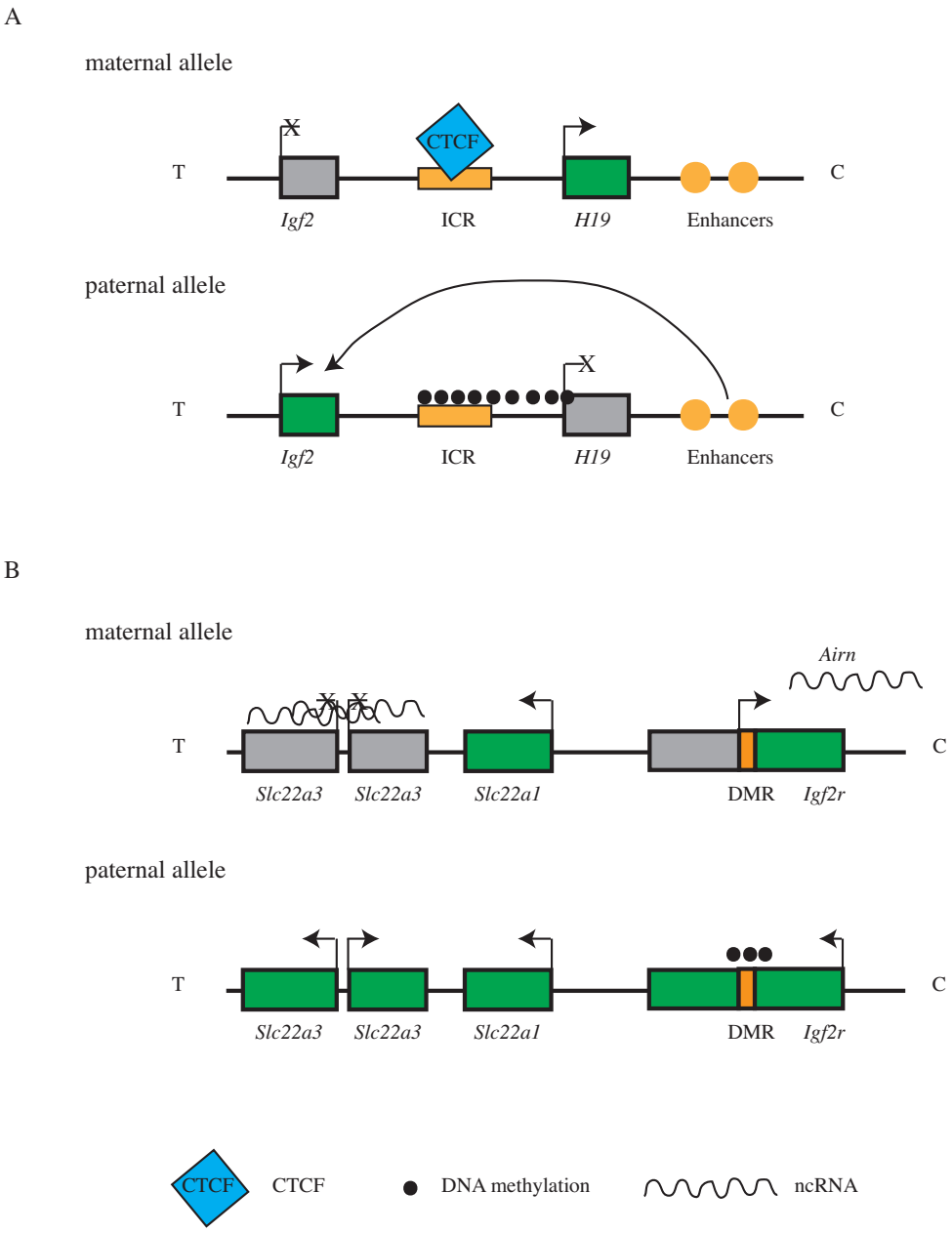


Figure 1.8: Insulator and ncRNA models of genomic imprinting.
(A) Insulator model for imprinting at the *H19*/*Igf2* locus. The maternal allele, with an unmethylated ICR is bound by CTCF, preventing enhancer activation of downstream *Igf2* gene. Conversely the methylated paternal ICR is unbound by CTCF, allowing enhancer activation of *Igf2*. (B) ncRNA model for imprinting at the *Igf2r* locus. The maternal allele with an unmethylated DMR transcribes the ncRNA *Airn* which prevents *Igf2r* transcription by blocking RNA polymerase II assembly, and coats the *Slc22a2* and *Slc22a3* genes. On the methylated paternal allele, *Airn* is not transcribed, allowing transcription of *Igf2r*. *Slc22a2* and *Slc22a3*. Green denotes transcribed genes, grey silent genes, and orange genomic regulatory elements. Modified from (Lee and Bartolomei, 2013).

1.2.3 Monoallelic Expression in the Immune System: Immunoglobulins

The immunoglobulins were the first autosomal genes shown to be monoallelically expressed. Experiments on rabbit leukocytes in the 1960s (figure 1.7) revealed that the products of different alleles were always present in different cells (Pernis et al., 1965), a discovery made prior to the isolation of the immunoglobulin genes themselves. One decade later, in 1976 (figure 1.7), Hozumi and Tonegawa discovered that this allelic exclusion was due to DNA rearrangement events at one of the two alleles (Hozumi and Tonegawa, 1976), a discovery which later won Tonegawa the Nobel prize. There are 7 distinct antigen receptor loci in the genome: the heavy chain locus (IgH); light chain loci (Igκ and Igλ) and T-cell receptors (TCR-α, -β, -γ and -δ); spread over 2.5Mb of the genome, all of which undergo DNA rearrangement and are monoallelically expressed in the respective B- and T-cells in the immune system. DNA rearrangement provides antigen receptor diversity which could not be encoded within the genome itself, and monoallelic expression ensures that each mature cell expresses one uniquely rearranged receptor, ensuring the specificity of the immune system. All antigen receptor loci contains variable (V), diversity (D) and joining (J) gene segments, each containing several segments which are recombined together to produce the variable region of the receptor. This rearrangement is lineage restricted and developmentally controlled, and is mediated in part by the DNA recombinase enzymes RAG1 and RAG2 (Oettinger et al., 1990; Schatz et al., 1989).

The immunoglobulin genes, like many other examples of monoallelic expression including the X-chromosome, imprinted genes and olfactory receptors, replicate asynchronously in S phase. This asynchronous DNA replication, in which one allele replicates in early S-phase and one in late S-phase, precedes and predicts the future allele which will be rearranged and expressed (Mostoslavsky et al., 2001). Thus, rather than a probabilistic model in which there is a random choice of alleles due to low probability of simultaneous efficient recombination, a deterministic or instructive model is favoured in which one allele is pre-marked by its earlier replication timing (Farago et al., 2012; Mostoslavsky et al., 2001). Upon initiation of recombination, the inactive allele associates with pericentromeric heterochromatin, whereas the early replicating allele looses its DNA methylation marks (Giambra et al., 2008) and gains H3 acetylation (Hesslein et al., 2003), resulting in a change in chromosome conformation rendering it accessible to RAG1 and

RAG2 (Chaumeil et al., 2013; Subrahmanyam and Sen, 2012). A subsequent maintenance phase involves a feedback loop which inhibits further rearrangement. In this way, the chromatin state and nuclear organisation of the antigen loci play a critical role in the establishment and maintenance of monoallelic expression (reviewed in (Cedar and Bergman, 2011; Subrahmanyam and Sen, 2012)).

Monoallelic expression occurs at another class of receptor genes in the immune system: the Ly49 Natural Killer (NK) cell receptors. These rodent-specific receptors are a class of C-type lectins which upon binding of host MHC class I molecules, thereby delivering inhibitory signals to the NK cell preventing it from killing the target cell. Humans, and other primates, do not have Ly49 receptors, instead the Killer cell Ig-like Receptor (*KIR*) genes are proposed to provide a similar function, and interestingly they are also similarly monoallelically expressed (Rouhi et al., 2006). The *Ly49* genes are encoded in a 1.4Mb cluster on chromosome 6 (Held et al., 1999; Takei et al., 2001). Each Ly49 receptor is only expressed in a subset of NK cells, however individual NK cells may express more than one Ly49 receptor, as detected by single cell RT-PCR analysis (Kubota et al., 1999). Expression of the Ly49 receptors is monoallelic (Held and Kunz, 1998; Held et al., 1995), and regulated in part by promoter DNA methylation and histone acetylation (Rouhi et al., 2006). While biallelic expression is also possible, it occurs at a much lower frequency equal to the product of the individual frequencies (Held and Raulet, 1997). The expression of alleles is determined independently for different Ly49 genes (Tanamachi et al., 2001), and the rate of co-expression of two different Ly49 receptors is equal to that which is expected if the two alleles are expressed independently (Valiante et al., 1997). This supports a model in which low probability independent stochastic activation of each allele of each receptor results in monoallelic expression of a small repertoire of receptors. In this way, monoallelic expression increases the number of distinct NK clones expressing different combination of Ly49 receptors (Held et al., 1999).

1.2.4 Generating Diversity in the Nervous System: Olfactory Receptors and Protocadherins

The olfactory receptors (OR) are the largest gene family in mammals, comprising approximately 1,400 functional genes in mice, or 350 in humans, arranged in 40 or more genomic clusters

(Young et al., 2002). In 1991 (figure 1.7) it was first proposed that the ORs were expressed in a monogenic manner in the main olfactory epithelium , and just a few years later (figure 1.7), the monoallelic nature of OR expression was discovered (Chess et al., 1994). The monoallelic and monogenic expression of the OR gene family ensures that each neuron expresses only one allele of one olfactory receptor, providing specificity in odor recognition, and correct axon guidance and wiring of the olfactory network. Since these original findings, extensive research into the molecular mechanisms of the monoallelic and monogenic nature of ORs has taken place over the last 20 years. Initially two models were proposed: deterministic and stochasticity (reviewed in (Magklara and Lomvardas, 2013)). In the determinism model, each of the 1,400 OR genes would be regulated by a unique combination of transcription factors acting at a unique *cis* regulatory sequence in their promoters. This was supported by the observation that ORs are expressed in 4 main regions of the olfactory epithelium, and each OR is expressed overall at the same frequency. However, experiments in which a reporter transgene placed under the control of an OR promoter never co-expressed with the endogenous OR, provided strong evidence against a deterministic model (Shykind, 2005). The alternative model, which is more widely accepted today, is that the choice of OR allele to be expressed occurs stochastically in the cell, and that feedback loops act to stabilise the initial OR allele choice and prevent additional allele activation.

How the initial selection and stabilisation of OR allele choice is made is only beginning to be understood at a molecular level. In 2004, somatic cell nuclear transfer experiments with post-mitotic olfactory neurons demonstrated that not only could post-mitotic cells re-enter the cell cycle, but the OR allele choice could be re-set in the cloned animal, indicating that irreversible DNA alterations did not take place (Eggan et al., 2004; Li et al., 2004). This was an important finding as it contrasted with, what was known at the time, the only other example of autosomal monoallelic expression, the immunoglobulin and T-cell receptor genes which undergo extensive DNA rearrangements. Next, it was proposed that the singularity of allele choice occurred through specific association with a 2kb enhancer element, *homology (H)*, on chromosome 14 adjacent to one of the OR gene clusters (Lomvardas et al., 2006; Serizawa et al., 2003). Chromosome conformation capture (3C) and RNA-DNA Fluorescence *In Situ* Hybridisation (FISH) experiments clearly demonstrated co-localisation of the active OR allele with the *H* element (Lomvardas et al., 2006). However, following experiments in which the *H* element was deleted

in mice, had no widespread effect on OR allele allele expression and affected only the adjacent OR genes (Fuss et al., 2007; Khan et al., 2011; Nishizumi et al., 2007). Thus association of a single allele with a locus control region is not the mechanism resulting in monoallelic expression of the OR genes. Instead, the current hypothesis is that the ORs are by default silent and present as distinct, compact and inaccessible heterochromatic macrodomains in the nucleus (Clowney et al., 2012), marked by the constitutive heterochromatic histone modifications H3K9me3 and H4K20me3 (Magklara et al., 2011). This nuclear aggregation is regulated, in part, by the absence of Lamin B1 Receptor (Lbr) protein as ectopic expression of Lbr causes OR gene disaggregation (Clowney et al., 2012). Stochastic activation of a single allele by lysine specific demethylase 1 (Lsd-1), results in the allele disassociating from the heterochromatic foci, and acquisition of the active histone modifications H3K4me3 (Magklara et al., 2011). Stabilisation of the allele choice is made through negative feedback loops, requiring expression of full length OR coding sequence (Fleischmann et al., 2008; Nguyen et al., 2007), in which OR induced activation of *Adcy3* leads to down-regulation of *Lsd-1*, preventing further OR allele activation (Lyons et al., 2013). The mechanisms of how a single OR allele, rather than 2 or 3 OR alleles, is initially activated is yet to be determined and will remain a future area of research in the coming years.

The clustered protocadherins are another example of a gene family which, through stochastic monoallelic expression, represents a candidate for generating neural diversity in vertebrates. Mammals contain approximately 50 protocadherin genes which are arranged into three tandem clusters on chromosome 18 in mouse: *Pcdh-α, Pcdh-β* and *Pcdh-γ.* Organisation of the *Pcdh-α* and *Pcdh-γ* clusters resemble that of the immunoglobulin and T-cell receptors, with tandem arrays of variable (V) regions followed by a set of three constant (C) regions. Each gene has a unique promoter and is expressed through splicing of the particular V exon to the C exons (Tasic et al., 2002; Wang et al., 2002). In contrast, the *Pcdh-β* cluster contains 22 members of unspliced single exon genes. Single-cell analysis of Purkinje cells first demonstrated that the *Pcdh-α* genes were expressed monoallelically and combinatorially (Esumi et al., 2005), such that each cell expresses approximately 2 different genes, one from each allele. The monoallelic and combinatorial expression was then extended to include genes in the *Pcdh-γ* (Kaneko et al., 2006) and *Pcdh-γ* (Hirano et al., 2012) clusters. As the protocadherins form heteromultimeric protein tetramers at the cell membrane (Yagi, 2008), the estimated diversity of combinatorial

co-expression of the three clusters would be more than 20 million combinations, more than sufficient to distinguish the 14 million cerebellar Purkinje cells in humans, analogous to the *Drosophila* Dscam1 proteins (Chen and Maniatis, 2013; Hirano et al., 2012). Regulation of the *Pcdh-α* cluster involves long range cis-regulatory enhancer elements, which compete with the individual variable exon promotors (Ribich et al., 2006), and similar DNA elements have been identified for the *Pcdh-β* (Yokota et al., 2011) and *Pcdh-γ* clusters (Guo et al., 2012). CTCF binding sites are present in the promoters of the *Pcdh-α* and *Pcdh-γ* clusters (Kim et al., 2007; Xie et al., 2007) and CTCF has been implicated in mediating the enhancer activity of the upstream cis-regulatory region of the *Pcdh-α* cluster (Kehayova et al., 2011). The current model suggest that CTCF and/or cohesin mediated DNA-looping interactions of the protocadherin promoters to the enhancers provides a means through which stochastic monoallelic expression of different isoforms can occur in each cell (Chen and Maniatis, 2013; Guo et al., 2012).

1.3 Random autosomal monoallelic gene expression

In addition to the classic examples described in the previous section, monoallelic expression has also been observed for autosomal genes that do not fall into any large clustered gene families. This phenomenon of random autosomal monoallelic gene expression was initially described at isolated examples, however in the last 7 years genome-wide studies have demonstrated that up to 10% of the genome may be monoallelically expressed in a given cell type (figure 1.7). Despite the identification of random monoallelically expressed genes, many unanswered questions remain regarding the molecular mechanisms of establishing and maintaining the allelic imbalance, and importantly, why have monoallelic gene expression for these genes at all.

1.3.1 Identification of autosomal monoallelic expression: from isolated examples to genome wide studies

Perhaps the first example of random monoallelic expression of an autosomal gene outside of the nervous and immune systems described, is that of albumin in hepatocytes (Michaelson, 1991, 1993). Immunofluorescence with sera raised against one of the two allelic forms of albumin, *Alb-1*^{c} provided direct evidence that most of the cells expressing albumin in the liver of heterozygous

Alb-1^{a}x*Alb-1*^{c} mice are in fact expressing only one of the two alleles (Michaelson, 1991, 1993). Moreover, clusters of multiple adjacent cells showed clonal expression of either allele. It was proposed that this cellular heterogeneity was due to a low probability of stochastic independent activation of the alleles, which once established is inheritable (Michaelson, 1993). The stochastic nature of gene expression was proposed to account for variations seen in other systems, including coat colour variegation of autosomal genes, for example at the pink eye dilution locus, and the globin genes (Michaelson, 1987; Searle, 1968; Silvers, 1958). Several years later, another example of low probability stochastic gene activation resulting in random monoallelic expression was described at the interleukin-2 (*Il-2*) locus in CD4+ T-cells (Hollander et al., 1998). Single cell analysis in heterozygous knock-out mice revealed that half the cells produced Il-2, where as the other half did not, and random monoallelic expression was confirmed by single-cell RT-PCR analysis in F1 hybrid wild-type mice (Hollander et al., 1998). Similarly, monoallelic expression has been reported for interleukin-4 (*Il-4*) (Bix and Locksley, 1998), as well as *Il-3*, *Il-5* and *Il-13* (Kelly and Locksley, 2000) and *Il-10* (Calado et al., 2006). Again, the monoallelic expression of the cytokines is likely due to stochastic low probability of expression from each allele, resulting in a mixed population of cells with either 0, 1 or 2 expressing alleles (Guo et al., 2005). In this case, it may not be monoallelic expression that is selected for, but the resulting low frequency of expressing cells that can be readily expanded in response to exogenous stimuli (Paixão et al., 2007).

Stochastic allele-specific expression was next extended to the *Pax5* transcription factor, which was initially reported to be monoallelically expressed in early progenitors and mature B cells, but biallelically expressed in immature B cells (Nutt et al., 1999). However, unlike the cytokines which exhibit asynchronous DNA replication timing (Ensminger and Chess, 2004; Hollander et al., 1998), *Pax5* replicated synchronously (Nutt et al., 1999). Furthermore, monoallelic expression of *Pax5*, was not strictly inheritable, with switching of alleles occurring between 1 and 4 weeks of culturing. A follow-up study using sensitive single-cell RT-PCR analysis, as well as FISH, confirmed monoallelic expression for the interleukins, but rejected it for *Pax5* (Rhoades et al., 2000), suggesting that the observed monoallelic expression was due to the failure to adequately detect both alleles of this low abundant transcript. Additional examples of random monoallelic expression of autosomal genes in a wide range of cell types continued to

emerge: the cell adhesion molecule *p120 catenin*, identified by its asynchronous DNA replication timing profile, was reported to be monoallelically expressed in both ES cells and fibroblasts (Gimelbrant et al., 2004); the glial fibrillary acidic protein *Gfap* was shown to be monoallelically expressed in cortical astrocytes (Takizawa et al., 2008a); *Igf2bp1*, a zinc finger protein required for imprinted expression of the *Igf2* gene, was reported to be monoallelically expressed in B cells (Thomas et al., 2011); and recently *Cubilin*, an endocytic receptor, was shown to be monoallelically expressed in renal proximal tubules and small intestine (Aseem et al., 2013). What is apparent from all of these isolated examples, is that a wide range of different types of genes can be monoallelically expressed in a wide range of different cell types. Interestingly, in essentially all examples, not all cells exhibit monoallelic expression, in contrast to what is observed for the classical examples of monoallelic expression including X-chromosome inactivation and olfactory receptor expression. Furthermore, the monoallelic expression patterns are inheritable, such that once established, the expression patterns are maintained across successive cell generations.

Depending on the organism, there can be anywhere between 100 to 10,000 copies of ribosomal RNA genes encoded in the genome. Mammals have approximately 400 copies of ribosomal RNA genes arranged into 5 chromosomal clusters in humans (6 in mice) (McStay and Grummt, 2008). Despite the large abundance of ribosomal RNA in the cell, not all the genes are transcribed by RNA polymerase I (Grummt, 2007). Instead, the genes exist in two states that are stably propagated through the cell cycle: an active open state that is transcribed; and an inactive nucleosomal state (Conconi et al., 1989). In the mouse pre-implantation embryo, the rRNA loci are biallelically expressed, which then, upon further differentiation, switch to monoallelic expression, effectively reducing the rRNA transcript levels in the cell by half (Schlesinger et al., 2009). Monoallelic expression is proceeded by asynchronous DNA replication timing (Li et al., 2005; Schlesinger et al., 2009), and requires differential DNA methylation at a specific CpG dinucleotide (Bird et al., 1981; Santoro and Grummt, 2001), histone modifications (Lawrence and Pikaard, 2004; Santoro et al., 2002; Zhou et al., 2002) and noncoding RNAs (Mayer et al., 2008, 2006) to maintain the differential expression between the alleles (reviewed in (McStay and Grummt, 2008)). It has been proposed that monoallelic expression ensures 50% of the alleles are expressed, thus providing a mechanism by which the early pre-implantation embryo can increase transcription to meet its higher energy demands (Schlesinger et al., 2009).

The development of high-throughput screening technologies was fundamental in enabling random autosomal monoallelic expression to be assessed at a genome-wide level. In 2007, the first genome-wide analysis of monoallelic expression was performed in clonal populations of human lymphoblast cell lines, as well as WI38 fibroblasts and placental tissue, by hybridising double-stranded cDNA to SNP arrays to determine allele-specific expression (Gimelbrant et al., 2007). Of the 3939 assessable genes in lymphoblasts, 85 (2.2%) were monoallelically expressed based on two or more informative SNPs, and another 286 (7.3%) based on a single informative SNP (Gimelbrant et al., 2007). Around the same time, a second group used micro-array based assays to detect simultaneous methylated and unmethylated DNA of a 65Mb region of mouse chromosome 7 in the CNS (Wang et al., 2007). In addition to the olfactory receptors, 5 of the 500 assessed genes showed simultaneous methylation and unmethylation and were confirmed as monoallelically expressed (Wang et al., 2007). Interestingly, one of the genes identified, *p* (*pink-eyed dilution*), was predicted to be monoallelically expressed 2 decades earlier (Michaelson, 1987). A follow-up study by the same group assessing dual DNA methylation patterns genome-wide of the entire mouse brain and identified 2,237 of 23,393 assessed genes (9.6%) marked by simultaneous DNA methylation and unmethylation, of which 12.4% (9 out of 122 assessed) were monoallelically expressed in clonal neural stem cells (Wang et al., 2010). The remaining genes likely displayed the dual DNA methylation pattern as a result of cellular heterogeneity in the brain. However, both the pilot and genome-wide screen suggest that 1-2% of genes expressed in the CNS are monoallelically expressed (Wang et al., 2010, 2007). The lower percentage of genes reported as monoallelically expressed in the mouse CNS compared to human cells (Gimelbrant et al., 2007), could be due to the assumption made that monoallelic expression is marked by allele-specific DNA methylation marks, which likely does not hold for all monoallelically expressed genes, or differences in the criteria used to define monoallelic expression.

Very recently, four additional genome-wide assessments of monoallelic gene expression have been reported in both mouse (Li et al., 2012b; Zwemer et al., 2012) and human (Jeffries et al., 2012; Lin et al., 2012) cells. The first followed a similar approach to the original genome-wide screen performed in human lymphoblasts (Gimelbrant et al., 2007), again using SNP arrays to detect monoallelic expression in clonal populations of mouse lymphoblasts and fibroblasts (Zwemer et al., 2012). Similar to the human, 212 of 1,358 assessed genes (15.6%) were reported

as monoallelically expressed in lymphoblasts, however only 20 (1.5% assessed genes) contained more than 1 informative SNP (Zwemer et al., 2012). A second study, again using array-based technology, reported significantly less monoallelic expression in clonal human neural stem cell lines derived from the cerebral cortex, striatum and spinal cord (Jeffries et al., 2012). In this case approximately 2% of assessed genes were monoallelically expressed, of which half were stochastically expressed from different alleles in different clones (Jeffries et al., 2012). Similar frequencies were observed in the other two RNA-sequencing based genome-wide screens, also performed in neural cell types (Li et al., 2012b; Lin et al., 2012), with 2.4% of autosomal genes (170 of 7,198 assessed) in mouse neural stem cells exhibiting monoallelic expression (Li et al., 2012b). Interestingly, the number of monoallelically expressed genes appears to increase upon differentiation: 314 genes in human iPS cell lines exhibit monoallelic expression, compared to 801 in differentiating neurons derived from the same cells (Lin et al., 2012), however overall frequencies were not reported, and these numbers include X-linked genes, thus more investigation into how monoallelic expression changes during differentiation is required.

1.3.2 Current understanding of monoallelic expression and outstanding questions

Despite the numerous isolated examples and handful of genome-wide screens identifying monoallelically expressed genes in different cell types, very little is known about the features of these genes, and in depth characterisation and mechanistic insights is limited. Several themes, however, have emerged regarding random autosomal monoallelic expression. To date, the majority of monoallelically expressed genes identified are inheritable, that is once established, the allelic expression patterns are maintained through mitosis such that clones of cells, all expressing the same allele, are made. However, this may be simply a reflection of the methodologies used to identify monoallelically expressed genes, particularly with the genome-wide screens. A recent notable example of random monoallelic expression which is not strictly inheritable, is that of the pluripotency gene *Nanog* in early development (Miyanari and Torres-Padilla, 2012). Random monoallelic expression of *Nanog* is observed from the 2-cell stage embryo, up until formation of the blastocysts in which it switches to biallelic expression in the naive epiblast, before reverting again to monoallelic expression following implantation. The switch from monoallelic

to biallelic expression is proposed to be linked to the transition towards a ground state of pluripotency (Miyanari and Torres-Padilla, 2012), and is reflected by observations that ESCs cultured in serum, which gives rise to a heterogeneous mix of cells of variable pluripotency, are predominantly monoallelic, where as those cultured in 2i medium which mimics the pluripotent ground state (Ying et al., 2008), are predominantly biallelic (Miyanari and Torres-Padilla, 2012). However, the existence and physiological relevance of monoallelic expression of *Nanog* to pluripotency has recently been questioned (Faddah et al., 2013; Filipczyk et al., 2013), as the distinction between gene pulsing and unstable monoallelic expression is somewhat ambiguous.

There are several features that the inheritable random autosomal monoallelically expressed genes identified thus far share (table 1.1 on page 51). First of all, the identified genes are scattered throughout the genome, and do not fall into any clusters. Furthermore, there is lack of chromosome coordination of these genes (Gimelbrant et al., 2007; Wang et al., 2007; Zwemer et al., 2012), such that adjacent genes are not necessarily both expressed from the paternal or maternal alleles. One exception to this feature are the interleukins, in which there is a 81% concordance of the expressed allele for the clustered IL-4, IL-13 and IL-5, however not IL-3 which is located 450kb away (Kelly and Locksley, 2000). Secondly, the genes exhibiting random monoallelic expression encode proteins of a wide-range of functions. While there is a small enrichment for cell surface and transmembrane proteins in some studies (Gimelbrant et al., 2007; Jeffries et al., 2012; Li et al., 2012b), others report no gene ontology enrichment (Zwemer et al., 2012). Interestingly, not all cells exhibit monoallelic expression for a particular monoallelically expressed gene. In all 7 genome-wide screens, the vast majority of monoallelic genes showed biallelic expression in at least one of the single-cell derived clones. This feature was also observed at individual genes including *Albumin* in hepatocytes (Michaelson, 1993), the interleukins (Guo et al., 2005), *p120 catenin* (Gimelbrant et al., 2004), and *Gfap* (Takizawa et al., 2008a). This suggests that monoallelic expression is not strictly required by the cell and instead likely reflects independent random stochastic activation of the two alleles in the cell. Finally, no clear rule regarding the effect of monoallelic expression on the total transcript level in the cell can be made. While there is a general trend for monoallelically expressing cells have less transcript levels than biallelically expressing cells (Gimelbrant et al., 2007), the reduction is approximately 30-35% (Li et al., 2012b), not the expected 50%, suggesting that reduced

transcript levels is not a general rule for all monoallelically expressed genes (Jeffries et al., 2012). Therefore the functional consequences, if any, for the majority of random autosomal monoallelically expressed genes remains unknown.

While there are some similarities between random autosomal monoallelic expressed genes and the classic examples of monoallelic expression, there are also marked differences (table 1.2). All types of monoallelic expression identified, with the exception of genomic imprinting and monoallelic expression resulting from genomic *cis* effects, exhibit a random choice of allele expressed. Additionally, the allelic imbalance is mitotically stable, and thus is inherited across cell division. However, random autosomal monoallelically expressed genes are not always expressed exclusively from one allele, as is observed for X-chromosome inactivation, genomic imprinting, olfactory receptors and immunoglobulins. Furthermore, the random autosomal monoallelic genes are scattered throughout the genome and do not fall into any clusters, unlike the other cases of monoallelic expression. While the classic examples of monoallelic expression have been extensively studied and are fairly well understood at the molecular level, detailed characterisation and mechanistic insights into random autosomal monoallelic expression is limited. Thus it remains unknown if many of the features associated with the classic examples of monoallelic expression hold true for the random autosomal monoallelically expressed genes (table 1.2). For example, global nuclear organisation is strongly implicated in X-chromosome inactivation (Chow and Heard, 2010), olfactory receptors (Clowney et al., 2012; Lomvardas et al., 2006), immunoglobulins (Kosak, 2002; Skok et al., 2001) and to a lesser extent in imprinted genes (Krueger et al., 2012), however to date only one study has examined the role of nuclear organisation of a random autosomal monoallelically expressed gene (Takizawa et al., 2008a). Additionally, how the alleles are distinguished molecularly through either DNA methylation and/or histone modifications, has not been extensively covered. While two genome-wide screens were based on the assumption that monoallelically expressed genes show differential DNA methylation (Wang et al., 2010, 2007), and another suggested that monoallelically expressed genes may have decreased CpG density and increased DNA methylation levels compared to biallelic genes (Jeffries et al., 2012), careful examination of contribution of DNA methylation to monoallelic expression has not been made. Furthermore, there has been only one report identifying histone modifications that are preferentially associated with monoallelically expressed genes, including a 2.9

Table 1.1: Features of random autosomal monoallelically expressed genes.

	Gimelbrant et al. 2007 Science	Wang et al. 2007 PLoS ONE	Wang et al. 2007 PLoS ONE	Li et al. 2012 PLoS ONE	Lin et al. 2012 PLoS ONE	Jeffries et al. 2012 Stem Cells	Zwemer et al. 2012 Genome Biology
technology platform	microarray	concurent methylated and unmethylated DNA	concurent methylated and unmethylated DNA	RNA-sequencing	RNA-sequencing	microarray	microarray
species	human	mouse	mouse	mouse	human	human	mouse
cell type	lymphoblast fibroblast placenta	CNS	brain neural stem cells	neural stem cells	iPS cells and derived neurons	neural stem cells from cerebral cortex, striatum and spinal cord	lymphoblast fibroblast
number genes / percentage of assessed genes	371 genes 5-10%	5 genes 1%	9 genes 1-2%	170 genes 2.4%	314 - 801 genes (includes X-linked genes)	143 - 203 genes ~2%	212 genes >10%
chromosome coordination	no	no	-	-	-	-	no
gene ontology	transmembrane proteins	-	-	glutathione receptors, annexins, protocadherins	-	transmembrane glycoproteins	none
proportion genes strictly monoallelic	<20%	40%	<40%	35%	low	low	low
DNA methylation	-	assumed	assumed	-	-	correlative evidence for higher methylation levels	-
Histone modifications	-	-	-	-	-	correlative evidence for increased H3K27me3, decreased H3K4me3, H3K9ac, H3K9me3	-
Expression levels of monoallelic clones	less	-	-	30-35% less	-	weak but significant decrease	-

fold increase in H3K27me3 repressive marks and approximately a 2 fold decrease in H3K4me3, H3K9ac and H3K9me3 (Jeffries et al., 2012), however direct evidence of preferential association of these marks between active and inactive alleles has not been performed. Asynchronous DNA replication timing is a distinguishing feature of classic monoallelic expression (Chess, 2012a). While the replication timing patterns of most random autosomal monoallelically expressed genes remain to be determined, selected examples, including the interleukins (Ensminger and Chess, 2004; Hollander et al., 1998) and the non-coding RNA *Asar6* (Donley et al., 2013), exhibit asynchronous DNA replication timing. However, asynchronous replication timing is generally coordinated across entire chromosomes (Dutta et al., 2009; Ensminger and Chess, 2004), where as the choice of active allele is not (Gimelbrant et al., 2007; Wang et al., 2007; Zwemer et al., 2012), suggesting that even if monoallelically expressed genes are asynchronously replicating, the early replicating allele may not necessarily be the active allele (Chess, 2012a). Thus replication timing may not be the critical feature of random autosomal monoallelically expressed genes. Finally, the dynamics of monoallelic expression during differentiation, and tissue-specificities of the random autosomal monoallelically expressed genes remains to be determined. Increasing our understanding of which genes undergo random monoallelic expression, and how this state is maintained across cell divisions, will provide insights into potential functional consequences of random autosomal monoallelic expression in mammalian cells.

1.3.3 Embryonic stem cell differentiation as a paradigm to investigate monoallelic expression

One of the outstanding unanswered questions regarding random autosomal monoallelic expression is when and how the monoallelic state is established. For other forms of random monoallelic expression, including X-chromosome inactivation and immunoglobulin allele-exclusion, the future active and inactive alleles are selected in early development, around the time of implantation from E3.5 to E8.5 (Heard et al., 1997). Embryonic stem cells (ESCs) are pluripotent cells derived from the inner cell mass of the E4.5 blastocyst. They provide a useful tool for studying this stage of early development due to their unlimited self-renewal in culture, and ability to differentiate into any of the adult somatic cell types. ESCs have been an invaluable tool for the X-chromosome inactivation field. From the 2-cell stage to early blastocysts, both of the two

Table 1.2: Comparison of random autosomal monoallelic expression with other types of monoallelic expression.

	X-inactivation	**Genomic Imprinting**	**Immuoglobulins**	**Olfactory Receptors**	**Random Autosomal**
Genomic distribution	X-chromosome	Autosome, clusters	Autosome	Autosome, clusters	Autosome, dispersed
Choice of allele	random	parental	random	random	random
Mitotic inheritance	yes	yes	yes	n.a.	yes
Number of genes	~1,000	~150	7	~1400	100s - 1,000s
DNA methylation	yes	yes	yes	yes	?
Histone modifications	yes	yes	yes	yes	?
Nuclear organisation	yes	yes	yes	yes	?
DNA replication timing	asynchronous	asynchronous	asynchronous	asynchronous	?
Non-coding RNAs	*Xist, Tsix...*	*Air, Kcnq1ot1...*	no	no	?
Function	dosage compensation	parental conflict hypothesis	receptor diversity	receptor diversity	?
Mechanism	sotchastic	deterministic	deterministic	stochastic	?

X-chromosomes are active, and is believed that the lack of dosage compensation at this time point is not lethal to the embryo as not many X-linked genes are expressed at this early stage of development (Heard et al., 1997). Indeed, the consequences of two active X-chromosomes is only lethal around 10 dpc (Takagi and Abe, 1990). The earliest signs of X-chromosome inactivation is the asynchronous DNA replication timing observed at the two X-chromosomes at E3.5 to E4.5 (Takagi et al., 1982). The silencing of one of the two X-chromosomes occurs at the late blastocyst stage, at 5.5 to 6.5 dpc (McMahon and Monk, 1983) and is complete by gastrulation at 6.5 dpc (Monk and Harper, 1979). Embryonic stem cells contain 2 active X-chromosomes of which one undergoes random inactivation upon differentiation (Martin et al., 1978; Rastan and Robertson, 1985), and thus have been essential for many of the studies investigating the molecular mechanisms of X-inactivation. Similar to the X-chromosomes (Takagi et al., 1982), the immunoglobulin and T-cell receptor loci also replicate asynchronously in ESCs, and it is almost always the early replicating allele that later undergoes recombination in the B cells (Mostoslavsky et al., 2001). Additionally, other forms of random monoallelic expression, including the interleukins and olfactory receptors, replicate asynchronously in human embryonic stem cells (Dutta et al., 2009), again suggesting that the choice of the future active allele may be determined at this stage of development. Embryonic stem cells have also been useful in studying the early events of imprinted gene expression (Latos et al., 2009). In this way, embryonic stem cells represent a good *in vitro* system for studying random monoallelic expression.

Chapter 2

Embryonic Stem Cell Pluripotency and Differentiation

Embryonic stem cells (ESCs) are derived from the inner cell mass of the blastocyst embryo (figure 2.1A). Mouse ESCs were first derived in 1981 (Evans and Kaufman, 1981; Martin, 1981), and human ESCs in 1998 (Thomson et al., 1998). As mentioned above, ESCs are unique in their ability to self renew indefinitely, as well as contribute to cell types of all three embryonic germ layers: endoderm, mesoderm and ectoderm. While embryonic development was originally thought to irreversible, the pluripotent state can be reprogrammed in differentiated somatic cells either by somatic cell nuclear transfer (Gurdon, 1962), in which an adult nucleus is transferred into either an enucleated oocyte or an embryonic stem cell recipient; nuclear fusion of a somatic cell with a pluripotent cell (Blau et al., 1983); or over-expression of a set of 4 genes: *Oct4*, *Sox2*, *Klf-4* and *C-myc* (Takahashi and Yamanaka, 2006). This indicates that cell fate decisions are not irreversible, and, while the differentiated state is generally stable, it is also dynamically controlled and can be changed (Yamanaka and Blau, 2010).

2.1 Embryonic Stem Cell culture and characterisation

Mouse ESCs are typically cultured on a feeder layer, typically mouse embryonic fibroblasts (MEFs), and form tightly packed round colonies containing tens to hundreds of cells (figure 2.1B). These cells express the pluripotency markers Oct4 (Pou5f1) and Nanog (figure 2.1C)

which is absent in the MEF feeder cells. Nanog expression levels are variable, a feature which is thought to be linked to their pluripotent potential with Nanog^{high} cells undergoing self-renewal and Nanog^{low} cells being poised for differentiation. It has been proposed that stochastic monoallelic expression of *Nanog* contributes to the two populations of cells (Miyanari and Torres-Padilla, 2012), although this has recently been disputed (Faddah et al., 2013; Filipczyk et al., 2013). ESCs are traditionally cultured in serum containing medium supplemented with leukemia inhibitory factor (LIF) which keeps the cells in a pluripotent state by activating the STAT3 signaling pathway, and serum containing bone morphogenetic protein 4 (BMP4) which inhibits the the differentiation process, and are induced to differentiate through FGF signaling. ESCs can also be cultured in serum-free medium supplemented with small molecule inhibitors which block Erk signaling downstream of the FGF receptor and Glycogen Synthase Kinase 3ß (Ying et al., 2008), and represent a more homogeneous population of cells which show properties more similar to that of the inner cell mass, including hypomethylation (Ficz et al., 2013), and homogeneous nanog expression (Miyanari and Torres-Padilla, 2012). With the exception of selected experiments in Chapter 3, ESCs were cultured in serum containing medium on MEF feeder cells.

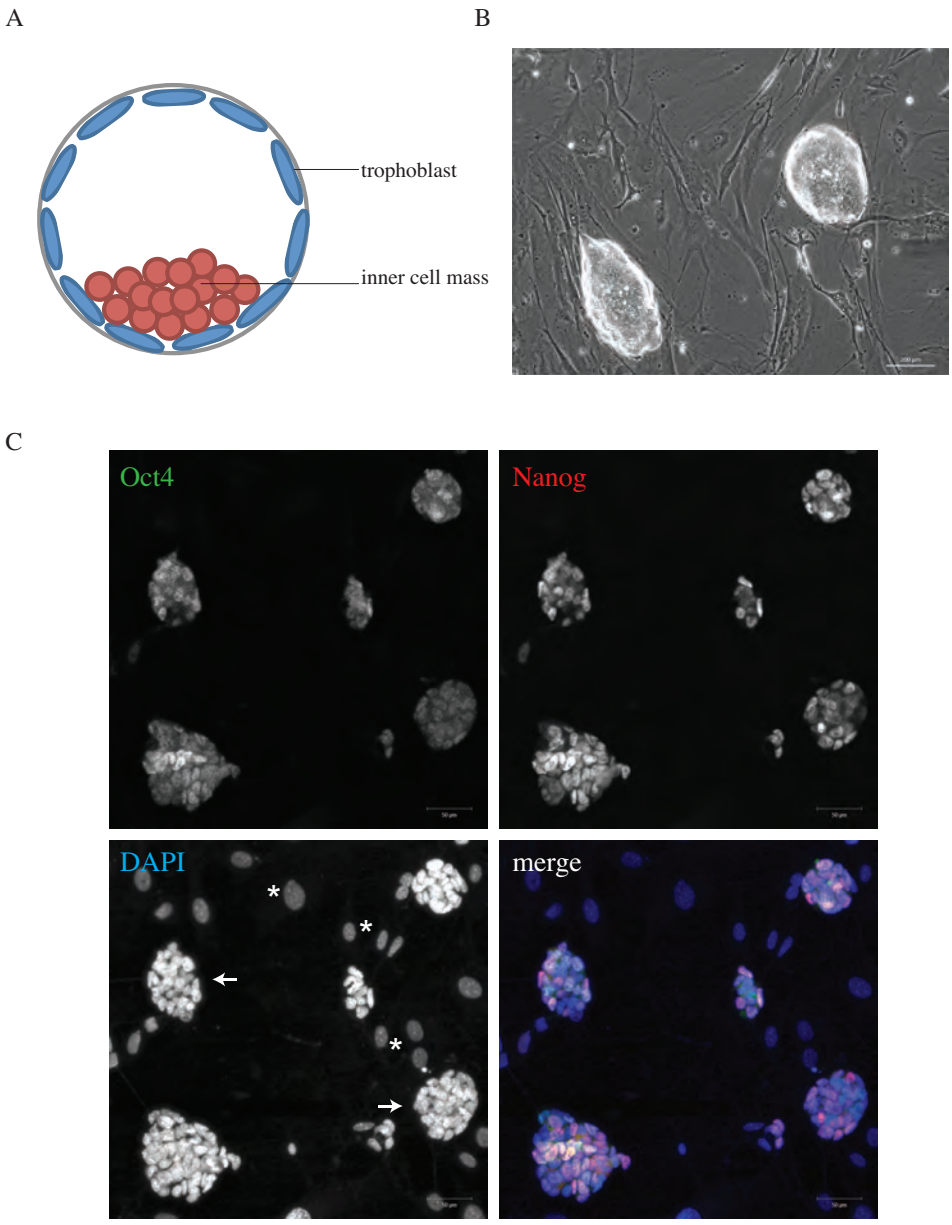


Figure 2.1: Pluripotency of mouse embryonic stem cells.
(A) Early blastocyst-stage embryo schematic showing inner cell mass from which ESCs are derived. (B) Light micrograph of ESCs grown on MEF feeder cells in culture. ESCs form round colonies containing 10s to 100s of tightly compacted pluripotent cells (arrows). Scale bar = 200μm. (C) Immunofluorescence staining of ESCs (arrows) grown on MEF feeder cells (*). ESCs express the pluripotency markers Oct4 (green) and Nanog (red). DAPI shows DNA. Scale bar = 50μm. Images represent single z-slice.

2.2 Embryonic Stem Cell differentiation paradigms

ESCs can be differentiated *in vitro* by removing them from the MEF feeder cells and withdrawing LIF from the medium. The process of differentiation can be sped up by addition of retinoic acid (RA), an analogue of vitamin A which through binding to the retinoic acid receptor induces and represses gene transcription of a large number of genes, including the HOX genes (Duester, 2008). By day 2 of RA induced differentiation, expression levels of *Nanog* are absent, and *Oct4* shutoff by day 4 (figure 2.2A on the next page). In contrast markers of the three germ layers as well as trophectoderm increase during RA induced differentiation (figure 2.2B-E), indicating that while the ESCs have lost pluripotency and differentiated, the resulting population of differentiated cells is very heterogeneous. In order to assess monoallelic gene expression during ESC differentiation, a homogenous population of differentiated cells is required. Neural progenitor cells (NPCs) were selected as the differentiated cell type for several reasons. Firstly, NPC differentiation is a fairly well established differentiation paradigm and is used by numerous labs. Therefore the protocols are fairly well established and reagents required readily available. Because of its wide use, there are also a large number of genome-wide data sets comparing ESCs to NPCs for many different features including histone-modifications (Mikkelsen et al., 2007), DNA methylation (Meissner et al., 2008), Lamin-associated domains (Peric-Hupkes et al., 2010) and DNA replication timing domains (Hiratani et al., 2008). Finally, monoallelic gene expression is known to exist in the nervous system, for example at the olfactory receptor and protocadherin gene clusters, therefore identification of potentially more autosomal monoallelically expressed genes could be of interest in the context of cell to cell variation and unique cell identity features of the nervous system.

The ESC to NPC paradigm established was based on that from Austin Smith's lab (Conti et al., 2005). A complete detailed protocol is provided in appendix C.1 on page 259. ESCs are collected and MEF feeder cells removed by soaking the cells on gelatin (figure 2.3). MEFs and other differentiated cell types will typically adhere to gelatin within 30 minutes, whereas ESCs take several hours. In this way the pluripotent ESCs can be isolated. Cells are then induced to differentiate in N2B27 medium on gelatin coated plates for approximately 6 days. This medium contains N2 and B27 supplements which contain, amongst other compounds, insulin, transferrin, progesterone, putrescine, selenite and retinoic acid. The concentration of cells initially seeded

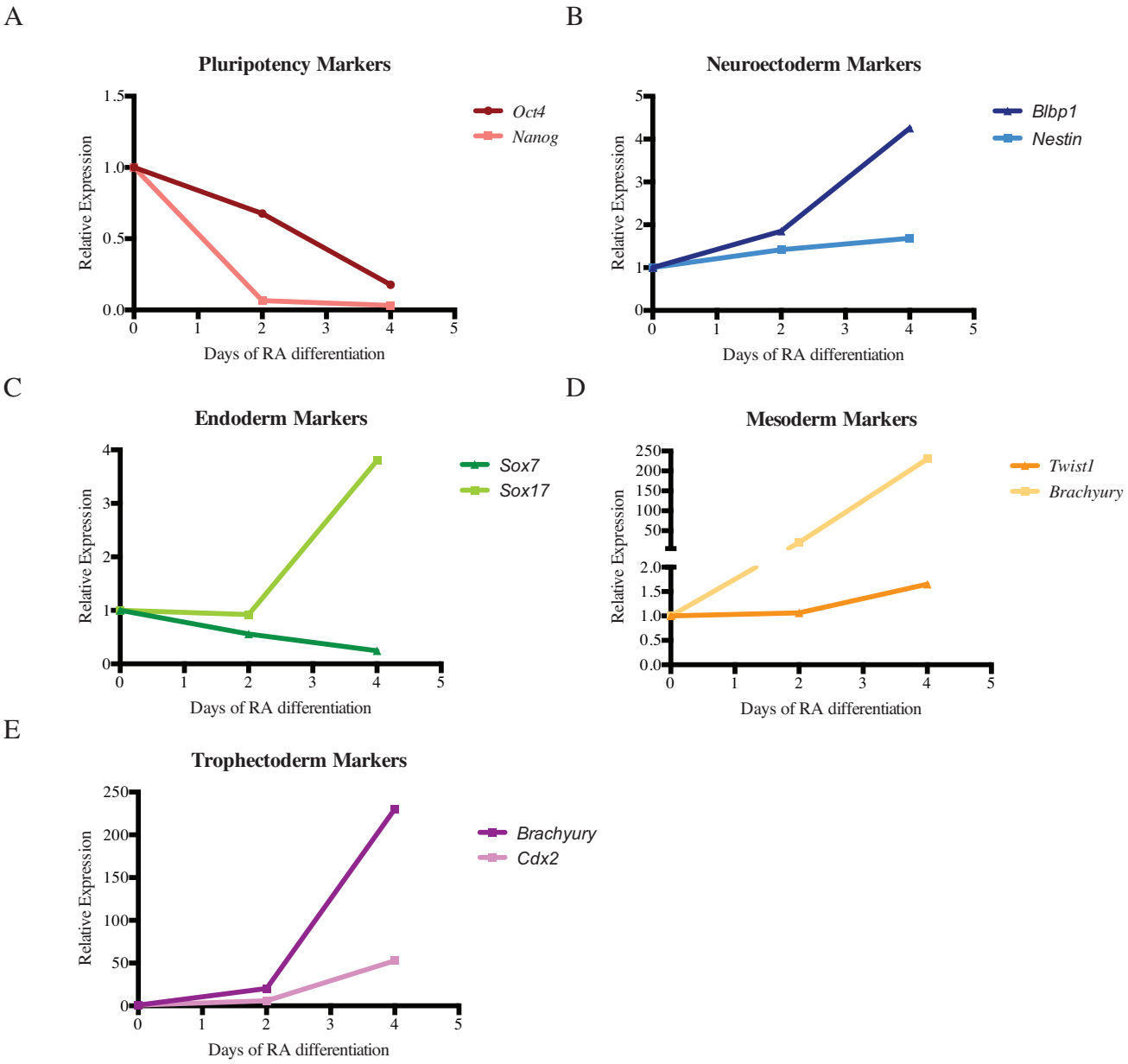


Figure 2.2: Retinoic induced differentiation of ESCs.
Expression of pluripotency (A), neuroectoderm (B), endoderm (C), mesoderm (D) and trophectoderm (E) markers during 4 days of retinoic acid induced differentiation.

was found to be critical, as it varies the efficiency of differentiation. Following 6 days, cells were collected and placed into suspension in N2 medium containing N2 supplement, EGF and FGF but lacking B27 supplement. Over the next 3-4 days, neurosphere aggregates will form and grow. These can be collected either by mild centrifugation or sedimentation, resuspended in fresh medium and placed back onto gelatin coated plastic. Within a few days the NPCs will be ready for passaging and within a few passages will represent a homogenous population of cells. The NPCs can be cryopreserved and sub-cloned, although the efficiency of colony formation is increased if B27 supplement is added to the medium.

The ESC derived NPCs have the typical small triangular morphology of NPCs (figure 2.3), and appropriate growth kinetics, requiring passaging approximately every 3 days. The NPCs lacked expression of the pluripotency markers *Oct4* and *Nanog*, which were present in ESCs, and gained expression of the neural markers *Nestin* and *Blbp1* (figure 2.4). Furthermore, immunofluorescence staining of the cells was performed using the neural intermediate filament marker Nestin, which was specifically present in the NPCs, but not the pluripotent ESCs (figure 2.5). Thus the differentiated homogenous population of cells have all the appropriate characteristics of NPCs.

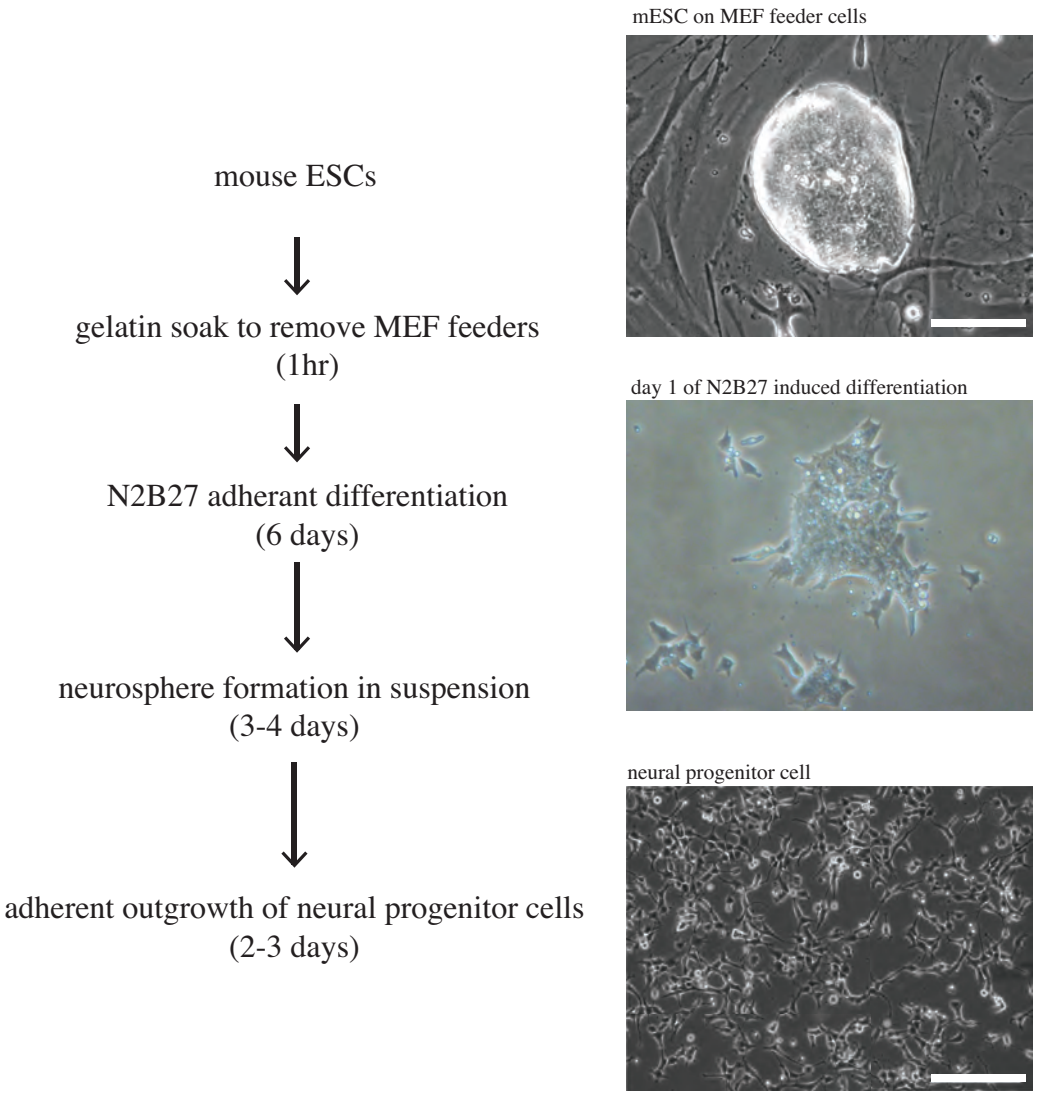


Figure 2.3: ESC to NPC differentiation protocol.
Outline of protocol used to differentiate mouse ESCs to NPCs. Representative light micrographs of different stages of the differentiation paradigm are shown. Scale bar = 200μm.

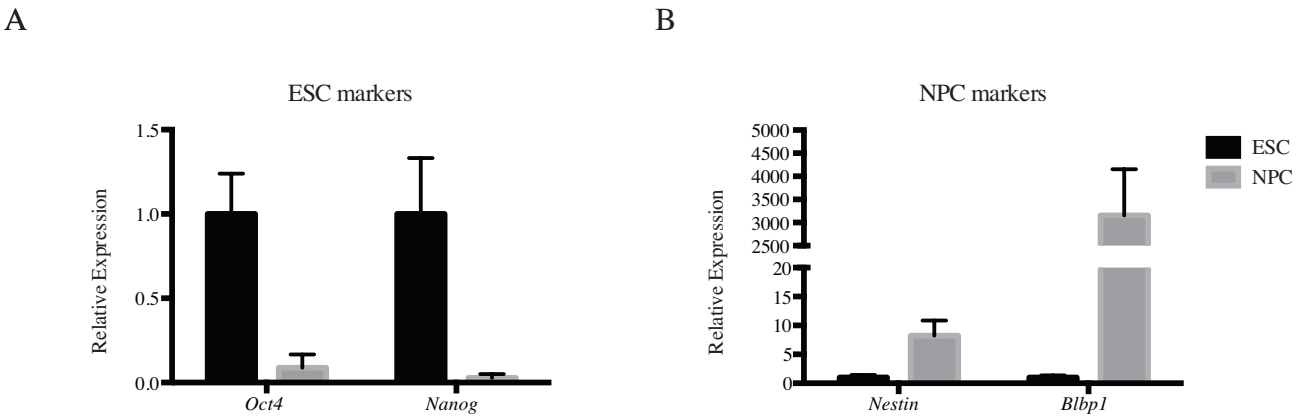


Figure 2.4: Expression of lineage marker expression upon ESC differentiation to NPC. Comparison of pluripotency (A) and neural (B) markers in ESCs (black) and NPCs (grey). Error bars represent SEM of 3 biological replicates.

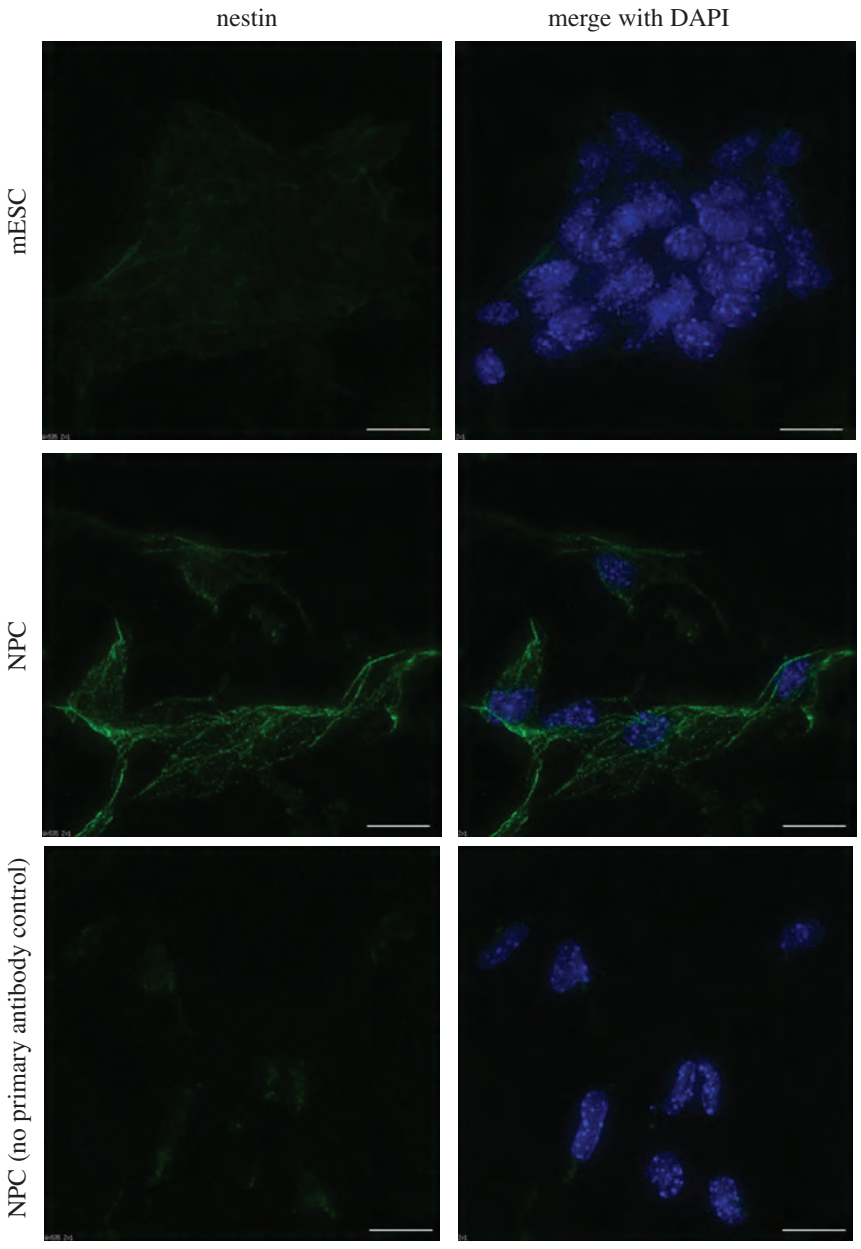


Figure 2.5: Nestin immunofluorescence staining of ESCs and NPCs.
Immunofluorescence staining for the neural marker nestin in ESCs (top row), derived NPCs (middle row). Bottom row represents negative control where no primary antibody was used. Scale bar = 10μm.

2.3 Nuclear dynamics of ESCs and differentiated cell types

ESCs are unique in their pluripotency which is reflected in a fluid and malleable nuclear structure. Indeed, live cell imaging of ESCs transiently transfected with mCherry-Lamin protein fusions demonstrates their hyperdynamic nucleus (figure 2.6 on the next page) over very short time scales. In contrast, differentiated MEFs have a more static and regularly shaped nucleus which lacks the protrusions and invaginations seen in ESC nuclei. NPCs are intermediates between the two cell types, while they still have dynamic nuclei, their shape is slightly more regular, although not to the same extent as the MEFs. One of the major dogmas at the time of these experiments, was that ESCs do not express Lamin A/C, one of the Lamin protein family members, and that the absence of Lamin A/C is not only a marker for the undifferentiated state (Constantinescu et al., 2006), but also be the reason for the dynamic nucleus and plastic chromatin associated with pluripotency (Mattout and Meshorer, 2010; Meshorer and Misteli, 2006). This motivate these original live-cell imaging studies, however, the expression of Lamin A or Lamin C in ESCs did not alter the nuclear dynamics (figure 2.6), nor did it change their gross morphology or growth kinetics in culture. This finding led to the interesting and controversial finding that ESCs do indeed express low levels of Lamin A/C and that the purported absence of Lamin A/C is not an essential attribute of pluripotency (Eckersley-Maslin et al., 2013).

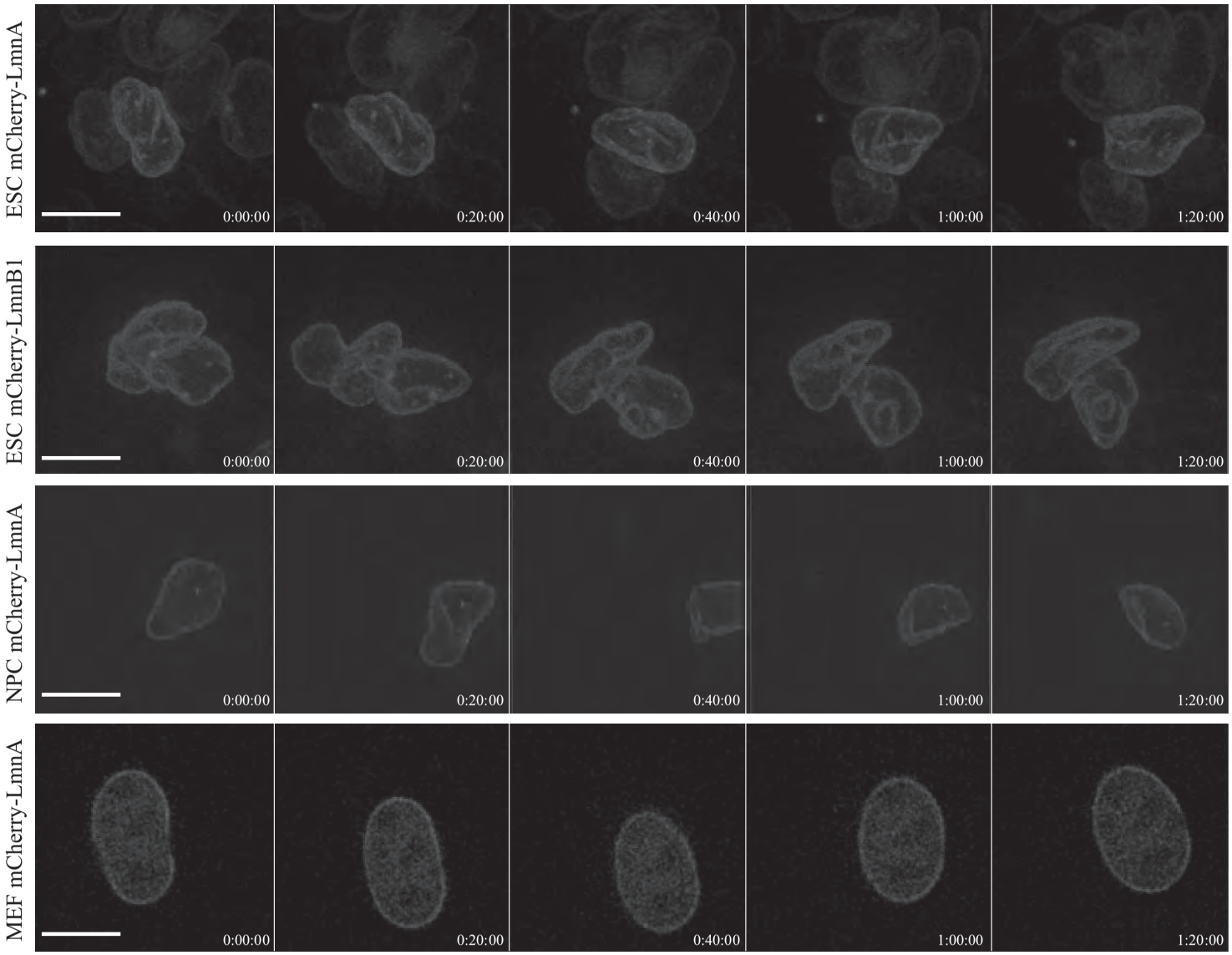


Figure 2.6: Live cell imaging of ESC, NPC and MEFs.
Movie montage at 20 minute intervals showing dynamic and irregular shape of ESC nuclei transiently transfected with either mCherry-LmnA (top row) or mCherry-LmnB1 (second row, 2 cells shown), compared to differentiated MEFs transiently transfected with mCherry-LmnA (bottom row). NPCs transiently transfected with mCherry-LmnA also show dynamic nuclei, although their shape is more regular (third row). Images represent projections of deconvolved image stacks. Image bar represents 10μm.

Chapter 3

Lamin A/C is expressed in pluripotent mouse embryonic stem cells

Embryonic stem cells are unique in their nuclear structure and organization, typically consisting of a more fluid and dynamic chromatin compared to somatic cell types (Denholtz and Plath, 2012; Fisher and Fisher, 2011; Li et al., 2012a; Meister et al., 2011) (figure 2.6). It has been widely purported that these unique features of ESCs are due to the absence of the nuclear intermediate filament Lamin A/C protein in the cell and many models have been made linking Lamin A/C absence to pluripotency (Li et al., 2012a; Melcer et al., 2012; Meshorer and Misteli, 2006). I was therefore interested to see what effect of expressing Lamin A/C in ESCs would have on chromatin, gene expression and pluripotency of these cells. However, surprisingly, Lamin A/C was clearly detected at both the mRNA and protein levels in both established and primary ESC lines. Furthermore, I showed that Lamin A/C is also readily detected in the inner cell mass of E3.5 blastocysts, thus refuting the dogma that lack of Lamin A/C is an essential feature of pluripotency. These observations were published as an article in *Nucleus* in January 2013 (Eckersley-Maslin et al., 2013).

Author contributions to the manuscript: MAE-M and DS designed experiments and wrote manuscript; MAE-M generated RNA-sequencing libraries and performed all Q-PCR, Western Blotting, Immunofluorescence experiments; JHB performed RNA-sequencing data analysis; ZL performed super-resolution OMX microscopy (not included in final manuscript).

3.1 Abstract

The pluripotent nature of embryonic stem cells (ESC) is associated with a dynamic open chromatin state and an irregular nuclear shape. It has been postulated that the absence of lamin A/C contributes to these features. However, we show that mouse ESCs express low, yet readily detectable, amounts of lamin A/C at both the RNA and protein levels. Full-length transcripts of both isoforms were readily detected by q-PC R and deep RNA sequencing. Additionally, protein expression was validated in multiple primary and established ESC lines by immunoblotting using several independent antibodies. Immunofluorescence labeling showed localization of Lamin A/C at the nuclear periphery of all Oct4/Nanog double-positive ESC lines examined, as well as in the inner cell mass of blastocysts. Our results demonstrate ESC s do express low levels of lamin A/C, thus models linking pluripotency and nuclear dynamics with the absence of lamin A/C need to be revisited.

3.2 Introduction

Embryonic stem cells (ESCs) are pluripotent cells derived from the inner cell mass of the pre-implantation blastocyst, which are unique in their ability to self-renew indefinitely as well as differentiate into essentially all cell types of the adult organism. ESC pluripotency is accompanied by a unique dynamic chromatin organization and nuclear morphology (reviewed in (Denholtz and Plath, 2012; Fisher and Fisher, 2011; Li et al., 2012a; Meister et al., 2011)). The irregular nuclear shape of ESCs contains many invaginations and protrusions which are highly malleable (Pajerowski et al., 2007), and the inter-membrane space of the nuclear envelope is wider and unevenly spaced compared to differentiated cells (Smith et al., 2011). Upon differentiation, this irregular nuclear shape transitions to a smooth and rigid form. Differentiation is further paralleled by changes in the association of blocks of chromatin with the nuclear periphery, including pluripotency genes such as nanog, (Peric-Hupkes et al., 2010) suggesting a tight connection between nuclear organization and the differentiation program.

Globally, the chromatin in ESCs is dispersed, loosely packed (Ahmed et al., 2010), with large, poorly defined heterochromatin domains and an increased mobility of heterochromatin proteins, such as HP1 (Meshorer et al., 2006). The chromatin is plastic and hyperdynamic,

compared to differentiated cell types (Meshorer et al., 2006). This openness of the chromatin is correlated with high DNAse I accessibility (Fisher and Fisher, 2011), and wide-spread transcription of both coding regions as well as repetitive elements of the genome (Efroni et al., 2008). Furthermore, ESCs contain extensive regions with a bivalent chromatin signature in which the transcription-associated H3K4me3 histone modification co-exists with the polycomb-group repressive H3K27me3 modification (Mikkelsen et al., 2007). These "poised" domains are largely enriched over lineage-specific genes and become resolved to either an active (H3K4me3) or inactive (H3K27me3) state during differentiation(Mikkelsen et al., 2007). The resolution of bivalent domains to an inactive state correlates with an increase in DNA methylation at these promoters (Mohn et al., 2008). The interconnection between the unique chromatin features and distinctive nuclear structure is thought to have an essential role in ESC pluripotency.

The major structural component of the nucleus is the nuclear lamina, which forms directly below the inner-nuclear membrane (Dechat et al., 2008; Zhao et al., 2009). The nuclear lamina is composed predominantly of a meshwork of type V intermediate filament proteins called lamins, which interact with chromatin, nuclear pore complexes and lamin-associated proteins{Wilson:2010cw}, and have important functions in gene regulation and chromatin organization (Dechat et al., 2010). The nuclear lamins can be subdivided into A- and B-types based on sequence homology and structural features. There are two major B-type lamins in most vertebrates, lamin B1 and lamin B2, which are expressed throughout development and are required for organogenesis and proper neural development (Coffinier et al., 2011; Kim et al., 2011). The A-type lamins are encoded by a single gene which is alternatively spliced yielding two isoforms, lamin A and lamin C, which differ in the length of their C-terminal tail with lamin A being 74 amino acids longer. Loss of lamin proteins in vivo results in severe developmental defects (Kim et al., 2011; Sullivan et al., 1999), and Lamin A/C mutations are associated with premature aging in individuals with progeria (Dechat et al., 2010). Lamin A/C has been reported to be completely absent in ESCs (Constantinescu et al., 2006), and levels increasing during embryogenesis (Röber et al., 1989; Schatten et al., 1985; Stewart and Burke, 1987). This purported absence of Lamin A/C in ESCs has served as a landmark for models and hypotheses explaining pluripotency, chromatin dynamics and ESC nuclear plasticity (Li et al., 2012a; Melcer et al., 2012; Meshorer and Misteli, 2006).

We sought to systematically re-investigate whether ESCs express lamin A/C. Interestingly, we show that lamin A/C is expressed in ESCs at both the RNA and protein levels, although at lower levels than in differentiated mouse embryonic fibroblasts (MEFs). Furthermore we show that Lamin A/C is expressed in the inner cell mass of early blastocysts, from which ESCs are derived. Our results show conclusively that ESCs express lamin A/C, and that its absence is not a marker of the undifferentiated pluripotent state.

3.3 Results and Discussion

Lamin A/C is expressed in embryonic stem cells In order to examine lamin expression levels, we performed real-time quantitative PCR on RNA derived from ESCs, neural progenitor cells (NPCs) and mouse embryonic fibroblasts (MEFs) (figure 3.1A). ESCs were separated from the MEF feeder cells (see materials and methods) and, as expected, expressed the pluripotency markers *oct4* and *nanog* (figure 3.1A). Importantly, we also detected *lamin A* and *lamin C* isoforms in ESCs at a similar level to NPCs, yet at a lower level than MEFs (figure 3.1A). The lmna promoter is marked by the active-transcription associated histone H3 Lysine 4 trimethylation (H3K4me3) mark (Mikkelsen et al., 2007) supporting gene transcription. Examination of whole genome polyA+ RNA-sequencing data (J.H.B., M.A.E-M, D.L.S., unpublished data), as well as published data sets from mouse (Marks et al., 2012; Mikkelsen et al., 2007) and human ESCs (Birney et al., 2007), confirmed full length *lamin A/C* mRNA was generated above thresholds used to define active gene transcription (figure 3.1B). Together, these data demonstrate that the *lmna* gene is actively transcribed to yield full-length mRNA in ESCs.

To confirm that *lamin A/C* mRNA transcripts are being translated into protein, we performed immunoblotting experiments using a series of well characterized antibodies recognizing specifically either Lamin A/C (Moir et al., 1994) or Lamin A alone (Dechat et al., 2007). All antibodies examined showed a clear signal in AB2.2 ESCs (figure 3.2A). Both monoclonal and polyclonal Lamin A/C antibodies showed a doublet band, which corresponds to the two protein isoforms, whereas the Lamin A antibody specifically detected the larger Lamin A isoform. Importantly, no signal was seen in an identically prepared lamin knock-out (lmna-/-) ESC line (Sullivan et al., 1999), confirming antibody specificity and the purity of the ESCs from MEF

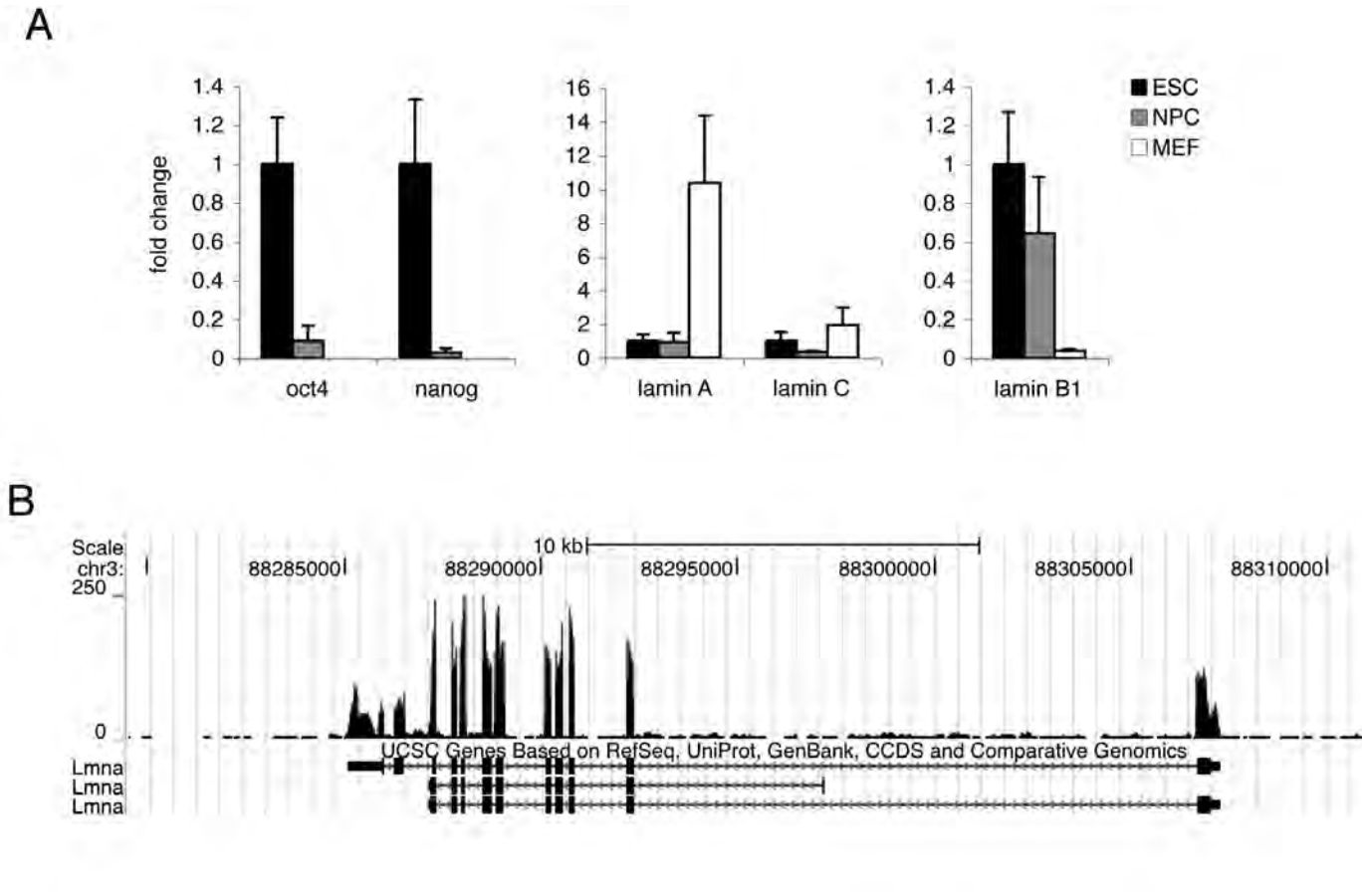


Figure 3.1: Lamin A/C is expressed in mouse Embryonic Stem Cells.
(A) Quantitative real-time RT-PCR of the pluripotency genes oct4 and nanog, and lamin A, lamin C and lamin B1 in ESCs (black), NPCs (grey) and MEFs (white). Error bars represent standard deviation of three biological replicates. Data was normalized to the geometric mean of three housekeeping genes. (B) Gene coverage plot showing expression of the full-length lmna gene in AB2.2 ESCs by whole genome deep RNA-sequencing of polyA+ RNA.

feeder cells (figure 3.2A, lane 2). Although this knock-out ESC line has been reported to expresses a truncated Lmna protein (Jahn et al., 2012), the epitopes recognized by the antibodies are not contained within this expressed portion. Both AB2.2 and lmna-/- ESCs expressed the pluripotency marker Oct4 (figure 3.2A). Lamin A/C protein was also detected in 3 early-passage ESC lines, 4 additional well-established ESC lines and an iPS cell line (figure 3.2B), confirming that Lamin A/C expression is not limited to the AB2.2 ESC line, nor an adaptation of long-term culturing, and is therefore a general feature of pluripotency. Furthermore, examination of previously published proteomics screens of ESCs (Graumann et al., 2008; Gundry et al., 2010) revealed Lamin A/C peptides. Combined, these results clearly demonstrate that lamin A/C is expressed at both the RNA and protein levels in multiple primary and well-established ESC lines.

Lamin A/C is correctly localized to the nuclear periphery in all cells within an ESC colony. We next performed immunofluorescence to determine expression patterns throughout individual cells in ESC colonies cultured either with MEF feeder cells in medium supplemented with 10% serum and leukaemia inhibitory factor (LIF), or in serum-free, feeder-free 3i medium (Ying et al., 2008) which maintains ESCs in a more uniform basal state of pluripotency. To ensure that the ESCs remained pluripotent, the ESCs grown on MEF feeders were not separated from the MEFs as above, but fixed on MEF feeder cells, which also serve as an internal control for immunofluorescence labeling. ESCs can be easily distinguished morphologically from MEFs as they grow in tight colonies, have a high nuclear to cytoplasmic ratio and have smaller and irregularly shaped nuclei (see for example figure 3.2C fourth panel). Lamin A/C immunofluorescence demonstrated correct localization to the nuclear periphery in both LIF (figure 3.2C) and 3i medium (figure 3.2C, bottom row). Importantly, all cells within the colony displayed Lamin A/C labeling, including the cells in the center of the colonies, confirming that Lamin A/C expression is not limited to the peripheral cells, which may occasionally include partially differentiated cells. The Lamin A/C positive ESCs co-expressed Oct4 and Nanog confirming their pluripotency (figure 3.2C, arrows). ESCs grown in 3i medium show uniform Nanog expression throughout all cells in the colony, consistent with this culture condition maintaining the cells in a ground-state of pluripotency (Marks et al., 2012; Ying et al., 2008). As a control, lmna -/- ESCs did not show any Lamin A/C labeling (figure 3.3, fifth row). Notably, ESCs have lower

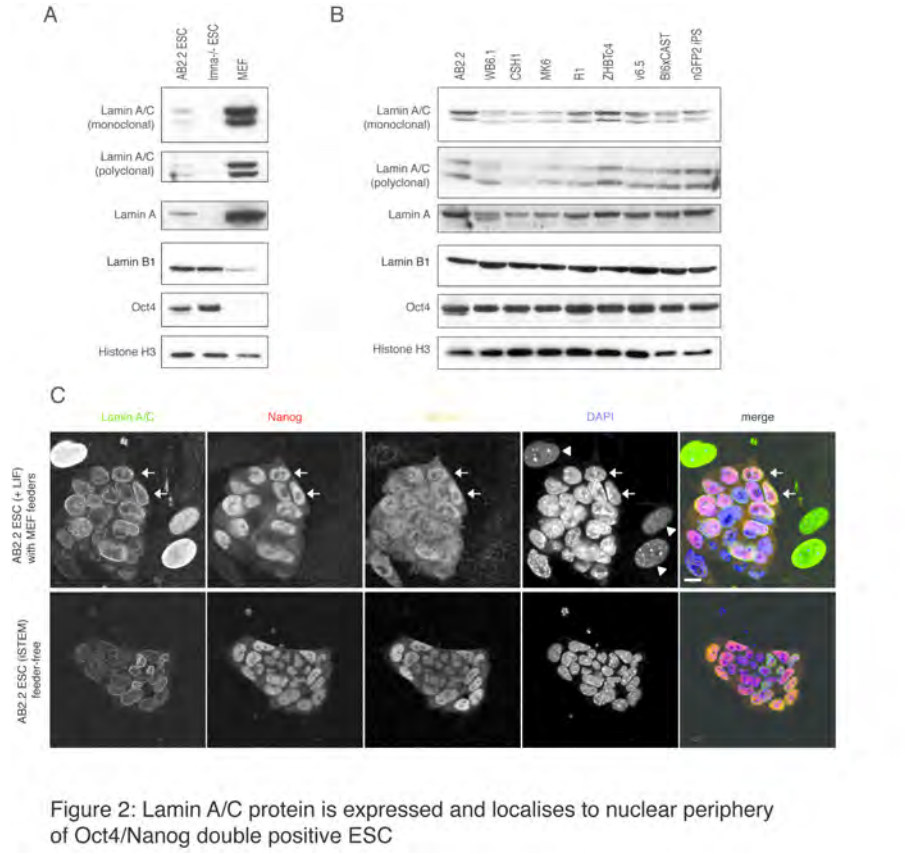


Figure 2: Lamin A/C protein is expressed and localises to nuclear periphery of Oct4/Nanog double positive ESC

Figure 3.2: Lamin A/C is expressed in mouse ESCs and localizes to the nuclear periphery of Oct4 positive ESCs.
(A) Lamin A/C protein expression is detected in AB2.2 ESCs by immunoblotting using three independent antibodies. Monoclonal (first row) and polyclonal (second row) Lamin A/C antibodies detect Lamin A (70kDa) and Lamin C (65kDa) isoforms. Lamin A/C knockout ESCs confirm purity of ESC protein preparation and immunoblotting specificity. All cell types express lamin B1. Oct4 is used to verify pluripotency and histone H3 is used as a loading control. (B) Extension of the immunoblotting analysis as above, in 9 independent ESC lines including AB2.2 (lane 1), three primary ESC lines (WB6.1, CHS1, MK6, lanes 2-4 respectively), established ESC lines (R1, ZHBTc4, v6.5, Bl6xCast, lanes 5-8 respectively), and an iPS cell line (lane 9). (C, D) Immunofluorescence was performed on ESCs grown on irradiated MEF feeder cells (top row) or in serum-free, feeder-free 3i medium (bottom row) for Lamin A/C (first column, green), nanog (second column, red) and Oct3/4 (third column, orange) and counterstained with DAPI (fourth column, blue). Representative image showing a colony of AB2.2 ESCs, which label for the pluripotency markers Nanog and Oct4 and have lamin A/C localization at the nuclear periphery. Arrows indicate the top of MEF nuclei. Images were taken on a Zeiss LSM710 confocal microscope. A single z-section through the center of the ESC colony is shown. Scale bar represents 10μm.

levels of Lamin A/C when compared to MEFs, which may explain why previous reports have failed to detected Lamin A/C in ESCs (Constantinescu et al., 2006), as ESC staining is very faint and could be mistaken for background staining when optimal exposures for MEF nuclei are used. However when compared to the negative control staining in which the primary antibodies were omitted (figure 3.3, last row), it is clear that the Lamin A/C signal observed is a bona fide localization signal. The localization of lamin A/C to the nuclear periphery of all cells in the ESC colony was further confirmed in other established ESC lines (figure 3.3). Immunofluorescence using a different antibody specifically against Lamin A also showed clear signal at the nuclear periphery in all cells in the colony in 5 separate ESC lines tested (figure 3.4). Our results convincingly show that Lamin A/C is correctly localized to the nuclear periphery in all pluripotent ESCs examined. Therefore, absence of Lamin A/C should not be used as a marker of the undifferentiated state.

Lamin A/C is expressed in blastocysts As Lamin A/C is expressed in embryonic stem cells, we tested whether we could detect Lamin A/C in the developing blastocyst. Examination of published single cell RNA sequencing data-sets of 2-cell, 4-cell, 8-cell and inner cell mass of blastocysts, in addition to ESCs, revealed *lmna* transcript was present at all developmental stages, above the significant expression threshold cut-offs used (Tang et al., 2011). We isolated fresh blastocysts at E3.5, and performed immunofluorescence labeling for Lamin A/C protein. We clearly detected Lamin A/C protein at the nuclear periphery of both nanog positive and nanog negative cells in the developing blastocyst (figure 3.5). The nanog positive cells represent the inner cell mass of the blastocyst from which ESCs are derived, demonstrating that the expression of Lamin A/C is not acquired upon ESC derivation, nor is it a cell-culture phenomenon.

Conclusion

In summary, we demonstrate that the dynamic nuclear structure of ESCs is not due to an absence of Lamin A/C. Our results show that lamin A/C is in fact expressed at both the RNA and protein levels in mouse ESCs as well as in the inner cell mass of the developing blastocyst. Additionally, immunofluorescence labeling clearly shows Lamin A/C expression in pluripotent Oct4/Nanog double positive ESC colonies, eliminating any concerns that the Lamin A/C detected by other assays is due to contamination of partially differentiated cells or MEF

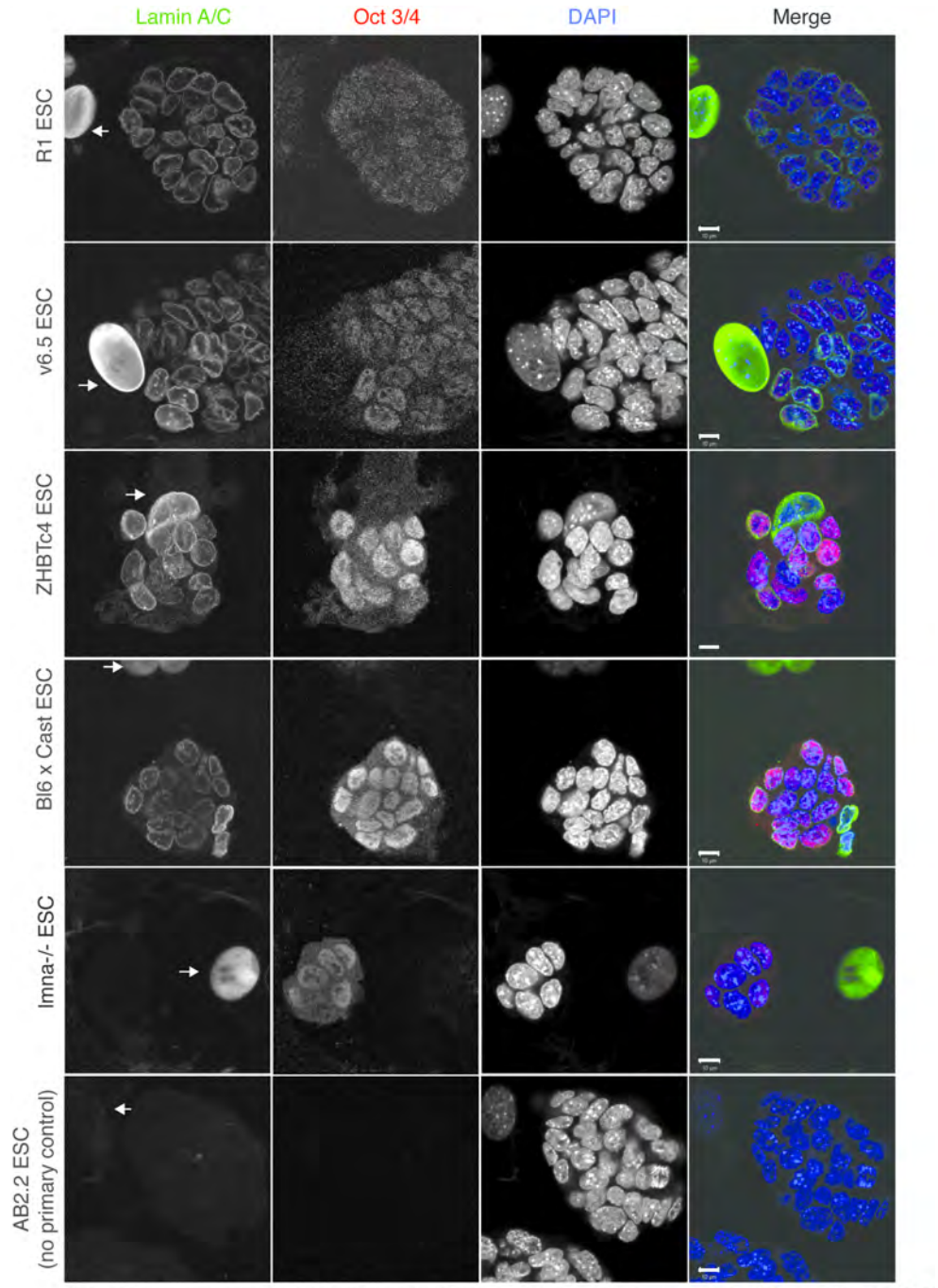


Figure 3: Lamin A/C localizes to the nuclear periphery of Oct4 positive ESC

Figure 3.3: Lamin A/C localizes to the nuclear periphery in Oct4 positive ESCs. Immunofluorescence against Oct4 and lamin A/C was performed in R1 (top), v6.5 (second row), ZHBTc4 (third row) and C57Bl6xCasteneous hybrid (fourth row) ESCs. The ESC colonies stain positively for Oct4 and Lamin A/C. Lmna-/- knockout ESC (fifth row) stain positively for Oct4 but not Lamin A/C which is present in MEFs (arrow), confirming antibody specificity. The bottom row shows a control immunofluorescence in AB2.2 ESCs in which primary antibodies were omitted. Scale bar represents 10μm. Images were taken on a Zeiss LSM710 confocal microscope. A single z-section through the center of the ESC colony is shown. Scale bars represent 10μm.

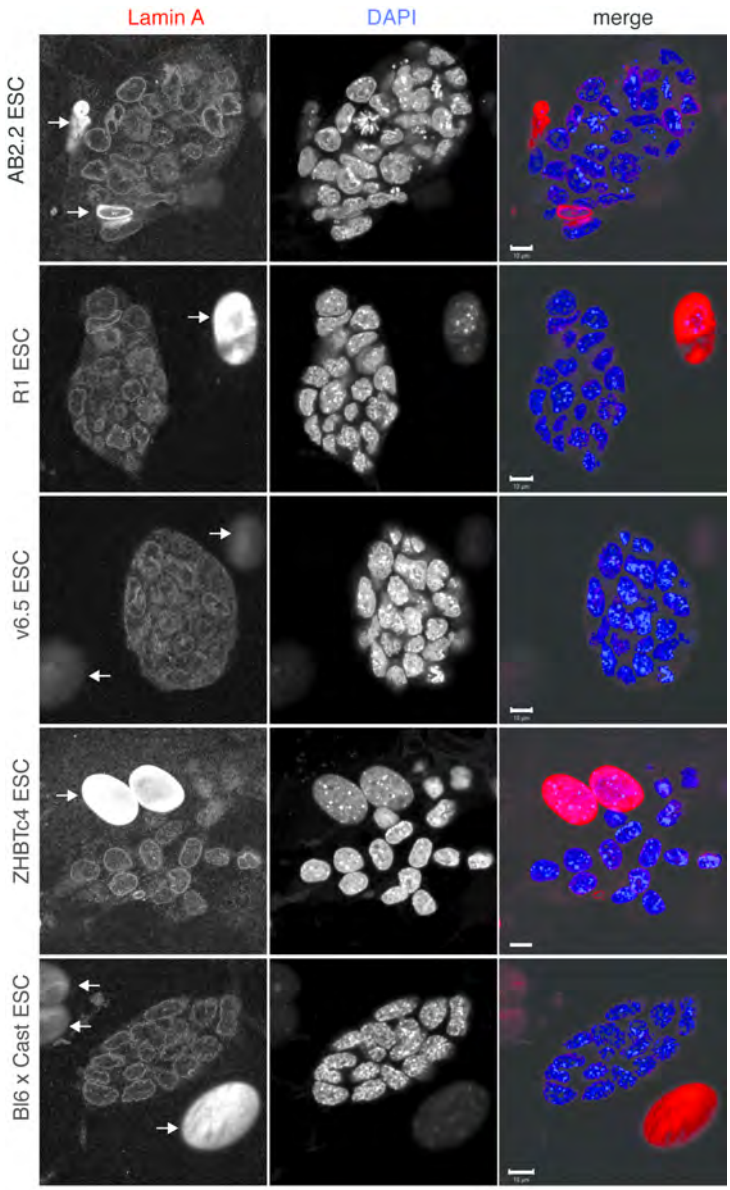


Figure 4: Lamin A is expressed in multiple ESC lines

Figure 3.4: Lamin A is expressed in multiple ESC lines.
Immunofluorescence using an antibody against lamin A (first column) shows localization to the nuclear periphery in all cells within the ESC colony in AB2.2 (top), R1 (second row), v6.5 (third row), ZHBTc4 (fourth row) and Bl6xCast (bottom row) ESC lines. Cells are counterstained with DAPI (second column). Images represent single 0.1μm sections through the center of ESC colony taken on a Zeiss LSM710 confocal microscope. Scale bar represents 10μm.

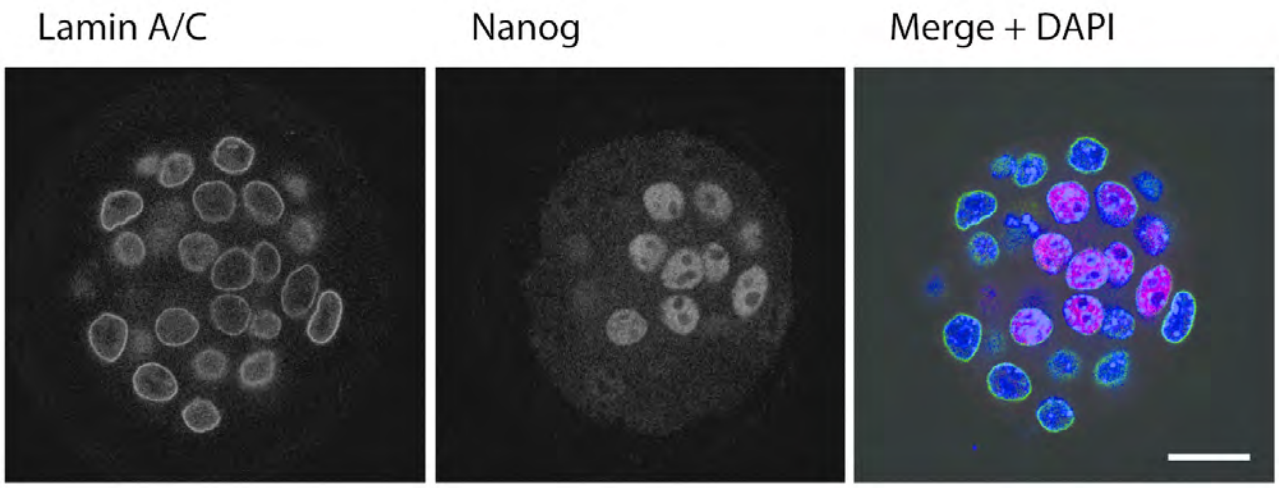


Figure 3.5: Lamin A/C is expressed in the inner cell mass of blastocysts. Immunofluorescence labeling of E3.5 blastocyst for Lamin A/C (left), nanog (center) and counterstained with DAPI (right). There is clear localization of Lamin A/C at the nuclear periphery of nanog positive cells, representing the inner cell mass as well as more peripheral nanog negative cells representing trophectoderm. Images were taken on a Zeiss LSM710 confocal microscope and represents a single section. Scale bar represents 20μm.

feeder cells.

A recent report has implicated Lamin A/C in ESC chromatin mobility (Melcer et al., 2012), however the effects seen were very moderate and limited to heterochromatin, with euchromatin remaining unchanged. Furthermore, overexpression of Lamin A/C did not result in any alterations in pluripotency gene expression nor ESC colony morphology (Melcer et al., 2012), consistent with our observations that Lamin A/C does not contribute to the link between nuclear structure and pluripotency. It remains to be determined whether the lowered levels of Lamin A/C compared to more differentiated cell types have a role in regulating the unique fluid chromatin and pluripotency of ESCs.

Importantly, our results provide insight into the absence of a phenotype in lamin B1/B2 knock-out ESCs (Kim et al., 2011), as the presence of a low level of lamin A/C would be sufficient to maintain cell viability and pluripotency. However, we do not rule out that total levels and, in particular, the ratio between the different lamin proteins, may play an important developmental role in either cell differentiation or cell fate decisions. The strong developmental defects seen in mice upon removal of either both lamin B1 and B2 (Kim et al., 2011), or lamin A/C (Sullivan et al., 1999), clearly demonstrate the importance of lamin proteins in coupling nuclear architecture to the gene expression program.

It remains to be determined how the highly irregular nuclear structure of ESCs is maintained in the presence of low levels of Lamin A/C, and how this relates to the pluripotent nature of the cells. While the lower levels of Lamin A/C in ESCs may contribute to their pluripotency, other nuclear envelope-associated factors may also be involved, and many are differentially expressed between ESCs and differentiated cells (J.H.B, M.A.E-M and D.L.S., unpublished data). In particular, Syne-1, a lamin-associated protein, has been implicated in regulating the changes in nuclear envelope spacing that accompany ESC differentiation (Smith et al., 2011). Additionally, ESC differentiation can be inhibited by preventing expression of the nuclear pore protein Nup210 (D'Angelo et al., 2012). How these, in addition to other protein(s) or factors, contribute to the dynamic nuclear structure of ESCs and their pluripotency will be an exciting area of research not only for the fields of chromatin, nuclear organization and stem cell biology, but also for cellular reprogramming and its clinical applications.

3.4 Materials and Methods

Cell Culture and Blastocyst Collection

ESCs were cultured using standard protocols in medium containing 1000U/ml Leukemia Inhibitory Factor (Millipore) with irradiated MEF feeders (GlobalStem) on gelatin-coated plates. ESCs cultured in iSTEM 3i media (Stem Cells Inc.) were passaged at least 6 times in serum-free, feeder-free conditions on gelatin-coated plates before being analysed. For immunofluorescence experiments, cells were grown either on gelatinized glass coverslips pre-seeded with MEF feeders for +LIF experiments, or gelatinized glass coverslips alone for 3i experiments. For RNA and protein isolation, ESCs were soaked twice for 1 hour on gelatin-coated plates to remove MEFs, then immediately processed. AB2.2 (129/SvEvBrd-Hprtb-m2) ESC line was kindly provided by A. Mills, CSHL; R1 (129X1 x 129S1) ESC and v6.5 (C57Bl/6 x 129/Sv) by S. Kim, CSHL; C57Bl/6 x Castaneous ESCs by C. Vakoc, CSHL; ZHBTc4 ESCs by A. Smith, Centre for Stem Cell Research, University of Cambridge, UK (Niwa et al., 2000); lmna-/- ESCs by C. Stewart, Institute of Molecular Biology, A*STAR, Singapore (Sullivan et al., 1999); nGFP2 iPSCs by R. Jaenisch, Whitehead Institute, MIT, USA (Wernig et al., 2008); WB6.1 (C57Bl/6 Cbrd/Cbrd/Cr), MK6 (C57B4) and CSH1 (129 x C57Bl/6) primary ESC lines were derived in-house and experiments performed within 8 passages. NPCs were derived from AB2.2 ESCs using a protocol adapted from (Conti et al., 2005). E3.5 blastocysts from normal female mice were kindly isolated and provided by S. Kim, CSHL and P. Jiang, CSHL. Blastocysts were fixed in 4% formaldehyde made fresh from paraformaldehyde and processed for immunofluorescence within 2 hours of collection.

RNA-Sequencing

10µg of total RNA was isolated using Trizol reagent (Ambion). PolyA+ RNA was isolated (Oligotex kit, Qiagen) and depleted of ribosomal RNA (ribominus kit). Stranded libraries were prepared using a protocol adapted from (Parkhomchuk et al., 2009) for paired-end sequencing on the Illumina GA IIx platform. Reads were mapped to the mouse mm9 reference genome using the Tophat spliced-read aligner and coverage computed using BedGraph2 (bedtools suite). Coverage tracks were uploaded to and visualized in the UCSC genome browser in BigWig format.

Real-time RT PCR

Total RNA was isolated using Trizol reagent (Ambion), treated with amplification grade RNAse-free DNAse I (Invitrogen) and converted to cDNA using random hexamer primers (Applied Biosystems RT reagents) using manufacturer's protocols. Quantitative real-time PCR was performed using SYBR green reagents (Applied Biosystems). Primer sequences are available upon request. Error bars represent standard deviation from three biological replicates normalized to the geometric mean of beta-actin, cycloB1 and pabpc1.

Antibodies, Western Blotting and Immunofluorescence

Rabbit anti-lamin A (323-11, (Dechat et al., 2007), 1:2,000) and rabbit-anti-lamin A/C (266, (Moir et al., 1994) 1:2,000) were kindly provided by R. Goldman, Northwestern University, Feinberg School of Medicine; mouse anti-lamin A/C (Active Motif 39287, 1:1,000); rabbit anti-laminB1 (abcam ab16048, 1:2,000); rabbit anti-Oct4 (Santa-Cruz sc9081, 1:2,000); rabbit anti-histone H3 (abcam ab1791, 1:10,000) were used. Western Blots were performed using anti-mouse or anti-rabbit secondary antibodies conjugated to HRP (1:10,000) and detected by ECL (Perkin-Elmer). For immunofluorescence, coverslips were fixed in 4% PFA for 20 minutes at room-temperature, permeablized in 0.5% Triton X-100 at 4 degrees for 5 minutes and blocked for 1 hour in 3% BSA. Blastocysts were fixed in 4% PFA for 30 minutes at room-temperature, permeabilised in 0.25% Triton X-100 for 15 minutes at room-temperature, then blocked in 10% FBS with 0.1% Triton X-100 for 1-3 hours at room-temperature. Cells were incubated with primary antibodies containing 1% BSA for either 1 hour at room temperature (Oct4 1:400, lamin B1 1:400) or overnight at 4 degrees (lamin A 1:200, lamin A/C 1:200). All primary antibody incubations for blastocysts were performed overnight. Overnight incubations are necessary to ensure efficient antibody accessibility into the tightly packed ESC colonies. Anti-mouse, anti-rabbit and anti-rat secondary antibodies conjugated to Alexa-488, Alexa-594 or Alexa-647 (Invitrogen) were used, and DAPI was used to counterstain DNA.

Imaging

Immunofluorescence imaging of single sections was performed using a Zeiss LSM710 laser scanning confocal microscope using a 63x 1.4N.A. oil-immersion objective using 405nm, 488nm, 594nm and 647nm lasers. Slides were mounted in antifade containing 10% glycerol and 1mg/ml p-Phenylenediamine (Sigma). Images represent single 0.1μm sections through the center of ESC colonies or individual blastocysts. No post-acquisition image processing was performed. Figures show representative single z-sections.

3.5 Acknowledgements

We thank R. D. Goldman, R. Jaenisch, S. Kim, A. Mills, A. Smith, C. Stewart, C. Vakoc for cell lines and reagents, S. Kim and P. Jiang for blastocyst isolations, C. Davis, J. Drenkow and T. Gingeras for assistance in RNA-Sequencing library preparation, S. Hearn for assistance in microscopy, G. Arun, M. Bodnar, M. Huebner, I. R. Kumaran, J. Li, C. Zepeda-Mendoza, B. Zhang of the Spector Laboratory, and R. D. Goldman and members of the Goldman Laboratory for discussions and comments. M. A. E-M is supported by Genentech Foundation Fellowship, and George A. and Marjorie H. Anderson Fellowship. J. H. B. is supported by a DAAD PostDoctoral Fellowship. This work was supported by the NIH, National Institute of General Medical Sciences (GM42694) to D. L. S.

Chapter 4

Identification and characterisation of monoallelically expressed genes

Bioinformatic analysis of the RNA-sequencing screen was performed through a collaboration with David Thybert, John Marioni and Paul Flicek at European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom. All other experimental results were generated by Mélanie Eckersley-Maslin with some help from Jan Bergmann (CSHL) for NPC sub-cloning and RNA-FISH analysis.

4.1 Allele-specific RNA-sequencing screen design

To identify in an unbiased manner monoallelically expressed genes, an allele-specific RNA-sequencing screen was performed. Allele information was assessed using ESCs from a F1 hybrid cross between the C57Bl/6J and CAST/EiJ mouse strains. This mouse strain has a high number of polymorphisms, with 82% of transcripts, corresponding to 78.8% of genes, containing at least 1 exonic single nucleotide polymorphism (SNP) and a median of 6 SNPs per transcript (figure 4.1 on page 83). Single cells from both ESCs and NPCs were isolated and expanded to generate 6 single-cell derived clones each. Assuming inheritance of monoallelic expression across cell divisions, all cells within each single-cell derived clone are expected to express the same combination of alleles. However, different clones should show random selection of alleles,

allowing the identification of mitotically inheritable random monoallelically expressed genes.

Figure 4.2 (page 84) shows a schematic representation of the screen design. Within each of the 6 ESC and 6 NPC single-cell derived clones, transcripts were classified as either monoallelic towards the C57Bl/6J allele (orange) or CAST/EiJ allele (blue), or biallelically expressed with equal number of reads originating from both alleles (orange + blue). Transcripts could also be classified as not-expressed or not-assessable (grey). Following classification of transcripts within clones, transcripts were then grouped into one of three classes based on their expression patterns across clones. Class A transcripts include the high confidence random monoallelically expressed transcripts. Class B transcripts were of lower confidence and were subjected to additional filtering to select additional high confidence monoallelic transcripts. Class C transcripts represent non-random monoallelically expressed genes.

RNA sequencing of poly(A)$^+$ purified RNA was performed and approximately 80 million reads were generated per clone, which were mapped to both the C57Bl/6J and CAST/EiJ transcriptomes using BWA (Li and Durbin, 2010), resulting in approximately 50 million mapped reads per clone. The C57Bl/6J transcriptome, containing 87,862 transcripts, was collected from ENSEMBL v59. The CAST/EiJ transcriptome was created by substituting the 423,197 exonic single nucleotide polymorphisms (SNP) between the C57BL/6J and CAST/EiJ genomes (Keane et al., 2011). A transcript was deemed expressed if it was in the top 80$^{\text{th}}$ percentile of the overall expression distribution, and assessable if there were at least 5 reads covering an informative SNP. For all expressed and assessable transcripts, the number of reads corresponding to each allele at each SNP position was used to determine whether there was evidence of allele-specific expression. Monoallelic expression was quantified using two metrics: a p-value calculated from a binomial test; and a d-score representing the ratio of allele expression. The p-value was calculated using a simple binomial distribution test, with the null hypothesis that there are equal number of reads mapping to both alleles for all SNPs in a given transcript. The d-score was determined by calculating the average weighted mean that each SNP within a transcript deviated from the expected 50:50 equal distribution. In this way SNPs with a higher coverage, in which there is less sampling noise, are given a higher weight. D-score values range between -0.5 and +0.5 with negative values corresponding to a bias towards the CAST/EiJ allele, and positive values corresponding to a bias towards the C57Bl/6J allele. A value of 0 reflects equal number of reads

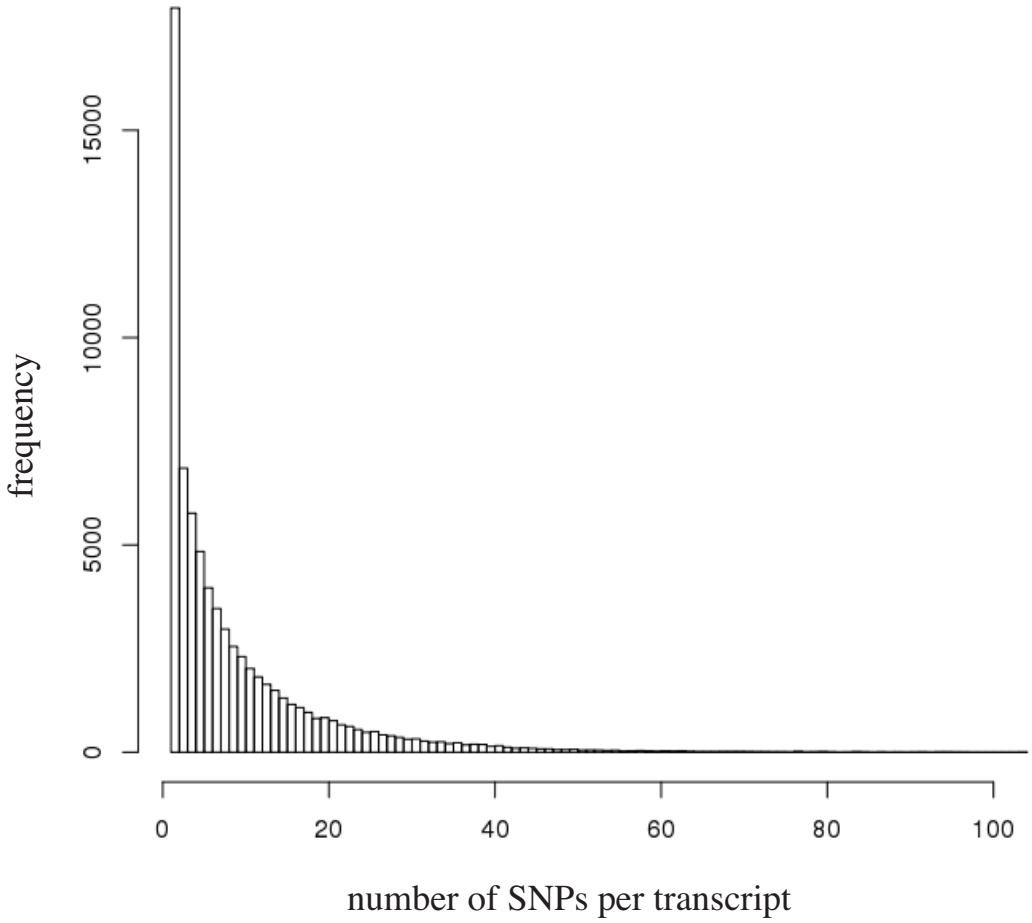


Figure 4.1: Distributions of exonic SNPs between C57Bl/6J and CAST/EiJ genomes. Histogram plot showing the distribution of frequency of SNPs within a transcript. 82% of transcripts contain at least one SNP. The median number of SNPs per transcript was 6.

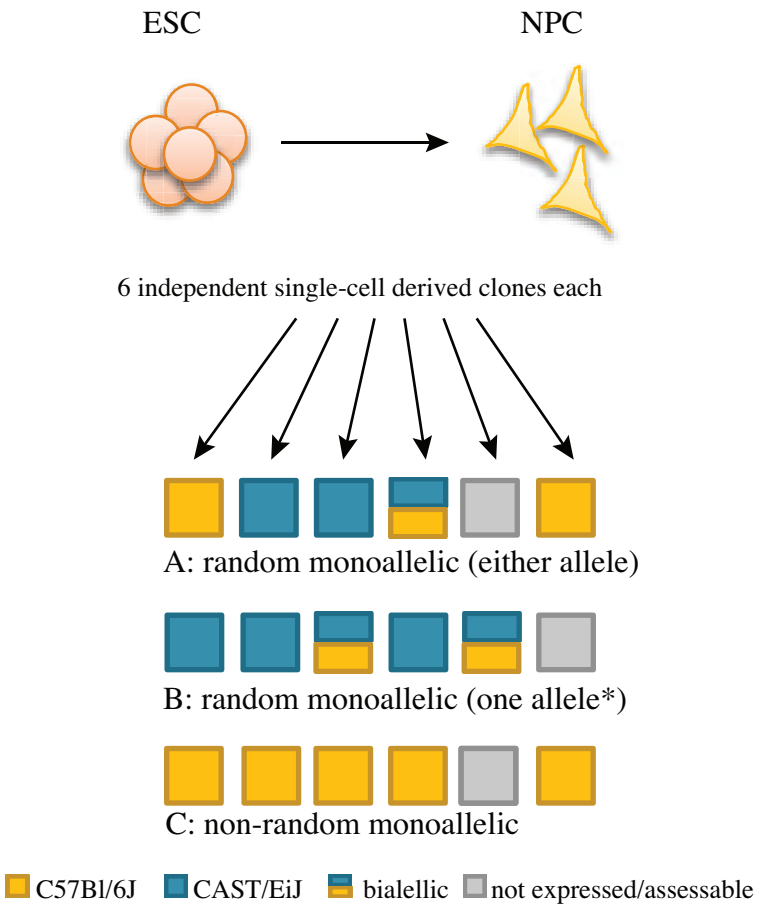


Figure 4.2: Schematic representation of the allele-specific RNA sequencing screen. C57Bl6/J x CAST/EiJ F1 hybrid ESCs were differentiated into NPCs and RNA from 6 single-cell derived clones from each cell type derived sequenced. Within each clone transcripts were classified as C57Bl/6J biased (orange), CAST/EiJ biased (blue), biallelic (orange + blue) or not expressed/assessable (grey). Based on patterns across clones, transcripts were grouped into three classes. Class A corresponding to random monoallelic with at least one clone biased to C57Bl/6J and one clone biased to CAST/EiJ. Class B corresponding to random monoallelic genes in which one clone was biased to either C57Bl/6J or CAST/EiJ. * Transcripts in this class were further filtered to include those in which there was evidence the second allele could be transcribed. Class C transcripts contain non-random monoallelic genes in which all clones are biased all towards the same allele.

from both alleles (biallelic expression). In this way, transcripts within each clone were classified into one of 6 classes based on the p-value and d-score values (figure 4.3 on the next page):

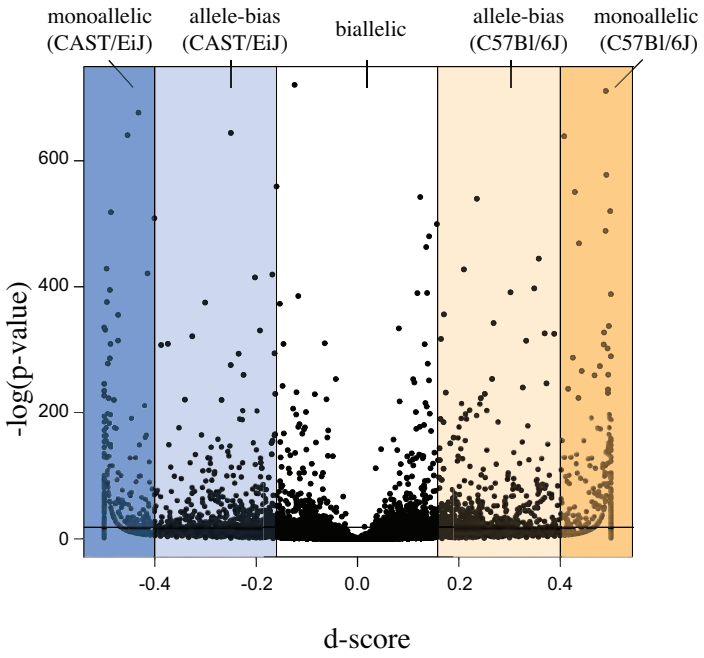
- Monoallelic: |d-score|$\geq$0.40, p-value $\leq 10^{-8}$
- Allele-biased: 0.18$\leq$|d-score|$<$0.40, p-value $\leq 10^{-8}$
- Biallelic: |d-score|$<$0.18 and/or p-value$>10^{-8}$
- Not expressed: expression is lower than 5.8 normalized reads per kilobase (NRPK)
- Not assessable: No SNV in the transcript with at least 5 reads coverage or the transcript is within a genomic region filtered for aneuploidy (see below)
- Other: all remaining transcripts

To control for possible loss of heterozygosity, mouse diversity SNP arrays were run on genomic DNA from each of the 6 NPC clones. At each SNP position, a BAF value, corresponding to the relative signal intensity of the C57Bl/6J allele versus CAST/EiJ allele, was calculated. Each chromosome was split into 1Mb sized bins containing an average of 85.5 SNPs per bin. The average BAF value for the bin was calculated and compared to an ESC polyclonal control sample in which the genomic integrity had been previously confirmed by karyotype analysis. Regions in which the distribution of BAF values differed significantly from the control were defined as aneuploid and classified as non-assessable in the screen pipeline (figure 4.4 on page 87). A total of 16 aneuploid regions were identified across the 6 NPC clones, and a subset confirmed by karyotype analysis.

Based on the classification of transcripts into the 6 categories described above, transcripts were assigned into one of three classes of monoallelically expressed genes. Those transcripts which were assigned to the 'not-expressed', 'non-assessable' and 'other' classes for a given clone were not included in the classification process.

- Class A contains transcripts where at least one clone showed allele specific expression (either monoallelic or allele-bias) of the C57Bl/6J allele and a second separate clone showed

A



B

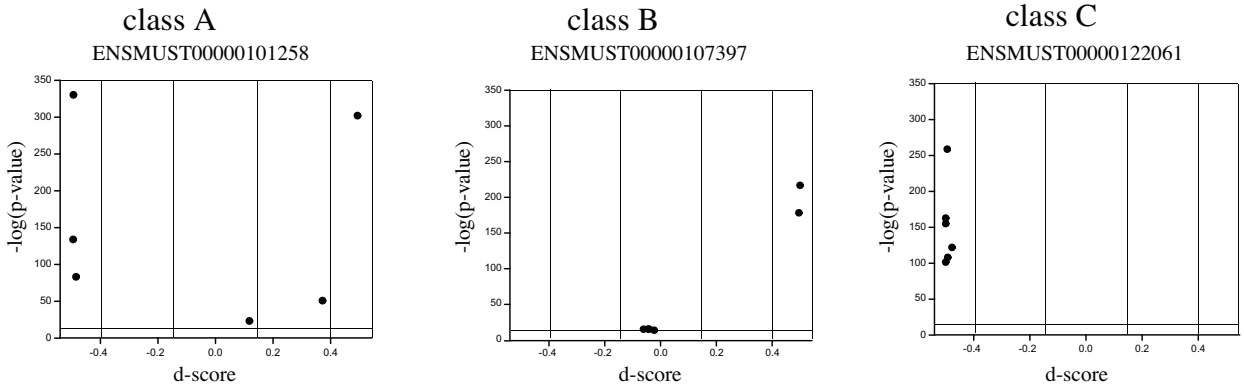


Figure 4.3: Classification of monoallelic transcripts.
(A) Classification of transcripts (black circles) within each clone based on d-score (x-axis) and p-value (y-axis). Transcripts could be biallelic (middle, white), monoallelic for CAST/EiJ allele (dark blue), allele-biased for CAST/EiJ allele (light blue), allele-biased for C57Bl/6J allele (light orange) or monoallelic for C57Bl/6J allele (dark orange). (B) examples of transcripts belonging to each of three classes based on transcript classification across clones. Class A contain transcripts in which one clone is biased towards C57Bl/6J allele, and one biased towards CAST/EiJ allele. Class B contain transcripts in which one clone is biased towards C57Bl/6J or CAST/EiJ allele, some clones remain biallelic. Class C contain non-random monoallelic genes in which all clones show bias towards the same allele.

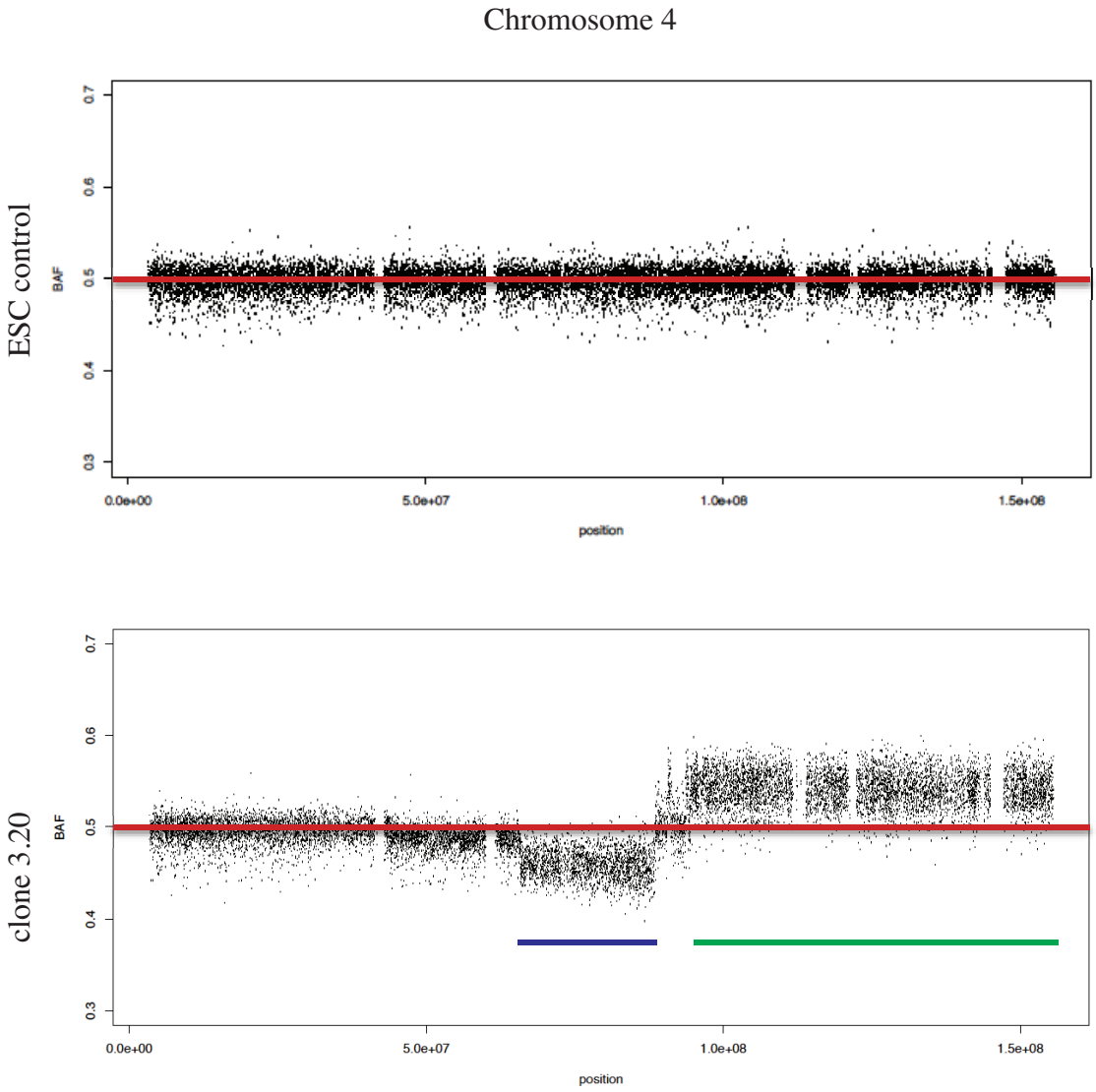


Figure 4.4: Identification of aneuploid regions with genomic SNP arrays.
Example of array data from chromosome 4 from ESC polyclonal control genomic DNA (top) and NPC clone 3.20 genomic DNA (bottom). Red line depicts BAF value of 0.5 corresponding to equal signal intensity coming from both C57Bl/6J and CAST/EiJ alleles. Each dot represents one SNP between C57Bl/6J and CAST/EiJ mouse strains. Control ESC polyploid sample shows a uniform distribution of BAF values surrounding 0.5. NPC clone 3.20 shows two aneuploid regions. One biased towards CAST/EiJ (blue bar) and the second biased towards C57Bl/6J (green bar). Transcripts within these regions were removed from subsequent analysis.

allele specific expression of the CAST/EiJ allele. Some clones may also be biallelic. Transcripts in this class represent high confident random monoallelic genes.

- Class B contains transcripts with at least one biallelic clone and at least one clone showing allele specific expression (either monoallelic or allele-bias) towards either the C57Bl/6J or CAST/EiJ allele. All clones showing allele-specific expression have the bias in the same direction. To select high confidence monoallelic genes, class B transcripts were further filtered using more stringent parameters to select those in which there was strong evidence that the second allele was transcribed in at least one clone. Filtered class B transcripts contain at least one clone with stringent allele specific expression (p-value$<10^{-10}$, |d-score|>0.35) and one clone with stringent biallelic expression (|d-score|<0.10) and represent additional high confidence random monoallelic genes.

- Class C contains transcripts in which allele specific expression was observed toward the same allele for all clones and no biallelic clones were observed. This class contains non-random monoallelically expressed genes, including imprinted genes and those biased due to strong cis genetic effects. Class C transcripts were not included in subsequent analysis.

The final set of random monoallelically expressed transcripts includes class A and filtered class B transcripts. Not all transcripts were necessarily assigned to one of the three classes above and genes could be present in more than one class if they were represented by transcripts classified in different classes.

4.2 Allele-specific RNA sequencing screen results

In mouse ESCs, 13,699 genes were expressed at sufficient levels and contained at least one SNP covered by at least 5 reads. Of these assessable genes, only one was classified as high-confidence class A (table 4.1 on page 90). An additional 66 filtered class B genes resulted in a total of 67 genes, or 74 transcripts, classified as random monoallelically expressed. This represents only 0.49% of assessable genes. Interestingly, this low number increased 5.6 fold during differentiation to 376 genes in NPCs (table 4.2 on page 90). This included 86 genes in class A and 302 genes in filtered class B, or 135 and 381 transcripts respectively, and represents 3.00% of assessable genes.

This increase in monoallelic expression during differentiation suggests that the establishment of monoallelic expression occurs upon cell-fate specification early in development. As expected, the expressed imprinted genes were correctly classified into class C. Furthermore, 11 protocadherins were included within the 376 NPC random monoallelically expressed genes, providing a good internal positive control for the screen analysis.

Notably, only 4 of the 376 random monoallelically expressed genes in NPCs (*Cyp7b1*, *Npl*, *Plin2* and *Rgs16*) showed random monoallelic expression in all assessable clones. Therefore for the remaining 98.9% of NPC monoallelically expressed genes, at least one clone was either biallelic and/or did not express the respective gene. This contrasts with imprinted and X-chromosome inactivated genes, where all cells exhibit strict monoallelic expression, and implies that, rather than being tightly regulated, random monoallelic expression is likely not an active decision required for cell survival or differentiation. Within a single clone, approximately 60% of the monoallelically expressed genes showed biallelic expression for both ESC and NPC (figure 4.5 on page 91), supporting the hypothesis that monoallelic expression is not a requirement for these genes but may reflect variation in gene expression regulation between two homologous alleles. Interestingly, if an expressed transcript was not biallelic in ESCs, it was predominantly allele-biased. In contrast, non-biallelic transcripts were predominantly monoallelically expressed in NPCs (figure 4.5). This observation, along with the reduced frequency of inheritable monoallelically expressed genes in ESCs compared to NPCs, is possibly a reflection of the increased plasticity of pluripotent ESCs compared to the lineage committed NPCs (Mattout and Meshorer, 2010).

Table 4.1: Summary of random monoallelically expressed genes in ESCs.

	Genes	Transcripts	% Genes
Class A	1	1	~
Class B	1665	2499	12.15
Class B filtered	66	73	0.48
Class C	547	721	3.99
Total Random Monoallelic (A+B filtered)	67	74	0.49

Table 4.2: Summary of random monoallelically expressed genes in NPCs.

	Genes	Transcripts	% Genes
Class A	86	135	0.69
Class B	1874	3004	14.95
Class B filtered	302	466	2.41
Class C	276	381	2.20
Total Random Monoallelic (A+B filtered)	376	602	3.00

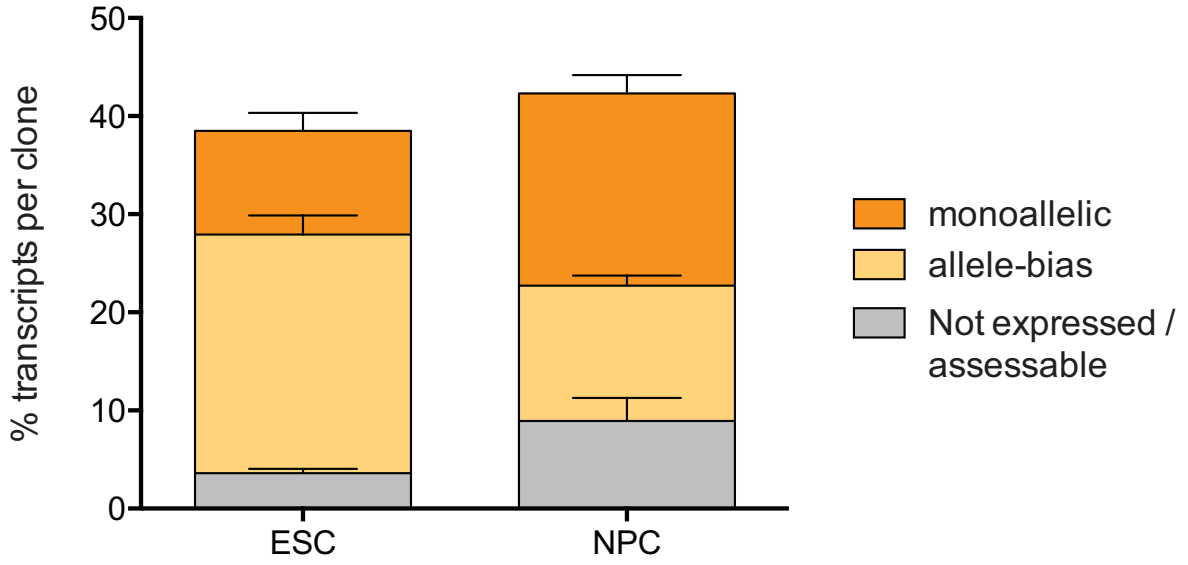


Figure 4.5: Status of monoallelically expressed genes within individual clones. Quantification of ESC and NPC monoallelic transcripts as monoallelic (dark orange), allele-biased (light orange) or not expressed/assessable (grey). Bars represent average ± standard deviation amongst 6 clones. ESC monoallelically expressed genes are predominantly allele-biased whereas NPC monoallelically expressed genes are predominantly monoallelically expressed.

4.3 Validation of monoallelically expressed genes

Three separate approaches were performed to validate the monoallelically expressed genes identified in the allele-specific RNA-sequencing screen. First, SNP-PCR, in which Sanger sequencing of PCR products containing informative exonic SNPs was performed, provided a fast method of verifying allele-specific expression in a population of cells. Secondly, RNA-polymerase II Chromatin Immunoprecipitation (ChIP) allowed molecular verification as to the transcriptional status of the alleles in a population of cells, independent of RNA analysis. Finally, combined RNA-DNA FISH, in which both the gene locus and transcriptional activity of the respective locus could be visualised, provided assessment of monoallelic expression at single-cell resolution.

SNP-PCR was performed on a subset of randomly selected monoallelically expressed genes from category A, B and C in both ESCs and NPCs. Validation of genes as monoallelically expressed through this method was invaluable in fine-tuning the p-value and d-score thresholds used in the allele-specific RNA-sequencing screen. RNA from each of the clones was isolated, cDNA synthesised and PCR performed to amplify exonic and/or intronic SNPs. Sanger sequencing was then performed on the PCR products and the expressed alleles deduced from analysis of the sequencing traces. Primers and SNPs were confirmed either on genomic DNA or cDNA generated from polyclonal population of cells in which both alleles are expected to be expressed. Following analysis of the sequencing traces, clones were classified as monoallelic or biallelic (figure 4.6 on the following page) prior to being compared to the RNA-sequencing screen results.

A total of 82 SNP-PCR products analysing 20 separate genes were assessed (figure 4.7 on page 95). This included 18 genes in NPCs and 2 genes in ESCs. Of the 18 NPC genes, 11 were from class A, 2 from class B, 2 from class C, and 2 were biallelic. Both of the ESC verified genes were from class B. For each gene, the results for the SNP-PCR and RNA-seq analysis are shown (top and bottom rows respectively) across assessed clones. 76 of the 82 SNP-PCR assays were in agreement with the RNA-sequencing screen, giving a 93% validation rate. Of those that did not validate, 2 gene-clone combinations had a |d-score| very close to the cutoff used of 0.180, resulting in them being classified as monoallelic in the RNA-sequencing screen and biallelic by SNP-PCR analysis, or vice-versa. Nevertheless, overall there was excellent agreement

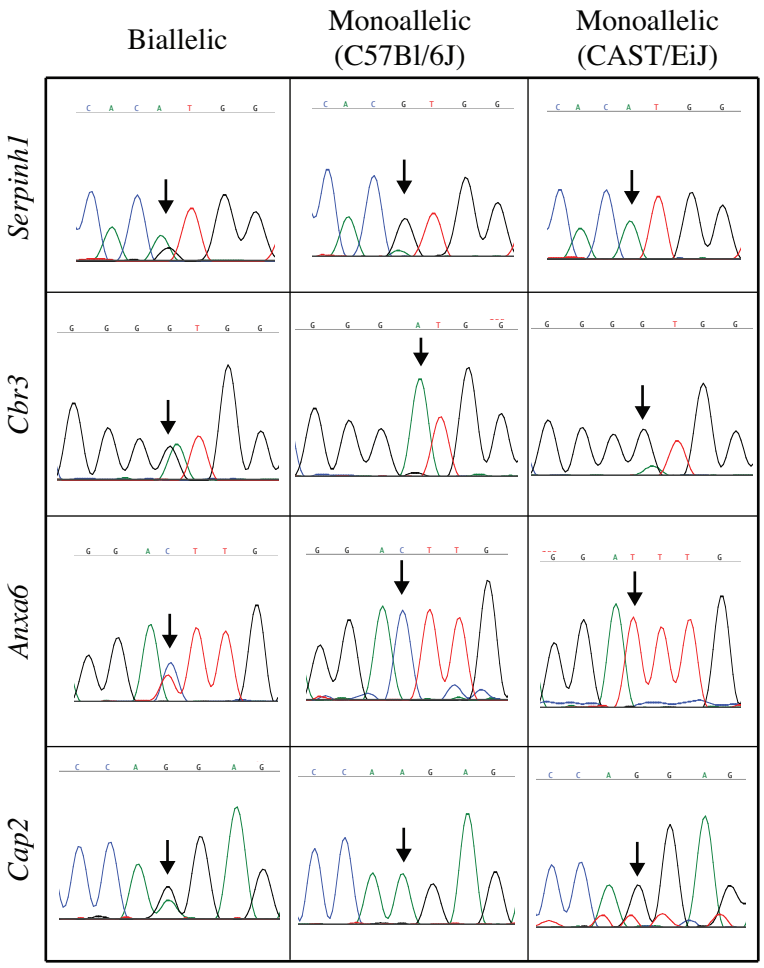


Figure 4.6: Selected SNP-PCR validation of monoallelically expressed genes.
Example of SNP-PCR traces for four separate monoallelically expressed genes: *Serpinh1*, *Cbr3*, *Anxa6* and *Cap2*, showing an example of a biallelic clone (left column), C57Bl/6J monoallelic clone (middle column) and CAST/EiJ monoallelic clone (right column). Arrows denote position of SNP. Blue = C, green = A, red = T, black = G.

between the RNA-sequencing results and SNP-PCR validation demonstrating the robustness of the approach.

RNA polymerase II ChIP was used to confirm that the monoallelic expression observed was indeed occurring at the level of transcription. Chromatin immunoprecipitation (ChIP) was performed using the 8WG16 monoclonal antibody (Thompson et al., 1989) that detects all forms of RNA polymerase II. Levels of RNA polymerase II pull-down were similar between monoallelic and biallelic clones within the body of four randomly selected genes (figure 4.8, three genes are shown). Due to informative SNPs within the amplicons, it was possible to distinguish which alleles were associated with RNA polymerase II. Importantly, for three of four genes tested, RNA polymerase II was specifically associated with only the active allele in monoallelic clones, compared to both alleles in biallelic clones (figure 4.8), supporting the findings that monoallelic expression is due to the exclusive transcription of only one of the two alleles in the cell.

Finally, monoallelically expressed genes were validated by performing combined RNA-DNA Fluorescent *In Situ* Hybridisation (FISH). By using fluorescently labeled probes targeting both exonic and intronic sequences of the target gene, the nascent RNA at the sites of transcription could be seen microscopically as a single fluorescent spot within the nucleus (figure 4.9 on page 97). These RNA-FISH spots co-localise with the gene locus visualized by subsequent DNA-FISH in the same cells, confirming that they are indeed sites of transcription. 6 out of 6 separate genes were confirmed as monoallelically expressed by RNA-FISH analysis (figure 4.9).

The percentage of expressing cells exhibiting monoallelic or biallelic expression was determined by counting RNA-FISH signals across a large number of cells (n$\geq$100) (figure 4.10 on page 99). For example, *Ror2* was detected by RNA-FISH in 83-90% of cells across experiments. Of these expressing cells, there was expression from only 1 allele in 84.3% of cells in a monoallelic clone, compared to only 27.0% of cells in a biallelic clone. Similarly, sites of transcription of *Acot1* were detected in 48% of cells of a monoallelic clone, of these cells, 89.6% showed transcription from only one of the two alleles. Likewise, *Acyp2* signal was detected by RNA-FISH in 62% and 73% of cells in a monoallelic and biallelic clone respectively. Of these expressing cells, 93.5% showed only 1 active allele in a monoallelic clone, confirming monoallelic expression in these cells. In contrast 57.5% of cells in a biallelic clone showed 2 active alleles as expected. Furthermore, expression from 1 allele was confirmed for 96.5% and 93.3% of expressing cells

A

Gene	Class	Assay	Clone 1	Clone 2	Clone 3	Clone 4	Clone 5	Clone 6
anxa6	A	SNP-PCR	biallelic	CAST/EiJ	biallelic		biallelic	C57Bl/6J
		RNA-Seq	biallelic	CAST/EiJ	biallelic		biallelic	C57Bl/6J
cap2	A	SNP-PCR		CAST/EiJ	C57Bl/6J	C57Bl/6J	C57Bl/6J	
		RNA-Seq		CAST/EiJ	C57Bl/6J	C57Bl/6J	C57Bl/6J	
cbr3	A	SNP-PCR	biallelic	C57Bl/6J	**C57Bl/6J**	biallelic	CAST/EiJ	C57Bl/6J
		RNA-Seq	biallelic	C57Bl/6J	**biallelic**	biallelic	CAST/EiJ	C57Bl/6J
fam111a	A	SNP-PCR			C57Bl/6J	C57Bl/6J		biallelic
		RNA-Seq			C57Bl/6J	C57Bl/6J		biallelic
fkbp7	A	SNP-PCR	biallelic			CAST/EiJ	biallelic	biallelic
		RNA-Seq	biallelic			CAST/EiJ	biallelic	biallelic
gas6	A	SNP-PCR	CAST/EiJ	**biallelic**	biallelic	CAST/EiJ	**biallelic**	biallelic
		RNA-Seq	CAST/EiJ	**C57Bl/6J**	biallelic	CAST/EiJ	**CAST/EiJ**	biallelic
npl	A	SNP-PCR	CAST/EiJ		CAST/EiJ	CAST/EiJ	C57Bl/6J	CAST/EiJ
		RNA-Seq	CAST/EiJ		CAST/EiJ	CAST/EiJ	C57Bl/6J	CAST/EiJ
pla2g7	A	SNP-PCR	CAST/EiJ	C57Bl/6J	CAST/EiJ	CAST/EiJ		C57Bl/6J
		RNA-Seq	CAST/EiJ	C57Bl/6J	CAST/EiJ	CAST/EiJ		C57Bl/6J
rgs16	A	SNP-PCR				CAST/EiJ	**biallelic**	
		RNA-Seq				CAST/EiJ	**CAST/EiJ**	
sgsm1	A	SNP-PCR		C57Bl/6J				biallelic
		RNA-Seq		C57Bl/6J				biallelic
tubb2a	A	SNP-PCR	biallelic	biallelic	CAST/EiJ	C57Bl/6J	CAST/EiJ	CAST/EiJ
		RNA-Seq	biallelic	biallelic	CAST/EiJ	C57Bl/6J	CAST/EiJ	CAST/EiJ
scd2	B	SNP-PCR	biallelic	biallelic	biallelic	biallelic	biallelic	CAST/EiJ
		RNA-Seq	biallelic	biallelic	biallelic	biallelic	biallelic	CAST/EiJ
rhoj	A+B	SNP-PCR	biallelic		CAST/EiJ	CAST/EiJ	biallelic	**C57Bl/6J**
		RNA-Seq	biallelic		CAST/EiJ	CAST/EiJ	biallelic	**biallelic**
serpinh1	B	SNP-PCR	CAST/EiJ	CAST/EiJ		biallelic	CAST/EiJ	biallelic
		RNA-Seq	CAST/EiJ	CAST/EiJ		biallelic	CAST/EiJ	biallelic
klhl22	C	SNP-PCR	C57Bl/6J				C57Bl/6J	
		RNA-Seq	C57Bl/6J				C57Bl/6J	
retsat	C	SNP-PCR	CAST/EiJ	CAST/EiJ	CAST/EiJ	CAST/EiJ	CAST/EiJ	CAST/EiJ
		RNA-Seq	CAST/EiJ	CAST/EiJ	CAST/EiJ	CAST/EiJ	CAST/EiJ	CAST/EiJ
bai2	biallelic	SNP-PCR			biallelic	biallelic		
		RNA-Seq			biallelic	biallelic		
tgds	biallelic	SNP-PCR	biallelic	biallelic				
		RNA-Seq	biallelic	biallelic				

B

Gene	Class	Assay	Clone 1	Clone 2	Clone 4
mcts2	B	SNP-PCR	**CAST/EiJ**	biallelic	CAST/EiJ
		RNA-Seq	**biallelic**	biallelic	CAST/EiJ
sema3c	B	SNP-PCR	biallelic	C57Bl/6J	C57Bl/6J
		RNA-Seq	biallelic	C57Bl/6J	C57Bl/6J

biallelic
C57Bl/6J monoallelic
CAST/EiJ monoallelic

Figure 4.7: Summary of SNP-PCR validation of monoallelically expressed genes. Paired results for monoallelically expressed genes across the NPC (A) and ESC (B) clones. Top row represents SNP-PCR result, bottom row result from RNA-sequencing screen. Only gene-clone combinations that were assessable and expressed and had a corresponding SNP-PCR trace are shown. Fields are colored according to the classification as biallelic (green), C57Bl/6J biased (orange) or CAST/EiJ biased (blue). Red boxes outline discrepancies between RNA-sequencing and SNP-PCR results.

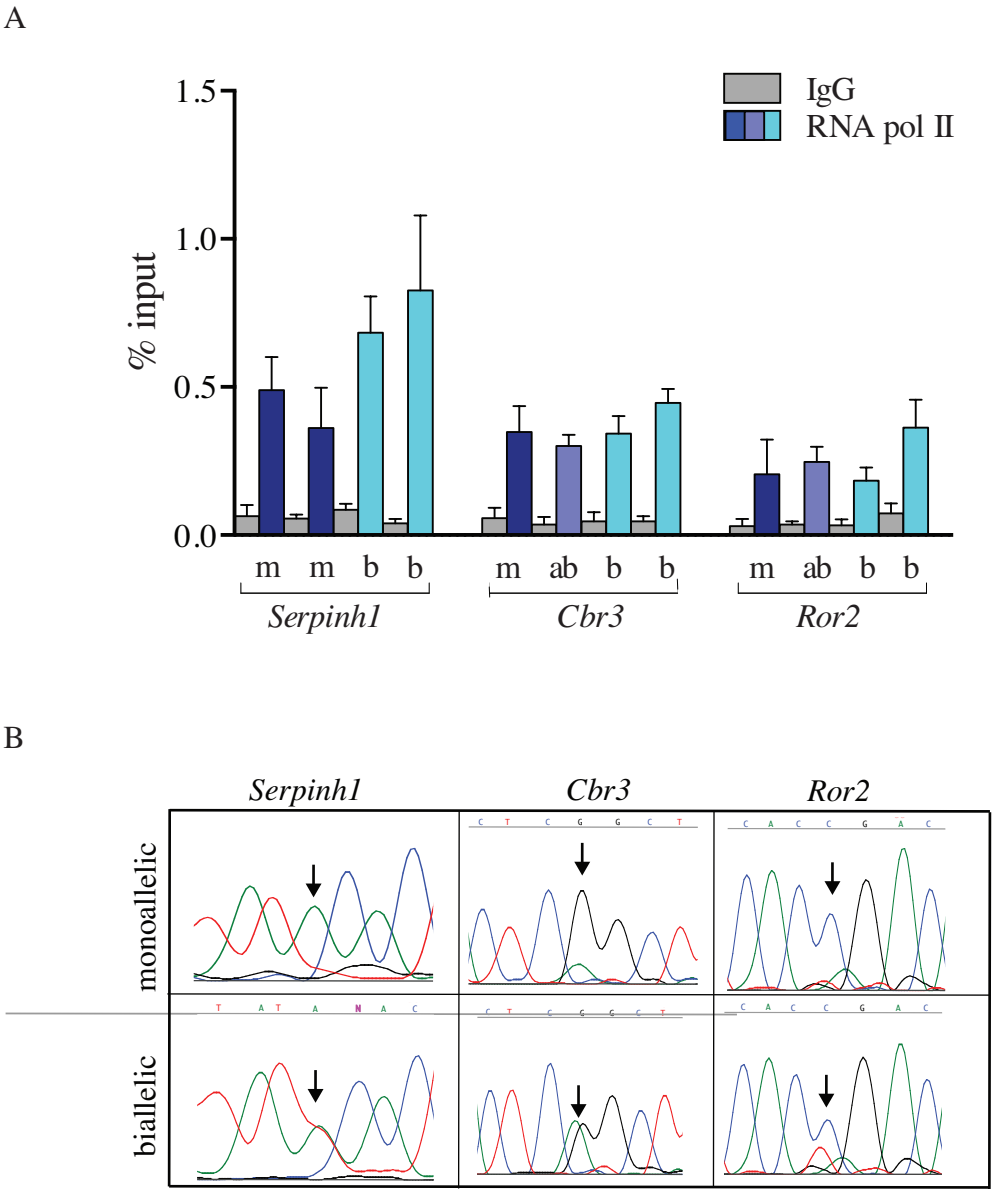


Figure 4.8: RNA polymerase II ChIP.
(A) Chromatin immunoprecipitation (ChIP) for RNA polymerase II large subunit (blue) or control IgG (grey) for three separate gene promoter regions between monoallelic (m, dark blue), allele-bias (ab, medium blue) or biallelic (b, light blue) clones. Error bars represent SEM of at least 3 biological replicates. (B) Representative traces of Sanger sequencing of ChIP products containing informative SNP (arrow) revealing associated alleles for monoallelic and biallelic clones.

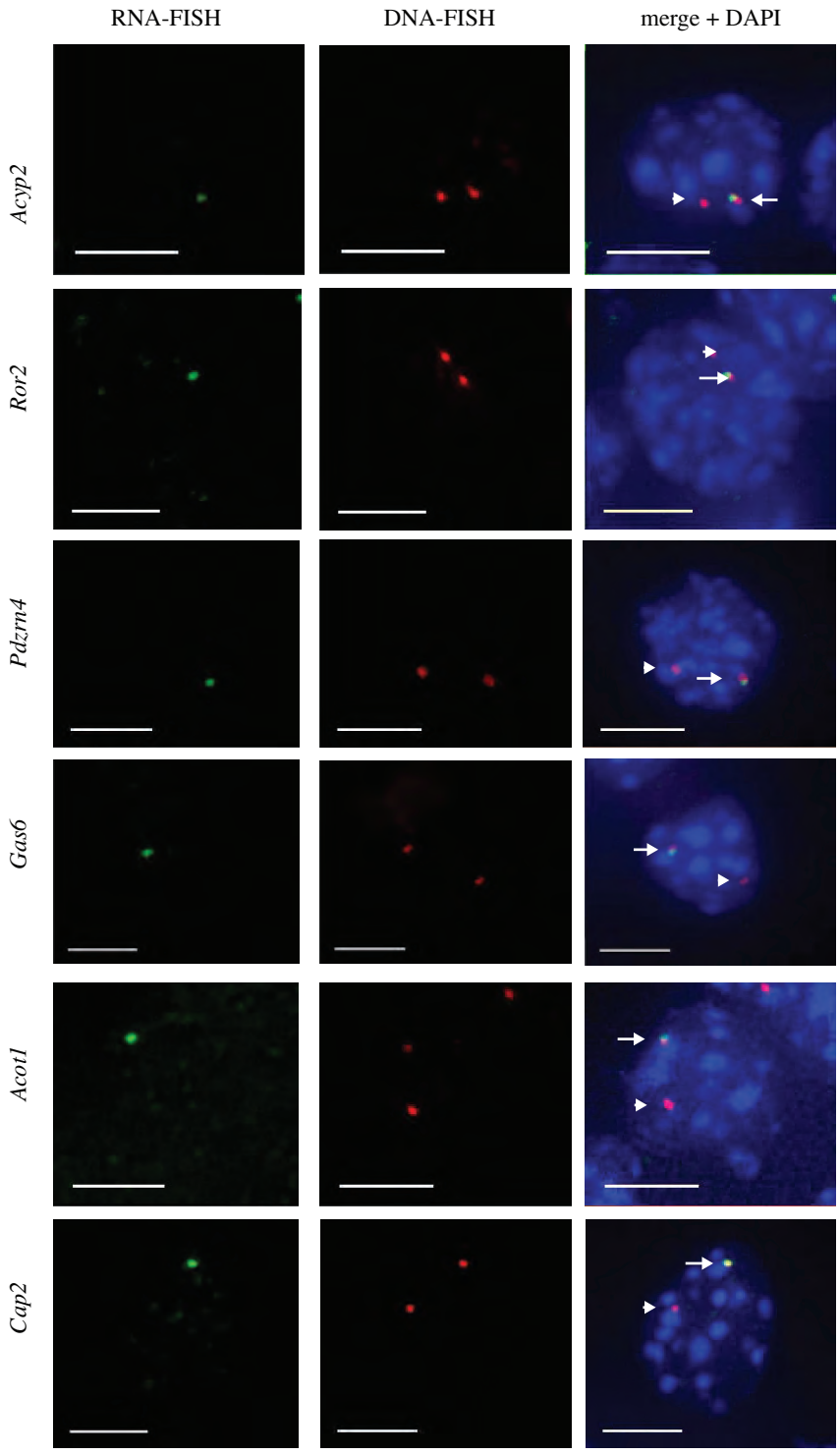


Figure 4.9: RNA-DNA FISH validation of monoallelically expressed genes. Representative 3D projections of RNA-FISH (left, green) and DNA-FISH (middle, red) image stacks. Right column shows a merge of RNA- and DNA-FISH with DAPI to visualize total DNA (blue). Arrows denote actively transcribing alleles, arrowheads inactive alleles. Scale bar represents 5μm.

in monoallelic clones for *Pdzrn4* and *Gas6* respectively. In this way, the RNA-FISH analysis validated precisely at single cell resolution the results by RNA-sequencing analysis, SNP-PCR and RNA polymerase II ChIP.

RNA-FISH analysis also proved valuable to understand the expression in clones exhibiting allele-bias on a population level where a smaller contribution of expression was observed from the second allele. Two possible interpretations could explain these allele-biased clones. Either both alleles are expressed in every cell although at different levels, or these allele-bias clones are a mixture of monoallelic and biallelic cells. Because the clones are derived from a founding single cell, this second scenario would suggest that, although mitotically stable, there may be a small frequency of reversion to biallelic expression. This reversion would be extremely rare, however, and would need to happen at a very early stage of clone expansion, as we do not generally observe loss of monoallelic expression, even with over 15 passages in culture. Based on the single-cell RNA-DNA FISH analysis, allele biased clones likely result from a combination of both processes within these clones.

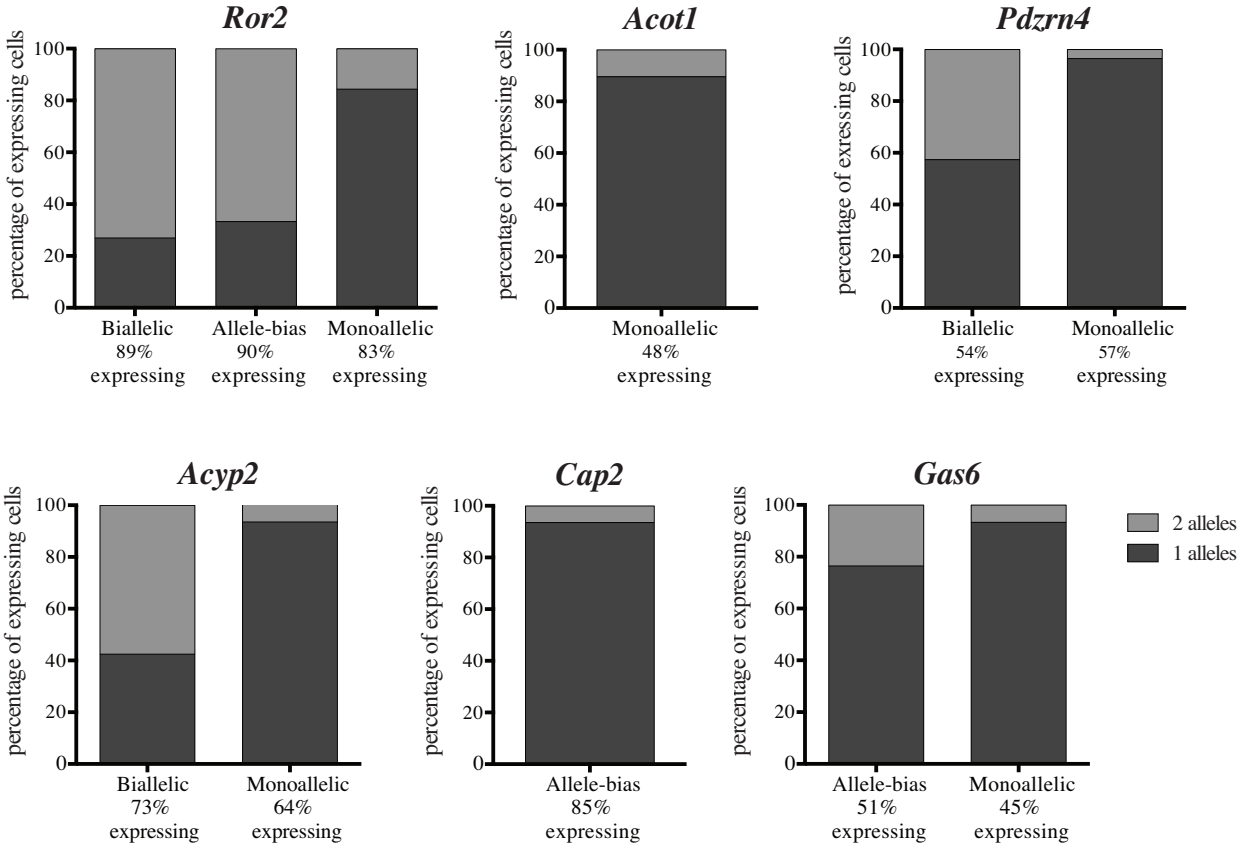


Figure 4.10: Quantification of RNA-DNA FISH analysis of monoallelically expressed genes. Quantification of independent clones for each of the 6 genes for NPC clones that were either biallelic, allele-biased or monoallelic for the respective gene. Percentage of expressing cells having either 1 (dark grey) or 2 (light grey) RNA-FISH signals representing monoallelic and biallelic cells respectively. At least 100 cells were analyzed per sample.

4.4 Genomic characteristics of monoallelically expressed genes

Next, the genomic characteristics of the random autosomal monoallelically expressed genes identified in the allele-specific screen were determined. First the level of expression of these genes was examined and compared to the expression level of all assessable genes. To assess the level of expression at the gene level, the average of expression level of all isoforms of the respective gene across all clones was calculated. Expression levels of the random monoallelic genes were compared to the set of assessable genes in both ESCs and in NPCs using a two-sided t-test. Importantly, the distribution of expression levels of the monoallelically expressed genes was not dramatically different from all assessable transcripts for both ESCs and NPCs (figure 4.11 on the following page). While a small statistically significant difference was seen in the expression level for NPCs, the difference is modest and unlikely to be of biological significance. There was no statistically significant difference in the expression level for monoallelically expressed genes versus all assessable genes in ESC, nor between ESC and NPC monoallelically expressed genes. This indicates that monoallelic expression is not a feature limited to poorly expressed genes, but can also occur for highly expressed genes.

The allele-specific RNA-sequencing screen used very stringent thresholds for defining whether a gene was expressed. To control whether these stringent thresholds may have removed some poorly expressed genes from the analysis, the expression level cutoff was varied and list of monoallelically expressed genes compared (figure 4.12 on page 103). The original threshold used to define whether a transcript was expressed, was 5.8 Normalised Reads Per Kilobase (NRPK). This threshold is equal to the average expression of the 20^{th} percentile of expression values across all clones. Reducing the threshold to 2.14 NRPK, corresponding to the 10^{th} percentile of expressed transcripts, or 1.08 NRPK, corresponding to the 5^{th} percentile of expressed transcripts, did not significantly change the outcome of the analysis. In ESCs, this resulted in an increase from 67 monoallelically expressed genes to 68 genes, and in NPCs, this resulted in an increase from 376 to 377 monoallelically expressed genes. Moreover, the proportion of all assessable genes that are monoallelically expressed did not change significantly. The influence of increasing the threshold to 10 NRPK, a common threshold used to define expressed transcripts, was also examined. This resulted in a loss of 10 monoallelically expressed genes in ESCs, and

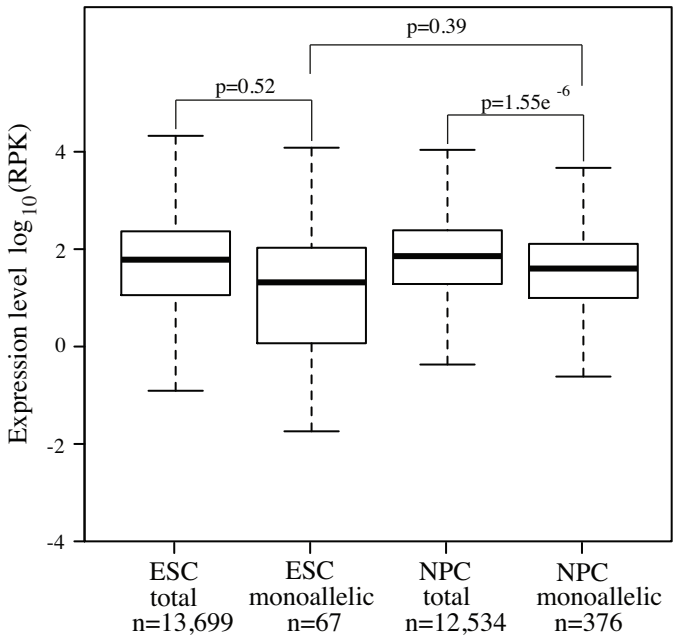


Figure 4.11: Expression level distribution of monoallelically expressed genes.
Expression level (normalised reads per kilobase, NRPK) distribution of all assessable genes (total) versus monoallelically expressed genes in both ESCs and NPCs. The list of assessable genes is defined as the genes which are expressed in at least 2 clones.

24 monoallelically expressed genes in NPCs, but again, the proportion of all assessable genes did not change significantly. Thus, the thresholds used to define expression had no noticeable impact on monoallelically expressed genes, and these genes are representative of all genes in terms of their expression levels.

Other forms of monoallelic gene expression, including imprinting and olfactory receptor allelic exclusion, occur in distinct domains or clusters in the genome. For example, more than 80% of the 144 genes imprinted in mouse are clustered into 16 distinct genomic regions containing 2 or more imprinted genes (Barlow, 2011). For these reasons, the genomic localisation of the monoallelically expressed genes in both ESCs and NPCs was assessed using the ENSEMBL karyotype viewer (Flicek et al., 2013). The monoallelically expressed genes were randomly distributed throughout the genome and did not fall into any clusters, nor were they preferentially located close to telomeres or centromeres (figure 4.13 on page 104). This is consistent with other reports (Gimelbrant et al., 2007; Zwemer et al., 2012), and distinguishes random autosomal monoallelic expression from other forms of monoallelic gene expression. The random autosomal monoallelically expressed genes were also similar to all assessable genes in terms of gene features, including the number of isoforms per gene and the number of exons (figure 4.14 on page 105). We also assessed the CG content at the promoters of the monoallelically expressed genes (figure 4.15 on page 106). AT rich (or CG poor) promoters have been previously reported as a signature of olfactory receptor gene promoters (Clowney et al., 2011). However there was no difference in the CG content of monoallelically expressed gene promoters compared to all assessable gene promoters (figure 4.15), suggesting that the AT rich promoter may be specific to olfactory receptors and not a general feature of monoallelically expressed genes.

Gene ontology analysis was performed on the monoallelically expressed genes in NPCs (figure 4.16 on page 108) using the tool DAVID (Huang et al., 2009). Analysis was performed using three different gene lists as a background reference, which enrichment of the monoallelically expressed genes was compared to. When the whole genome was used as background (figure 4.16A), there was a significant enrichment for cell adhesion, biological adhesion and membrane proteins. To control that these genes were enriched specifically in the monoallelically expressed genes and not in NPCs in general, gene ontology analysis was also performed using either the union or intersection of expressed genes as a background. The choice of background gave different results.

		ESC			NPC		
		Genes	**Transcript**	**% Genes**	**Genes**	**Transcript**	**% Genes**
5 th percentile (1.08 NRPK)	class A	1	1	~	86	136	0.63
	class B	1679	2529	11.04	1918	3063	14.15
	class B filtered	67	74	0.44	303	468	2.24
	class C	546	718	3.59	254	348	1.87
	Total A+B filtered	**68**	**75**	**0.45**	**377**	**604**	**2.78**
10 th percentile (2.14 NRPK)	class A	1	1	~	86	136	0.65
	class B	1679	2529	11.35	1910	3053	14.37
	class B filtered	67	74	0.45	303	468	2.28
	class C	546	718	3.69	258	356	1.94
	Total A+B filtered	**68**	**75**	**0.46**	**377**	**604**	**2.84**
20 th percentile (5.80 NRPK)	class A	1	1	~	86	136	0.69
	class B	1665	2499	12.15	1874	3004	14.95
	class B filtered	66	73	0.48	302	466	2.41
	class C	547	721	3.99	276	381	2.20
	Total A+B filtered	**67**	**74**	**0.49**	**376**	**602**	**3.00**
10 NRPK	class A	1	1	~	86	135	0.72
	class B	1618	2424	12.66	1656	2612	13.94
	class B filtered	56	61	0.44	277	427	2.33
	class C	548	731	4.29	308	424	2.59
	Total A+B filtered	**57**	**62**	**0.45**	**352**	**562**	**2.96**

Figure 4.12: Effect of varying expression level thresholds on monoallelically expressed genes. Number of monoallelically expressed transcripts and corresponding genes in ESCs and NPCs, and percentage of all assessable genes, when four separate expression level thresholds were used. The original threshold used corresponded to the 20thpercentile (5.8 NRPK) of expression values. Reducing this threshold to the 10th(2.14 NRPK) or 5th(1.08 NRPK) percentile, or increasing it to 10 reads per kb, did not significantly change the number or percentage of monoallelically expressed genes.

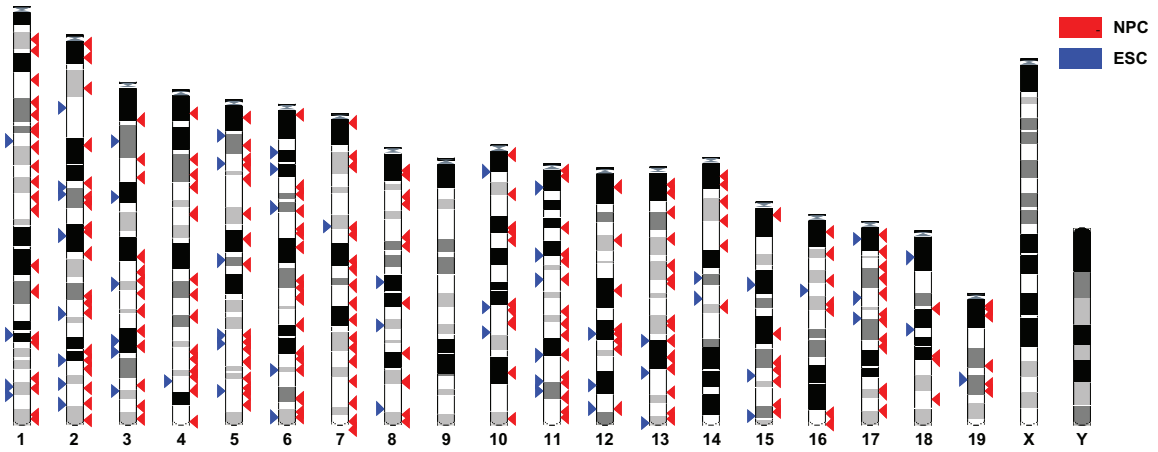


Figure 4.13: Genomic localisation of monoallelically expressed genes.
Genomic location of ESC (blue triangles) and NPC (red triangles) monoallelically expressed genes across all chromosomes. Genes do not fall into any distinct clusters.

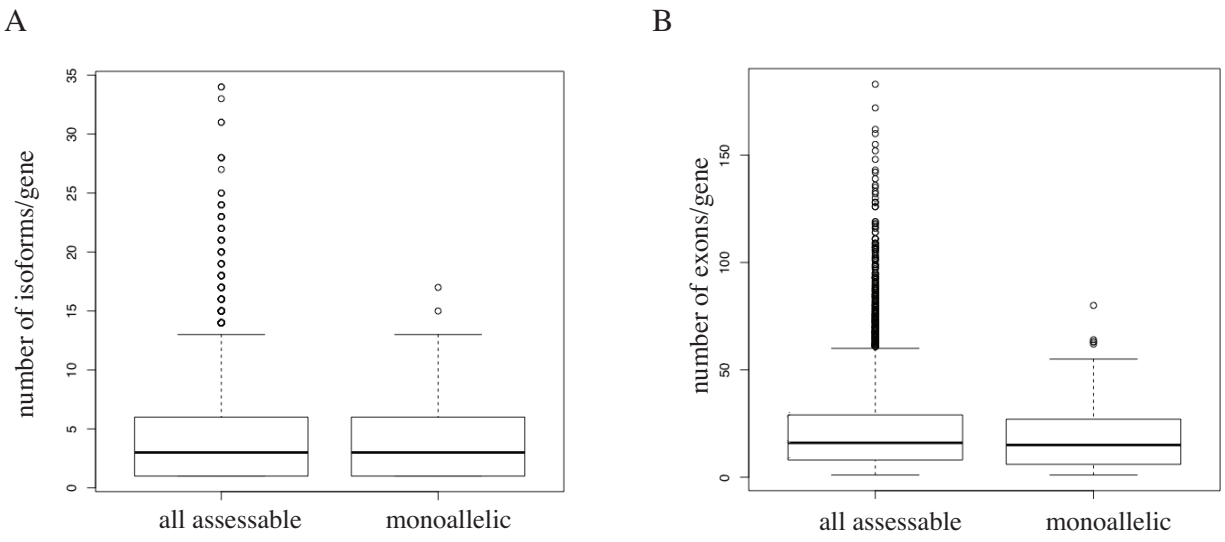


Figure 4.14: Gene structure of monoallelically expressed genes.
(A) number of isoforms per gene for all assessable genes or monoallelically expressed genes in NPCs. (B) number of exons per gene for all assessable genes or monoallelically expressed genes in NPCs. Distributions are not statistically significant (p-value > 0.5).

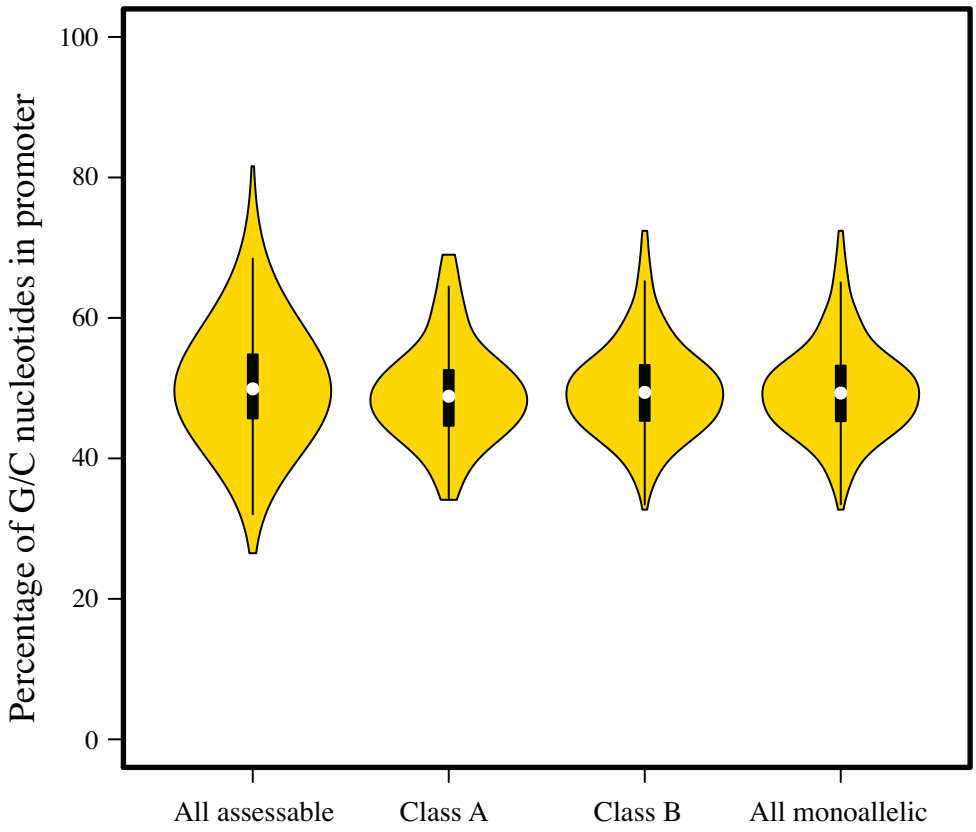


Figure 4.15: GC content of monoallelically expressed gene promoters.
Percentage of G/C nucleotides in the 1kb region surrounding the transcription start site of all assessable, class A, class B and all monoallelically expressed genes in NPCs. The distribution is centered around 50% and is not significantly different between the different groups of genes.

When comparing to the union of expressed genes, that is genes which are expressed in at least one NPC clone, cell adhesion , signal peptides and membrane were still enriched. However, comparing with the intersection of expressed transcripts, that is genes which are expressed in all NPC clones, cell adhesion and membrane ontologies are no longer enriched. Instead there is a small enrichment for signal peptides, glycoproteins and disulfide bonds, although the significance is low with false discovery rates between 0.02 and 0.03. Therefore there is a small enrichment of cell adhesion, signal and membrane proteins, although this feature of monoallelically expressed genes should not be over emphasised.

Interestingly, monoallelically expressed genes appear to be under less evolutionary constraint, when compared to all assessable genes (figure 4.17A on page 109). Genomic Evolutionary Rate Profiling (GERP) of both promoter and exonic sequences of the monoallelically expressed genes was performed to determine the extent of evolutionary constraint of these sequences. This measures the level of constraint at each base position by comparing the rate of nucleotide exchange to what would occur at a neutral sequence not under selection. A higher GERP score corresponds to a higher level of constraint. At a GERP threshold of 1.4, the exonic sequences of monoallelically expressed genes showed significantly less constraint than all assessable genes, indicating that they are under less evolutionary constraint, or more rapidly evolving. This was also seen at the promoter sequences, defined as 1kb sequence surrounding the transcription start site to a smaller yet still statistically significant extent. The difference was seen at all GERP thresholds. However, it is unclear whether the decreased evolutionary constraint seen is a real feature of monoallelically expressed genes, or if it is a consequence of the screen design. The allele-specific RNA-sequencing screen requires adequate coverage of exonic SNPs to provide allele information. A transcript with a higher number of exonic SNPs would be assigned a more significant p-value, and thus be more likely to be classified as monoallelically expressed, than a transcript with only one SNP, for example. Indeed, monoallelically expressed genes have a higher number of exonic SNPs per transcript than all assessable genes (figure 4.17). Whether this is a result of biases introduced in the allele-specific RNA-sequencing screen, or if it is a consequence of a potential decrease in evolutionary constraint of monoallelically expressed genes in general, cannot be resolved at this stage and requires additional analysis.

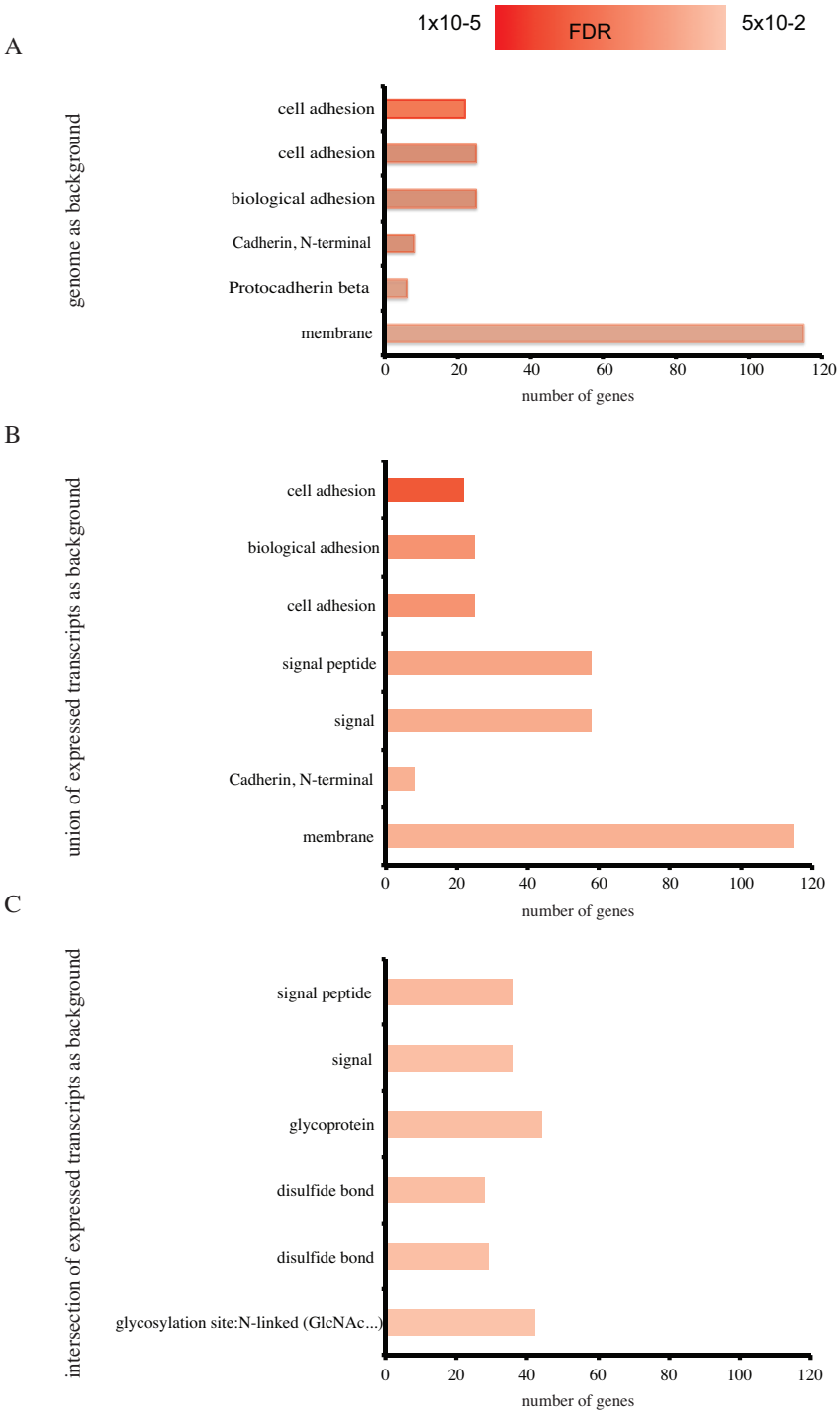


Figure 4.16: Gene ontology analysis of NPC monoallelically expressed genes.
Gene ontology enrichment of monoallelically expressed genes in NPCs compared to (A) genome, (B) union of expressed genes or (C) intersection of expressed genes, as background. Y-axis lists significant categories, X-axis lists the number of genes in each category. Shade of red represents false discovery rate (FDR) with darker colours having higher significance.

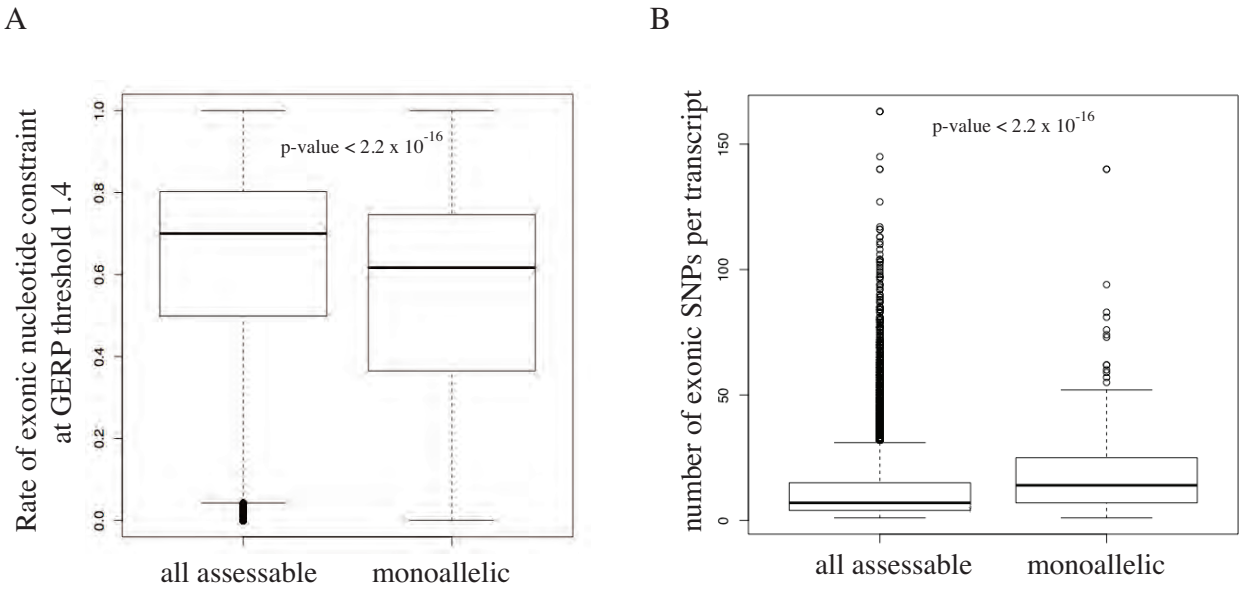


Figure 4.17: Genomic evolutionary rate profiling of monoallelically expressed genes. (A) rate of nucleotide constraint at GERP threshold 1.4 of all assessable and monoallelically expressed genes. (B) number of exonic SNPs per transcript for all assessable and monoallelically expressed genes.

4.5 Dynamics of monoallelic expression during differentiation

Next, the dynamics of monoallelic expression during cell differentiation was investigated. As the number of monoallelically expressed genes increases 5.6 fold between ESCs and NPCs, the overlap in monoallelically expressed genes between these two cell types was first compared. Interestingly, very few monoallelically expressed genes were found in common between the two cell types; of the 376 monoallelically expressed genes in NPCs, only 6 genes were monoallelically expressed in undifferentiated ESCs (figure 4.18A on page 111). In fact the majority (71.8%) of NPC monoallelically expressed genes were biallelically expressed in ESCs (figure 4.18B, figure 4.19), with only 8.8% switching from a biallelic state in ESCs to monoallelic state in NPCs. 17.8% of NPC monoallelically expressed genes were not expressed in the pluripotent ESCs. Likewise most (55.2%) ESC monoallelically expressed genes become biallelically expressed in NPCs, and approximately one third are no longer expressed (figure 4.18, 4.19). There was also notable variability in the expression biases across clones for a given transcript, with some clones showing biallelic expression and others monoallelic expression (figure 4.19), indicating that random autosomal monoallelic expression is not required by the cell. Furthermore, during differentiation there is an increase in monoallelic expression, coinciding with the loss of pluripotency and gain of lineage commitment. Therefore, monoallelic expression, while maintained across cell divisions, is not maintained during differentiation.

The expression level changes of the ESC and NPC monoallelically expressed genes was also compared during differentiation (figure 4.20 on page 114). DESeq (Anders and Huber, 2010) was used to perform the differential expression analysis of transcripts between ESCs and NPCs. Transcripts with a fold change $\geq$2 fold and a FDR$<$0.05 were defined as differentially expressed. The majority of ESC monoallelically expressed genes either decreased (50%) or did not change (32.4%) during differentiation, with only a small proportion (17.6%) increasing in expression level upon differentiation to NPCs. This is of interest as 71.8% of ESC monoallelically expressed genes are biallelically expressed in NPCs. Furthermore, of the NPC monoallelically expressed genes, only 13.1% are expressed at lower levels in the NPCs compared to ESCs, despite the fact that 55.2% switch from biallelic to monoallelic expression during differentiation. Expression levels of approximately half of the monoallelically expressed genes remain unchanged during

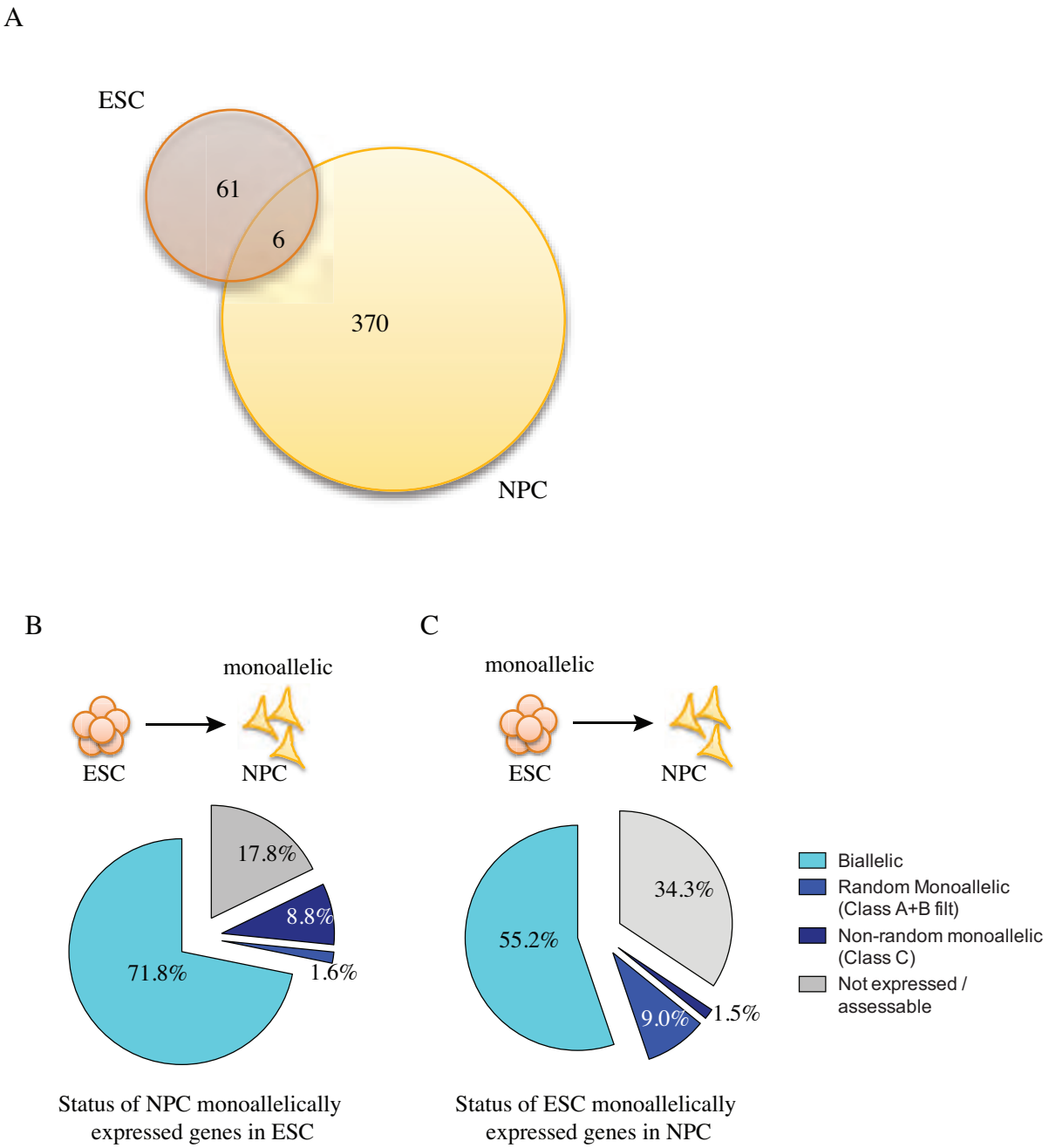


Figure 4.18: Dynamics of monoallelic expression during differentiation.
(A) Overlap of monoallelically expressed genes between ESCs and NPCs. (B) Status of the NPC monoallelically expressed genes in ESCs. (C) Status of the ESC monoallelically expressed genes in NPCs.

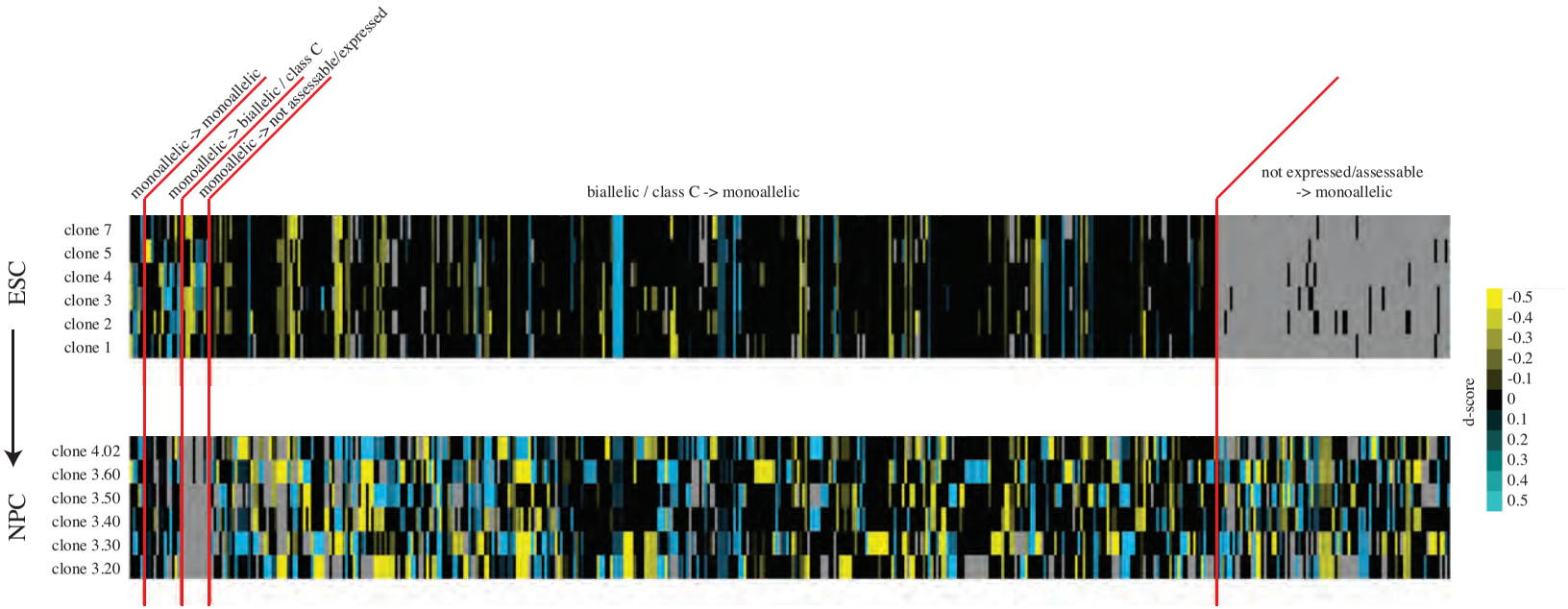


Figure 4.19: Heat map of d-scores for all transcripts showing variation across clones and dynamics during differentiation.
Heat map showing d-scores for each transcript in each clone, with blue representing bias towards C57Bl/6J allele (+0.5), yellow representing bias towards CAST/EiJ allele (-0.5), black denoting biallelic expression, and grey not-assessed or expressed transcripts. Each row represents a single clone in either ESC (top 6 rows) or NPC (bottom 6 rows). Red lines group transcripts into those that (A) remain monoallelic in both ESC and NPC, (B) switch from monoallelic expression in ESC to biallelic or non-random monoallelic expression (class C) in NPC, (C) switch from monoallelic expression in ESC to not-assessable or expressed in NPC, (D) switch from biallelic to monoallelic expression during ESC to NPC differentiation, or (E) switch from non-assessable or not-expressed in ESC to monoallelic expression in NPC. Note the large extent of variation between clones for individual transcripts.

differentiation, and another 40% increase, despite them switching to monoallelic gene expression in NPCs. This suggests that monoallelic expression is not a mechanism for reducing transcript levels in the cell and instead may reflect a stochastic nature of gene expression.

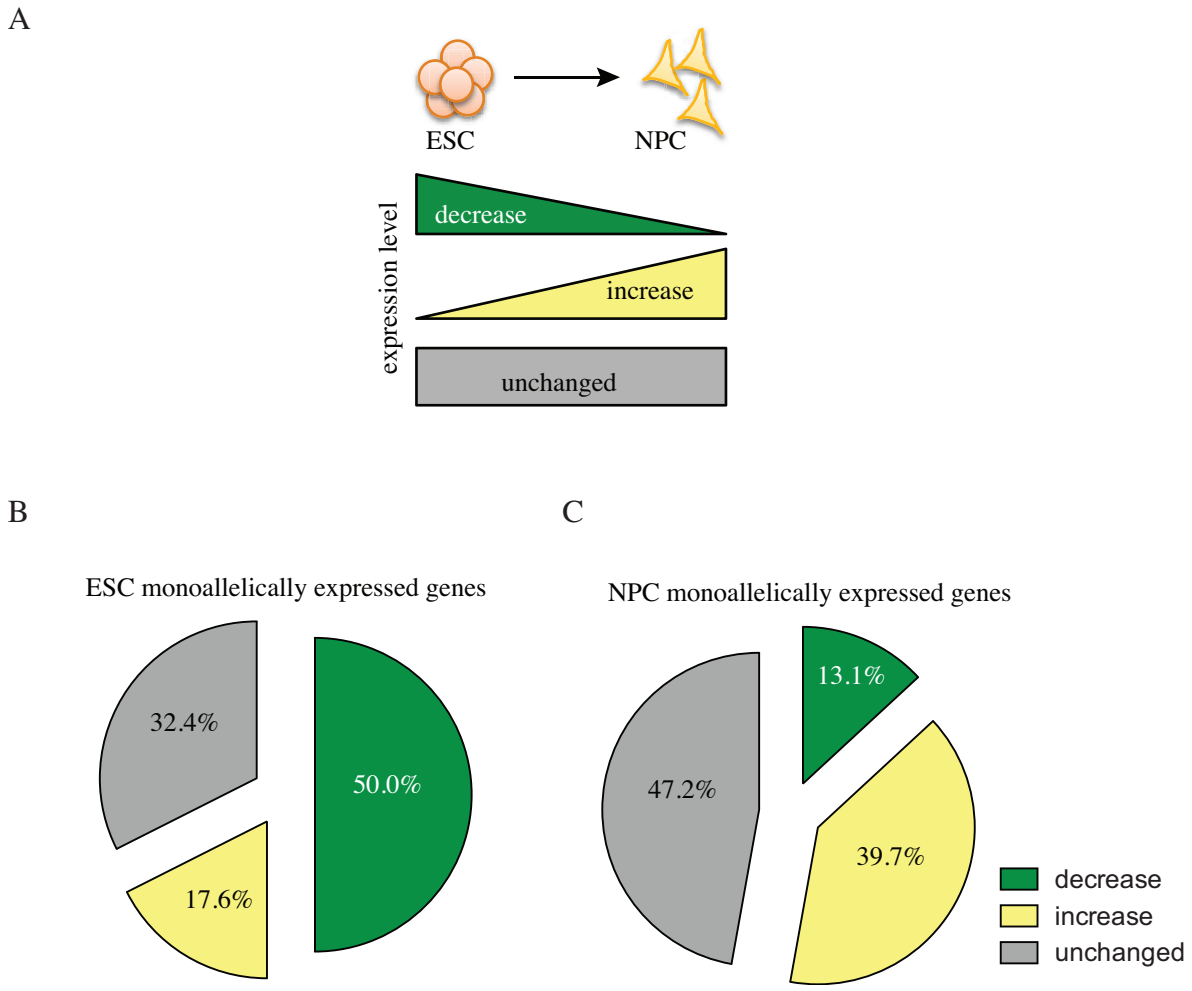


Figure 4.20: Expression level changes of monoallelically expressed genes upon differentiation. (A) Expression levels can either decrease, increase or remain unchanged during ESC to NPC differentiation. (B) Expression level changes for ESC monoallelically expressed genes. (C) Expression level changes for NPC monoallelically expressed genes.

Chapter 5

Inheritance of monoallelic gene expression

One intriguing aspect of monoallelic expression is that the transcriptional imbalance between the active and inactive alleles is maintained across cell generations. The molecular mechanisms describing how the active and inactive alleles are inherited through the cell cycle have not been previously described for random autosomal monoallelically expressed genes. I therefore sought to characterise the random autosomal monoallelically expressed genes identified in this study, based on three potential mechanisms for inheritance: DNA methylation; histone modifications and variants; and nuclear organisation. Additionally an epigenetic drug screen was initiated to identify molecules which may provide insight into how the inheritance is fulfilled.

5.1 DNA methylation

DNA methylation is the most widely accepted mechanism through which the transcriptional state of a gene can be inherited and maintained in daughter cells (reviewed in (Smith and Meissner, 2013)). Furthermore, differences in DNA methylation at imprinting control regions are responsible for distinguishing maternal and paternal alleles for imprinted gene clusters (reviewed in (Kelsey and Feil, 2013)), and DNA methylation differentially marks regions of the active and inactive X-chromosomes in female cells (reviewed in (Schulz and Heard, 2013)).

Initially, a potential role for DNA methylation in regulating random autosomal monoallel-

ically expressed genes was assessed bioinformatically. DNA methylation in mammals typically occurs at symmetrical CpG dinucleotides (Gruenbaum et al., 1981; Sinsheimer, 1955). The CpG density at the promoters of the monoallelically expressed genes was computed by calculating the number of CpGs in a 1kb region surrounding the transcription start site. Interestingly, there was a significant reduction in CpG density at the promoters of monoallelically expressed genes compared to all assessable genes (figure 5.1 on the following page). This is of particular interest as the overall CG content is similar between monoallelically expressed genes and all assessable genes (figure 4.15 on page 106). However, as mentioned previously in the case of evolutionary conservation of these genes 4.17 on page 109, it is unclear whether the decreased CpG density is potentially an consequence of the selection bias of transcripts containing a higher SNP frequency in the allele-specific screen. Further analysis is needed to determine if this is a true feature of monoallelically expressed genes, as has been previously suggested (Jeffries et al., 2012).

Next, bisulfite analysis of the promoters of monoallelically expressed genes was performed. Bisulfite treatment of DNA converts unmodified cytosines to uracil, which are then read out as thymine following PCR amplification and Sanger sequencing. Methylated cytosines, as well as hydroxymethylated cytosines, remain unaffected in the bisulfite conversion reaction, providing a digital readout as to the methylation status of single molecules at single base resolution. If DNA methylation was differentially marking the active and inactive alleles, one would expect monoallelic clones to specifically contain a mixture of methylated and unmethylated CpG sites, as is observed for imprinted genes. Biallelic clones, however, would not show this mix of two states. The levels of DNA methylation were first compared at the promoters of 5 monoallelically expressed genes containing CpG islands at their transcription start sites. Samples included NPC monoallelic, allele-biased and biallelic clones for the respective gene, as well as for ESCs (figure 5.2 on page 119). The promoters in ESCs were unmethylated for 4 out of 5 genes tested, with *Cpped1* showing hypermethylation, despite being expressed in ESCs. Upon differentiation, there was in general a gain in methylation levels at the promoters, although the extent varied between genes and clones examined. *Gas6* switched from an unmethylated state in ESCs to methylated in NPCs regardless of the transcriptional status of the allele. Likewise *Anxa6* was similarly methylated across the different NPC clones, although the extent of methylation correlated with the overall transcriptional output of the clone. *Cpped1*, which was methylated

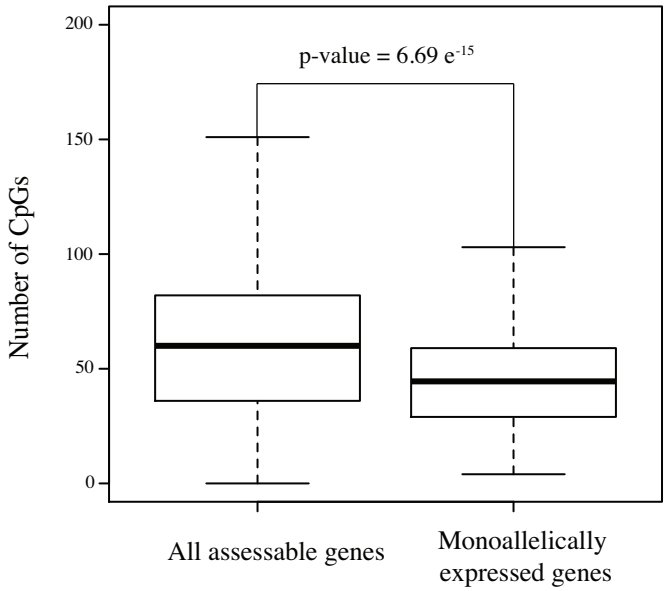


Figure 5.1: Decreased CpG density at monoallelically expressed gene promoters. Number of CpGs in 1kb sequence surrounding the transcription start site of all assessable genes compared to monoallelically expressed genes.

in the ESCs, remained mostly methylated in the NPC clones, with the exception of a biallelic clone which showed a mixture of methylated and unmethylated reads, despite both alleles being transcribed. Interestingly, methylation patterns for both *Npl* and *Cbr3* in a monoallelic clone showed a clear segregation between methylated and unmethylated reads. In the case of *Cbr3*, the reads were able to be assigned to either the active C57Bl/6J allele or inactive CAST/EiJ allele due to the presence of a A-T SNP within the amplicon analysed. This confirmed that the unmethylated reads indeed corresponded to the active allele, where as the methylated reads corresponded to the inactive allele. Both the biallelic and allele-biased clones for *Cbr3* contained unmethylated reads for both alleles, consistent with them being transcribed.

As the monoallelically expressed genes were observed to have reduced CpG density at their promoters when compared to all assessable genes (figure 5.1), bisulfite analysis was performed for an additional 5 genes which did not have CpG islands at their promotors (figure 5.3 on page 120). Again, both ESCs as well as NPC clones which were biallelic, allele-biased or monoallelic for the respective gene were analysed. Similarly to the CpG high promotors, ESCs were largely unmethylated (figure 5.3, top row). There was an increase in DNA methylation upon differentiation to NPCs at the promotor of *Cacng5*, such that the majority of CpGs became methylated, regardless of the transcriptional status of the two alleles. Another gene, *Rhoj*, remained mostly unmethylated in all NPC clones assessed. However, for two other genes, *Serpinh1* and *Fkbp7*, the monoallelic clones showed approximately 50% methylated reads and 50% unmethylated reads. Furthermore, in the case of *Fkbp7*, this was only observed for the monoallelic clone, and not for the allele-biased or biallelic clone. Therefore, of the 10 genes assayed overall, 4 showed evidence supporting allele-specific DNA methylation correlating with transcriptional status in the monoallelic clones.

As there were several examples supporting allele-specific DNA methylation correlating with transcriptional output, I next tested to see whether the DNA methylation was necessary to maintain the monoallelic state of the cells. 5-azacytidine is a small chemical analogue of the nucleic acid cytidine, and was first synthesised as a potential chemotherapeutic agent for cancer (Cihák, 1974). At low doses, it inhibits the activity of the DNA methyltransferases, thus treatment over several cell cycles leads to passive genome wide DNA demethylation. At higher doses, 5-azacytidine incorporates into both DNA and RNA and ultimately causes cell death. First, the

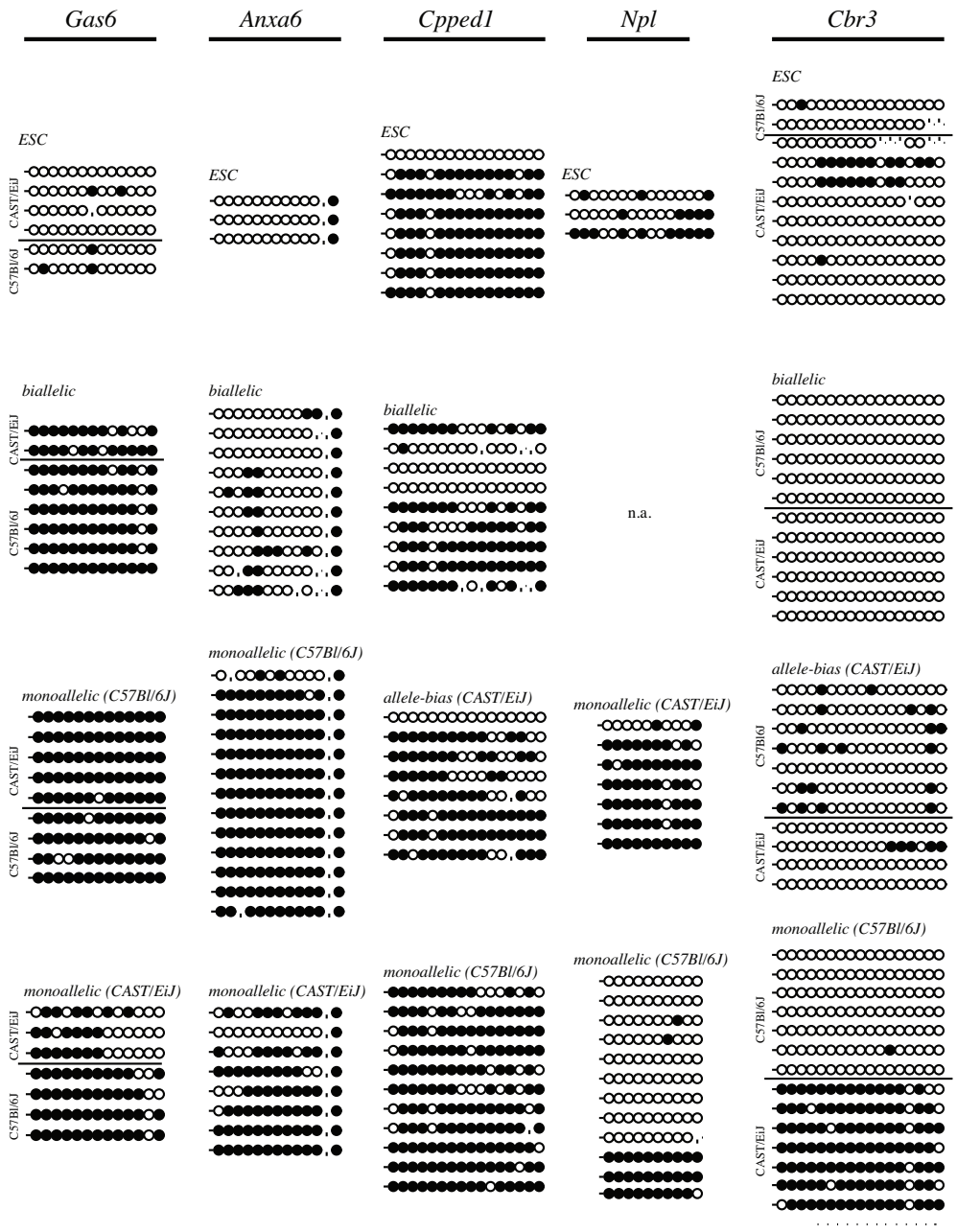


Figure 5.2: Bisulfite analysis of CpG high promotors.
Reads from bisulfite analysis performed at the promoters of CpG high monoallelically expressed genes in ESCs (top row) and NPC clones classified as biallelic (second row), allele-biased or monoallelic (third and fourth rows) for the respective gene. 5 separate genes were analysed. Open circles depict unmethylated CpGs, filled circles methylated CpGs. Allele information is noted when available. n.a. not assessable.

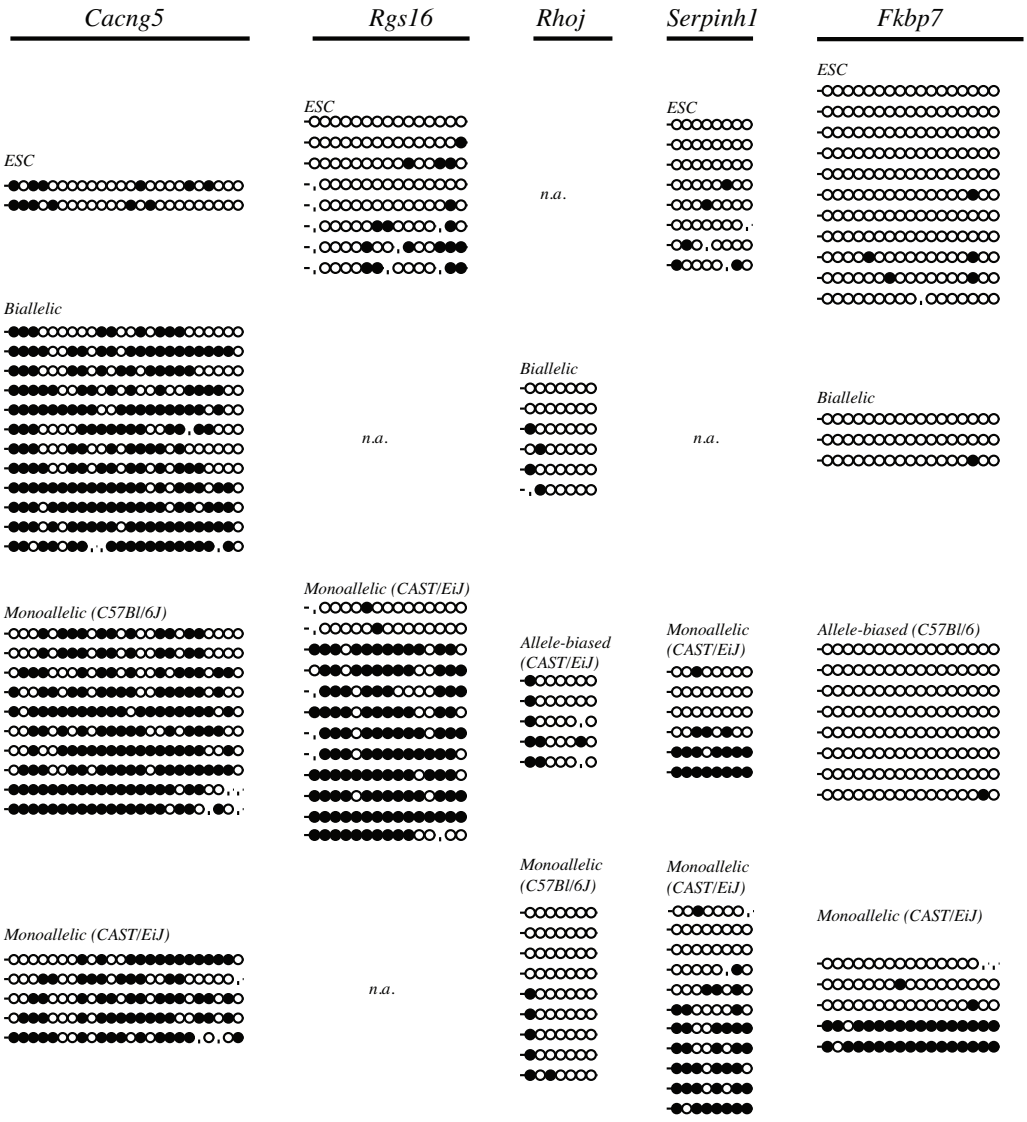


Figure 5.3: Bisulfite analysis of CpG low promotors.
Reads from bisulfite analysis performed at the promoters of CpG low monoallelically expressed genes in ESCs (top row) and NPC clones classified as biallelic (second row), allele-biased or monoallelic (third and fourth rows) for the respective gene. 5 separate genes were analysed. Open circles depict unmethylated CpGs, filled circles methylated CpGs. n.a. not assessable.

dosage and duration of 5-azacytidine was determined for NPCs. Cells were treated with 10nM, 200nM, 500nM, 1μM or 5μM 5-azacytidine for 48 hours and genomic DNA collected and level of DNA methylation assayed by bisulfite-PCR. However, following bisulfite conversion, there was still significant levels of DNA methylation in these cells, possibly due to the short time period of treatment. Cells were treated with either 1μM or 10μM for 5 days, with fresh medium containing 5-azacytidine changed every 2 days. During this time cells underwent several cell divisions and were passaged at least once. Genomic DNA was collected and bisulfite converted. Analysis of the *Cbr3* promotor revealed that while 1μM was not sufficient, 10μM 5-azacytidine led to demethylation of the inactive allele (figure 5.4 on the next page). Next, the transcriptional output of the two alleles was assessed by SNP-PCR. However, demethylation of the inactive allele did not lead to biallelic expression of *Cbr3* (figure 5.5 on page 123). Nor did 5-azacytidine lead to activation of the silent allele for 9 additional genes tested, including the differentially methylated *Serpinh1*, *Fkbp7* and *Npl* (figure 5.5). Based on the SNP-PCR results, there is no evidence that DNA methylation, although in some instances correlative, directly regulates monoallelically expressed genes. Genome-wide RNA-sequencing in several 5-azacytidine treated clones is currently underway to determine whether this is a general feature for all monoallelically expressed genes.

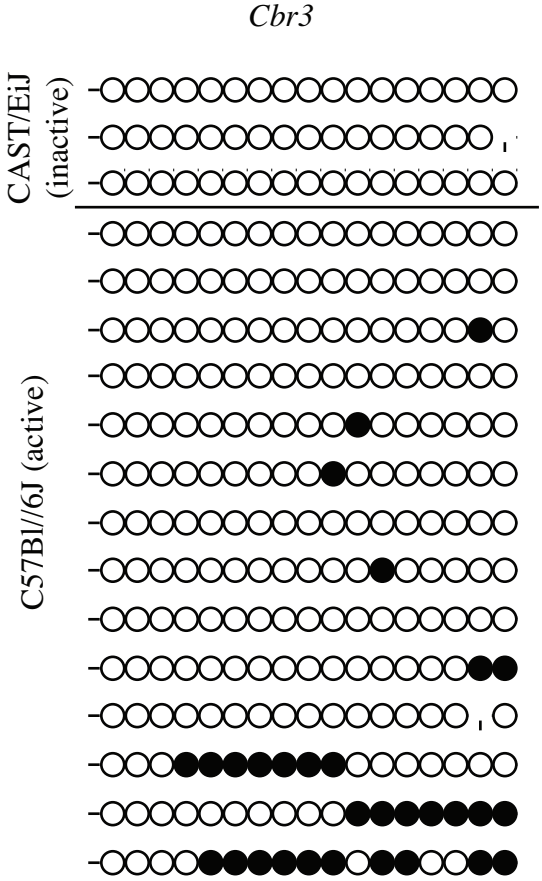


Figure 5.4: 5-azacytidine treatment leads to demethylation of the inactive *Cbr3* allele. NPCs were treated with 10μM 5-azacytidine for 5 days, genomic DNA collected and bisulfite analysis of the *Cbr3* promotor performed. Open circles represent unmethylated CpGs, closed circles methylated CpGs. 5-azacytidine treatment successfully led to the demethylation of the previously methylated CAST/EiJ allele.

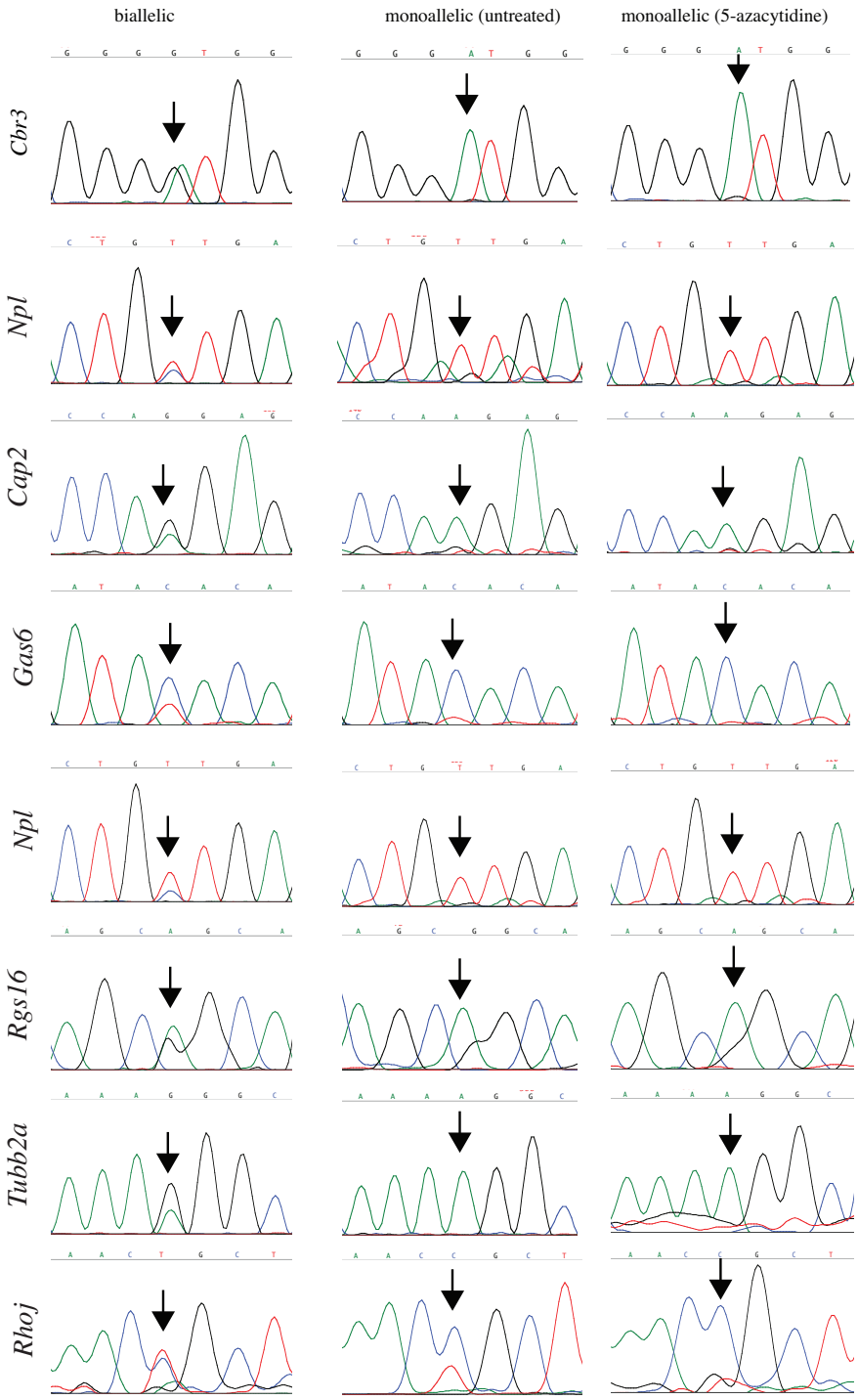


Figure 5.5: Effect of 5-azacytidine treatment on monoallelic expression.
Sanger sequencing traces of 7 monoallelically expressed genes for a biallelic clone (left column), monoallelic or allele-biased clone (middle column) and the same monoallelic clone treated for 5 days with 5-azacytidine (right column). Arrows denote SNP position. blue = C, black = G, green = A, red = T.

5.2 Histone modifications

While the mechanisms by which histone modifications may transmit transcriptional states through mitosis remain unclear, the preferential association of certain marks with different genomic features has been extensively documented (reviewed in (Black et al., 2012)). Therefore, the histone modifications at the promotors of these genes was investigated by chromatin immunoprecipitation (ChIP) to determine whether there were modifications specific to the active and/or inactive alleles. First, the chromatin preparation conditions for NPCs were optimised. Cells were fixed in 1% formaldehyde for 5 minutes at room temperature and quenched with glycine. Nuclear lysis was verified by staining with methyl-green in which complete removal of the cytoplasm could be observed under the light microscope. Nuclei were sonicated using the bioruptor for 8, 12, 16 and 20 cycles of 30 seconds on the highest setting followed by 30 seconds off. The chromatin smear was checked by loading sonicated chromatin onto a gel (figure 5.6 on the next page). Chromatin preparations should ideally contain the bulk of material at nucleosomal resolution, at approximately 200bp. From the sonication time course, 20 cycles was selected for all further histone ChIP experiments.

Allele-specific ChIP was performed by including an informative SNP in the region assayed by ChIP-PCR. In this way, following quantitative PCR analysis of the levels of pulldown, PCR products were subjected to Sanger sequencing and the associated alleles with the respective histone modification determined. A total of 9 separate histone modifications were assayed across 4 independent NPC clones, with at least 3 biological replicates performed on separate days each. For each pull-down, a total of 4 monoallelic gene promotors, 4 monoallelic gene bodies were assayed by Q-PCR, as well as control regions including one biallelic gene (*Gapdh1*), one inactive gene (*Nanog*), or previously characterised active and inactive genomic regions (Mikkelsen et al., 2007). First, the association of histone H3 lysine 4 di- and tri-methylation (H3K4me2, H3K4me3) were investigated at the promotors of monoallelically expressed genes. These modifications are associated with active gene promotors. As a control, H3K4me2 and H3K4me3 were both present at the promotor of *Gapdh1*, but not at the promotor of the transcriptionally inactive *Nanog* gene (figure 5.7 on page 127). Both H3K4me2 and H3K4me3 were present at the promotors of the 4 monoallelically expressed genes tested. In general, there was a tendency for

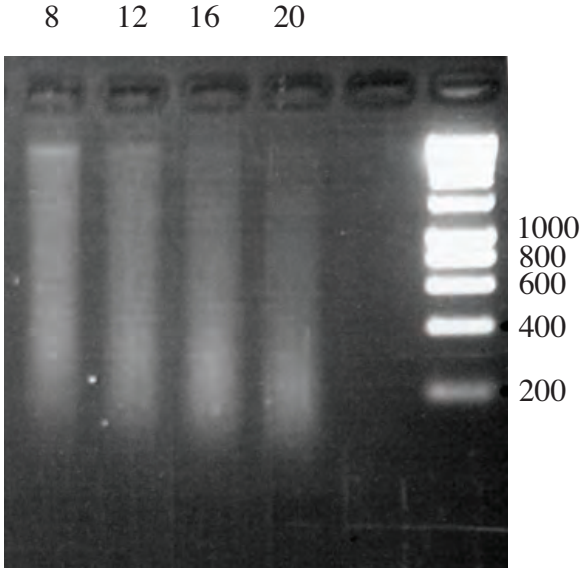


Figure 5.6: Optimisation of sonication conditions for chromatin immunoprecipitation. Nuclei were sonicated for 8, 12, 16 and 20 cycles of 30 seconds on maximum setting. Chromatin was loaded onto an agarose gel to check size distribution of DNA. Last lane contains DNA ladder, sizes are as marked in base pairs.

the biallelically expressing clones to have a greater level of association with H3K4me2/3 than the monoallelically expressing clones. Importantly, Sanger sequencing of the PCR products following ChIP-PCR revealed that while both alleles were associated with H3K4me2/3 in the biallelic clones, only the actively transcribing allele was marked by H3K4me2/3 in the monoallelically expressed clones (figure 5.8 on page 128). Thus the actively transcribing allele is specifically marked by both H3K4me2 and H3K4me3.

Next, allele-specific ChIP was performed against a panel of histone modifications generally associated with silent chromatin. H3K9me3 and H4K20me3 are typically associated with constitutive heterochromatin, where as the Polycomb mediated H3K27me3 with facultative heterochromatin and bivalent domains. Also tested was another mark of constitutive heterochromatin H4K20me1 which is implicated in dosage compensation in mammals (Kohlmaier et al., 2004), *C. Elegans* (Vielle et al., 2012) and *Drosophila* (Conrad and Akhtar, 2011). Association of H3K9me3 at the promotors of inactive genes was greater in monoallelically expressed clones compared to biallelic clones where levels were comparable to an actively transcribing control region (figure (5.9)A on page 129). In contrast, H3K27me3 was not associated with any of the genes tested, with the exception of *Cbr3* which showed a small pulldown (figure 5.9B). Furthermore, analysis of genome-wide published ChIP-chip data of H3K4me3 and H3K27me3 at promotor regions in NPCs (Mikkelsen et al., 2007), revealed that only 1.3% of monoallelically expressed promotors assessed were associated with H3K27me3, and another 1.3% with the bivalent H3K4me3/H3K27me3 signature (figure 5.9C). Instead the vast majority are marked by H3K4me3, consistent with the findings above. H4K20me1 was expressed at similar levels across all samples (figure 5.10A on page 130), where as H4K20me3 was not present at any of the regions examined (figure 5.10B). Importantly, Sanger sequencing of the PCR products revealed that H3K9me3 was specifically associated with only the inactive allele in monoallelically expressing clones (figure 5.8). H4K20me1, however, was present on both alleles at all genes tested indicating that it was not specifically associated with the inactive allele. Therefore H3K9me3 marks the promoter of the inactive allele of monoallelically expressed genes.

Histone acetylation is generally associated with open active chromatin. Therefore, ChIP analysis was performed using a pan-H4 acetylation antibody, as well as antibodies against H4K16ac and H3K14ac. Levels of H4 acetylation at the promotors of the genes were simi-

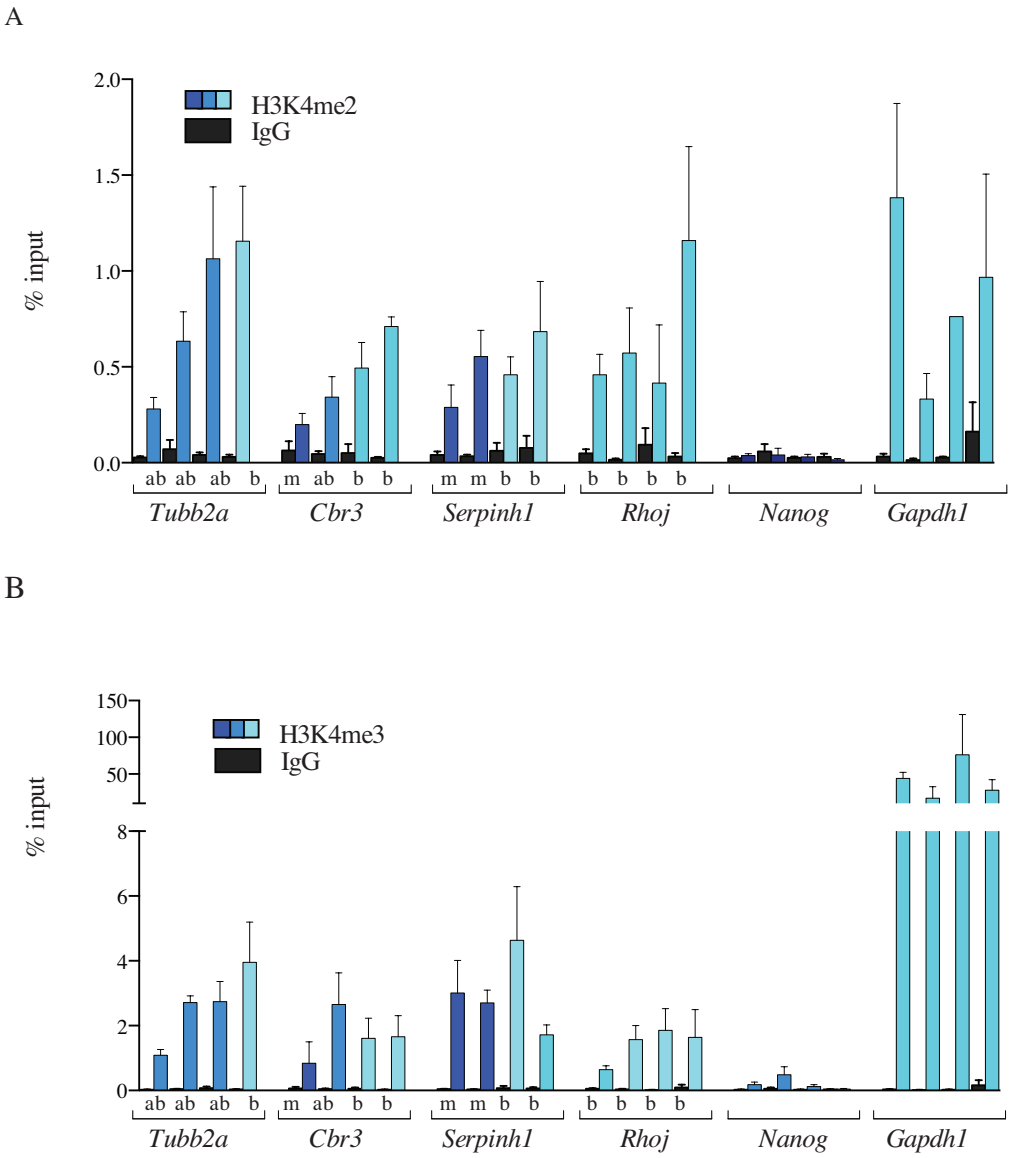


Figure 5.7: Chromatin Immunoprecipitation of H3K4me2 and H3K4me3.
Chromatin Immunoprecipitation analysis for H3K4me2 (A) and H3K4me3 (B). Analysis of regions within 200bp of the transcription start site for four separate monoallelically expressed genes are shown, in addition to transcriptionally silent *Nanog* and active *Gapdh1* (grey). Quantification of pulldown as percentage of input for individual clones which are either monoallelic (m, dark blue), allele-bias (ab, medium blue) or biallelic (b, light blue) for the respective clone are shown. IgG (black) shows non-specific pulldown. Error bars represent SEM of 3-4 biological replicates.

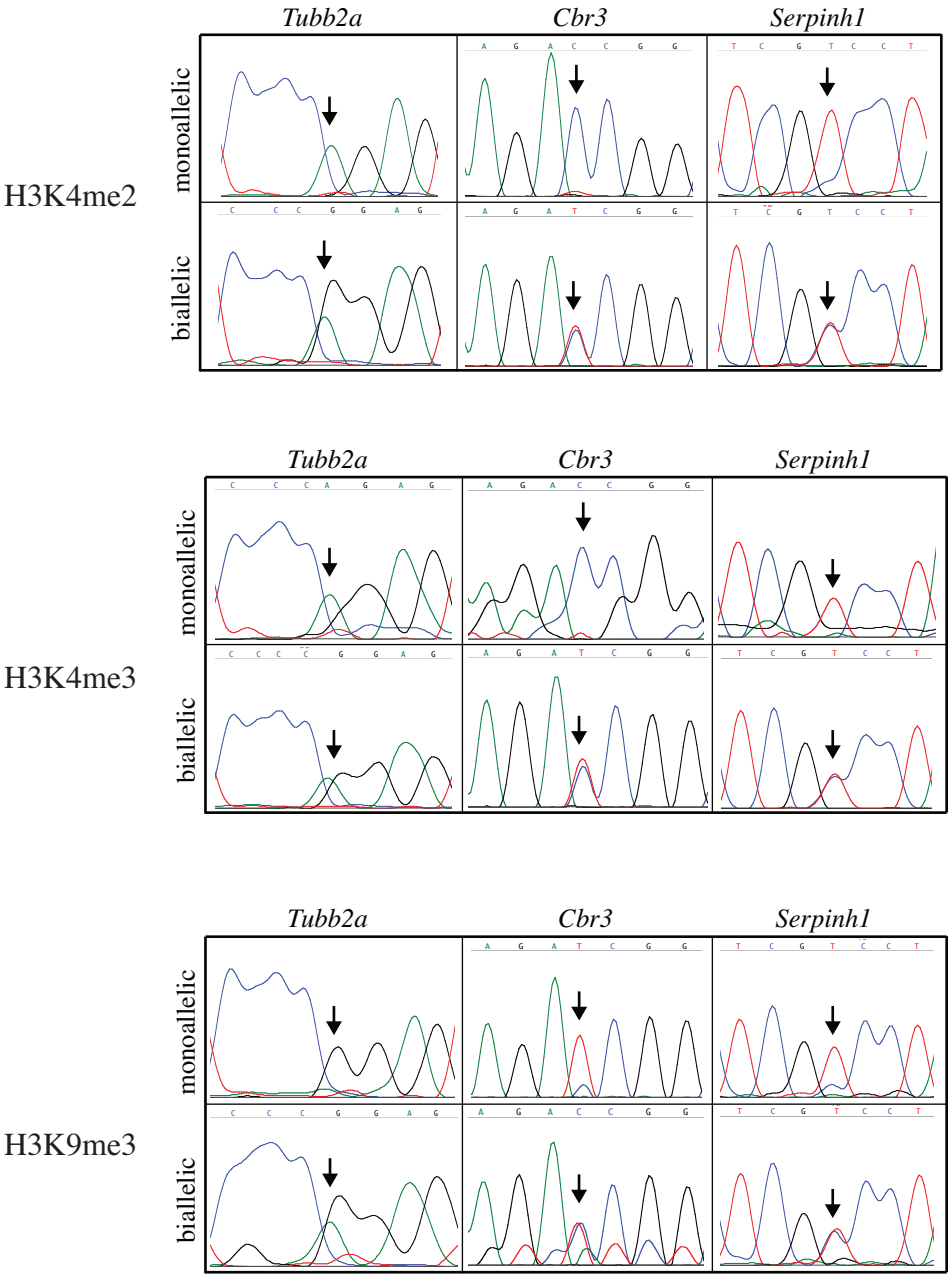


Figure 5.8: Sanger sequencing of ChIP-PCR products.
Representative Sanger sequencing traces of ChIP-PCR products from H3K4me2 (top), H3K4me3 (middle) and H3K9me (bottom) ChIP experiments in either a monoallelic clone or a biallelic clone for 3 separate gene promotors. Arrows denote SNP position. blue = C, black = G, green = A, red = T.

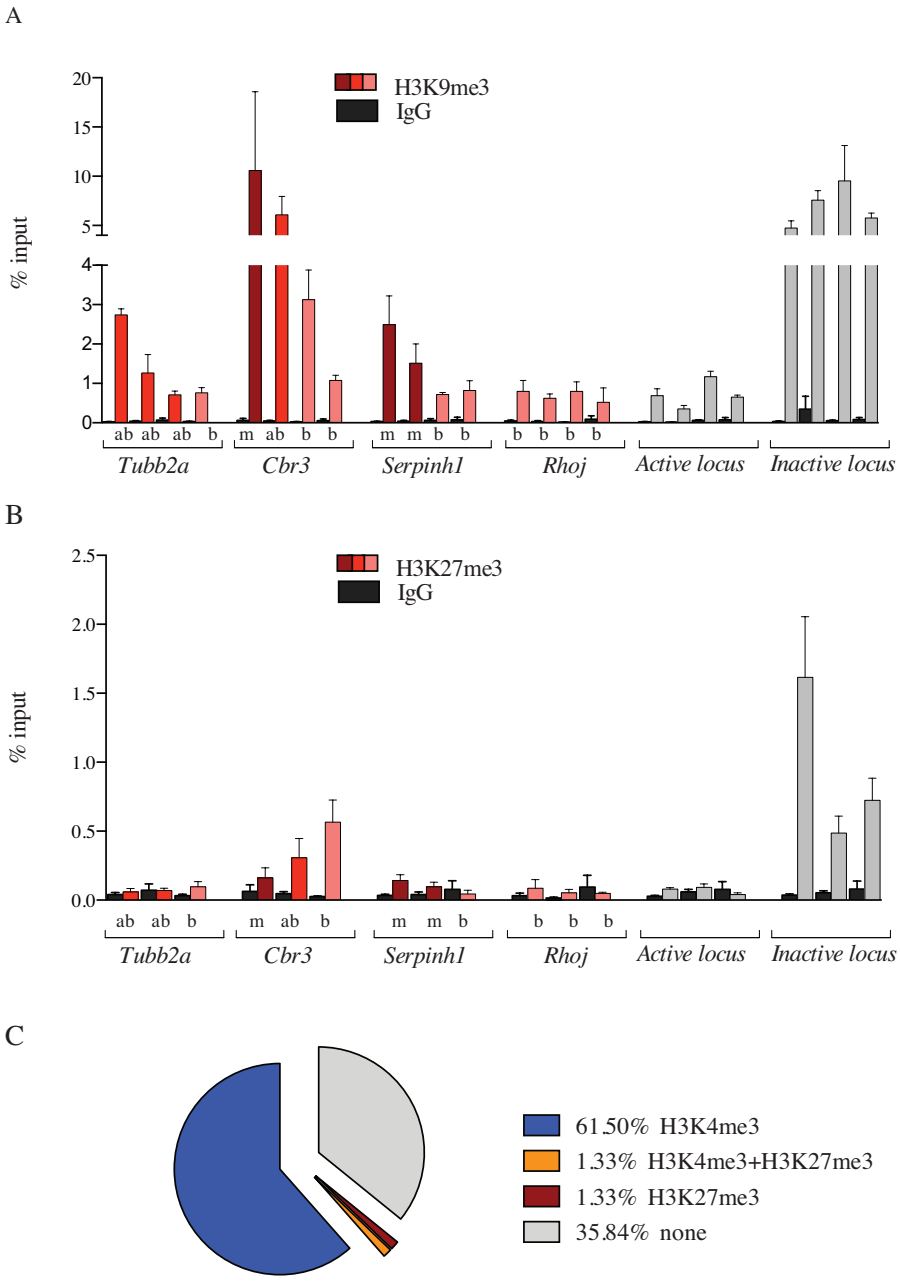


Figure 5.9: Chromatin Immunoprecipitation of H3K9me3 and H3K27me3.
Chromatin Immunoprecipitation analysis for H3K9me3 (A) and H3K27me3 (B). Analysis of regions within 200bp of the transcription start site for three-four separate monoallelically expressed genes are shown, in addition to transcriptionally active and silent regions (grey). Quantification of pulldown as percentage of input for individual clones which are either monoallelic (m, dark red), allele-bias (ab, medium red) or biallelic (b, light red) for the respective clone are shown. IgG (black) shows non-specific pulldown. Error bars represent SEM of 3-4 biological replicates. (C) Genome wide analysis of H3K4me3 and H3K27me3 distribution at the promotors of monoallelically expressed genes in NPCs.

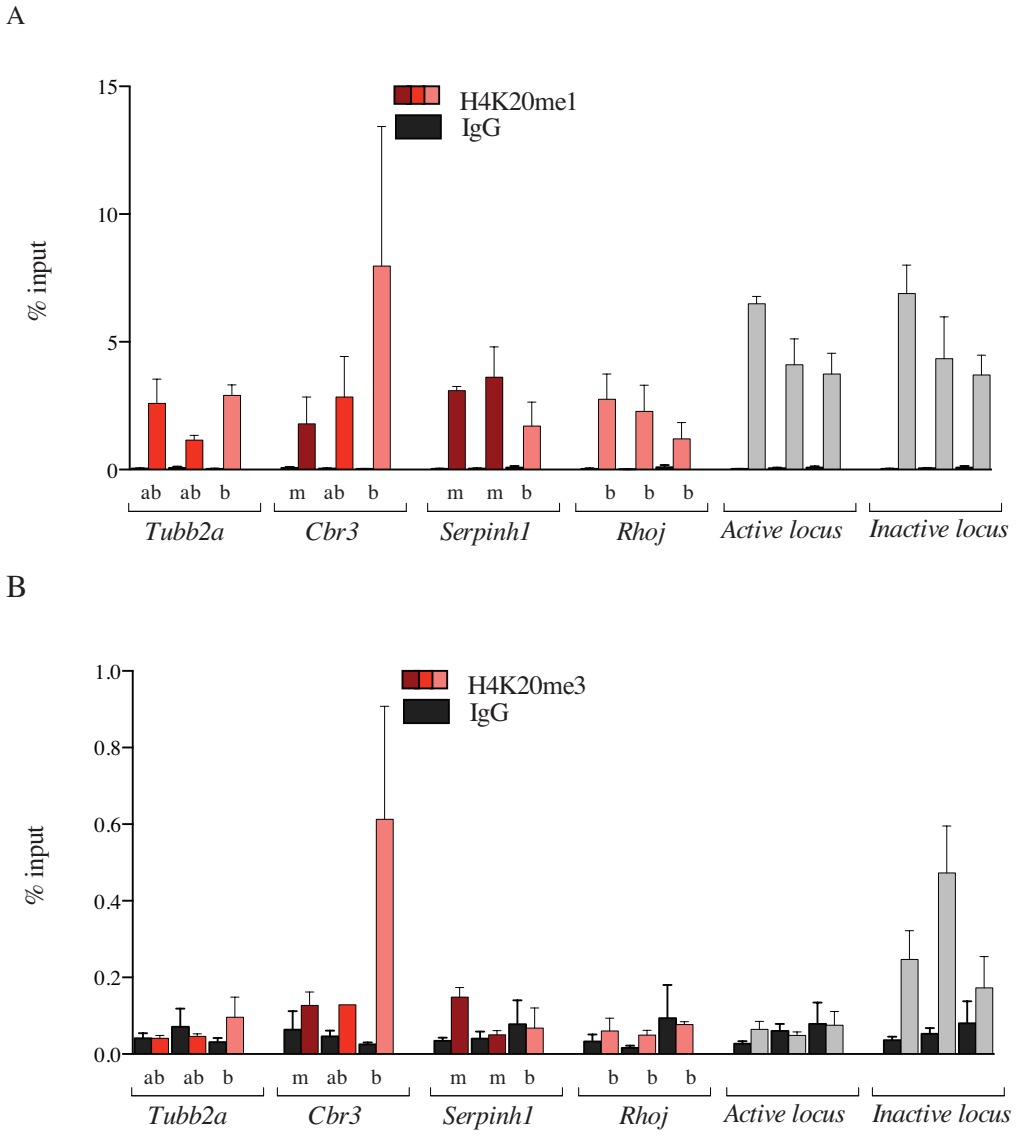


Figure 5.10: Chromatin Immunoprecipitation of H4K20me1 and H4K20me3.
Chromatin Immunoprecipitation analysis for H4K20me1 (A) and H4K20me3 (B). Analysis of regions within 200bp of the transcription start site for three separate monoallelically expressed genes are shown, in addition to transcriptionally active and silent regions (grey). Quantification of pulldown as percentage of input for individual clones which are either monoallelic (m, dark red), allele-bias (ab, medium red) or biallelic (b, light red) for the respective clone are shown. IgG (black) shows non-specific pulldown. Error bars represent SEM of 3-4 biological replicates.

lar between monoallelically and biallelically expressing clones for the four genes tested (figure 5.11A). Furthermore when Sanger sequencing was performed on the PCR products, both alleles were found associated with this mark in the monoallelically expressed clones. Similar results were observed for both H4K16ac (figure 5.11B) and H3K14ac (figure 5.12A), which showed similar levels of association between different clones, and no specificity between active and inactive alleles by Sanger sequencing analysis. I also investigated H3K36me3 at the promotors of these genes (figure 5.12 on page 133B), which also did not show any allele-specificity. However H3K36me3 is enriched over the gene body. Analysis of H3K36me3 association in exons of these genes is currently in progress.

In summary, of the 10 histone modifications examined, three showed allele-specific association (table 5.1 on page 134). Specifically H3K4me2 and H3K4me3 were present on the active allele, and H3K9me3 on the inactive allele. Genome wide ChIP-seq analysis is currently underway to determine whether this is a general rule for monoallelically expressed genes. If so, it may be possible to predict monoallelic expression based solely on chromatin marks, although dual association of both H3K4me2/3 and H3K9me3 observed by ChIP could either be a consequence of monoallelic expression, or result from a mixture of active and inactive cells within a population. Whether the marks identified are actively involved in maintaining the transcriptional bias between the alleles across cell divisions, or whether they are a secondary effect of gene transcription, remains to be determined and will be the focus of future studies.

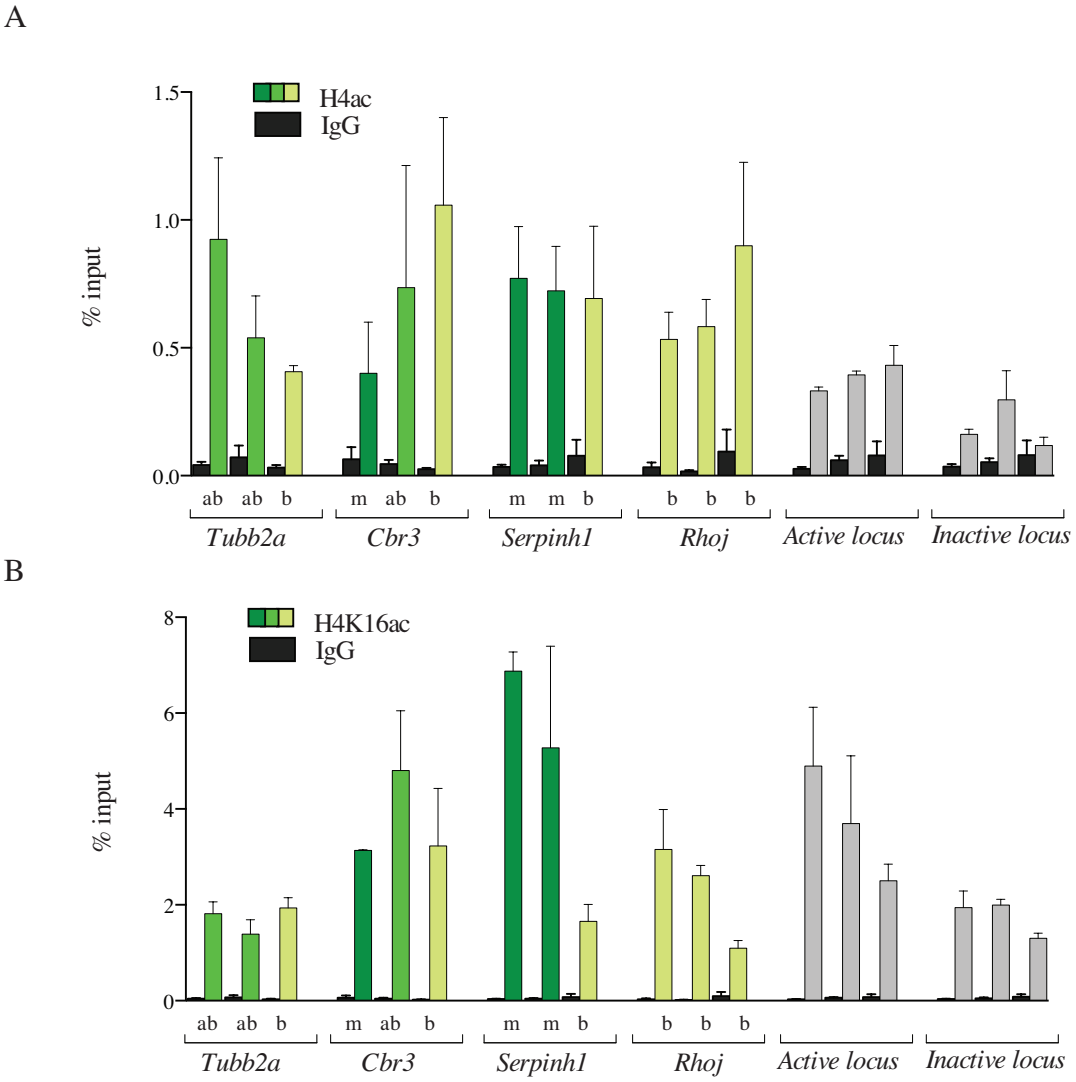


Figure 5.11: Chromatin Immunoprecipitation of histone acetylation modifications. Chromatin Immunoprecipitation analysis for panH4 acetylation (A) and H4K16ac (B). Analysis of regions within 200bp of the transcription start site for three separate monoallelically expressed genes are shown, in addition to transcriptionally active and silent regions (grey). Quantification of pulldown as percentage of input for individual clones which are either monoallelic (m, dark green), allele-bias (ab, light green) or biallelic (b, yellow) for the respective clone are shown. IgG (black) shows non-specific pulldown. Error bars represent SEM of 1-4 biological replicates.

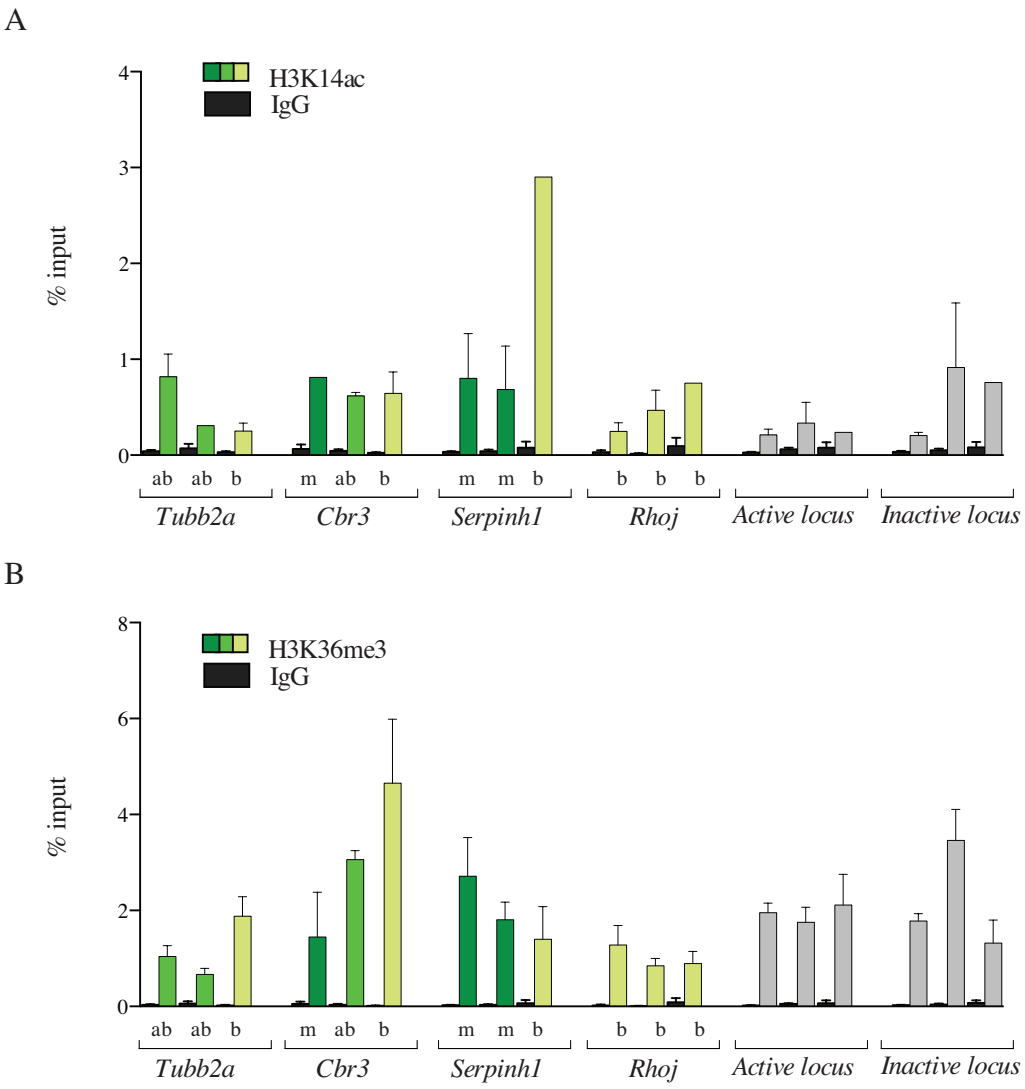


Figure 5.12: Chromatin Immunoprecipitation of H3K36me3.
Chromatin Immunoprecipitation analysis for H3K14ac (A) and H3K36me3 (B). Analysis of regions within 200bp of the transcription start site for three separate monoallelically expressed genes are shown, in addition to transcriptionally active and silent regions (grey). Quantification of pulldown as percentage of input for individual clones which are either monoallelic (m, dark green), allele-bias (ab, light green) or biallelic (b, yellow) for the respective clone are shown. IgG (black) shows non-specific pulldown. Error bars represent SEM of 1-4 biological replicates.

Table 5.1: Summary of ChIP analysis.

Modification	Active Allele	Inactive Allele	Comments
H3K4me2	✓	✗	Specific for active allele
H3K4me3	✓	✗	Specific for active allele
H3K9me3	✗	✓	Specific for inactive allele
H3K27me3	-	-	Not present
panH4ac	✓	✓	No change and present on both alleles
H4K16ac	✓	✓	No change and present on both alleles
H4K20me1	✓	✓	No change and present on both alleles
H4K20me3	-	-	Not present
H3K14ac	✓	✓	No change and present on both alleles
H3K36me3	✓	✓	No change and present on both alleles

5.3 Nuclear organisation

The nucleus is a highly structured and dynamic organelle containing many distinct domains and bodies, (reviewed in(Hübner et al., 2013; Spector, 2006; Zhao et al., 2009)), and the position of genes in relation to these structures has been correlated with their transcriptional activity (reviewed in (Hübner et al., 2013)). Monoallelically expressed genes provide a unique system for investigating the role of nuclear organisation on gene expression, as the active and inactive transcriptional status of an endogenous locus can be analysed simultaneously within the same nucleus. The position of the active and inactive alleles was analysed in relation to two nuclear structures associated with silent chromatin: the nuclear periphery and heterochromatic foci (figure 5.13A on the following page). Combined RNA-DNA FISH was performed for a panel of 6 monoallelically expressed genes in NPCs and three-dimensional image stacks through the entire nucleus collected. Association with the nuclear periphery and heterochromatic foci was determined when an allele which was touching or overlapping the structure identified by DAPI staining. At least 30 active-inactive allele pairs were analysed for each monoallelically expressed gene. Approximately 30% of alleles were associating with heterochromatic foci (figure 5.13B). If the heterochromatic foci were actively involved in repressing the silent allele, a higher frequency of association of the inactive allele with this domain would be predicted. However, there was no clear difference in associations between the active and inactive alleles for the genes analysed, with the exception of *Acyp2*, in which the active allele had reduced frequency of association, while the inactive allele showed association frequencies similar to active alleles of other genes. Similar results were observed for the association with the nuclear periphery (figure 5.13C). Levels of association were similar for the active and inactive alleles of each gene, with the exception of *Ror2* in which the inactive allele showed a higher frequency of association compared to the active allele. However this frequency was similar to that of the active allele of *Acyp2* and *Pdzrn4*. Biological replicates need to be performed to determine whether the small differences observed for *Acyp2* and *Ror2* are real. However, in general, there was no striking evidence for preferential positioning of the inactive allele towards heterochromatic foci nor the nuclear periphery. Furthermore there was no evidence of for allelic pairing of the active and inactive alleles.

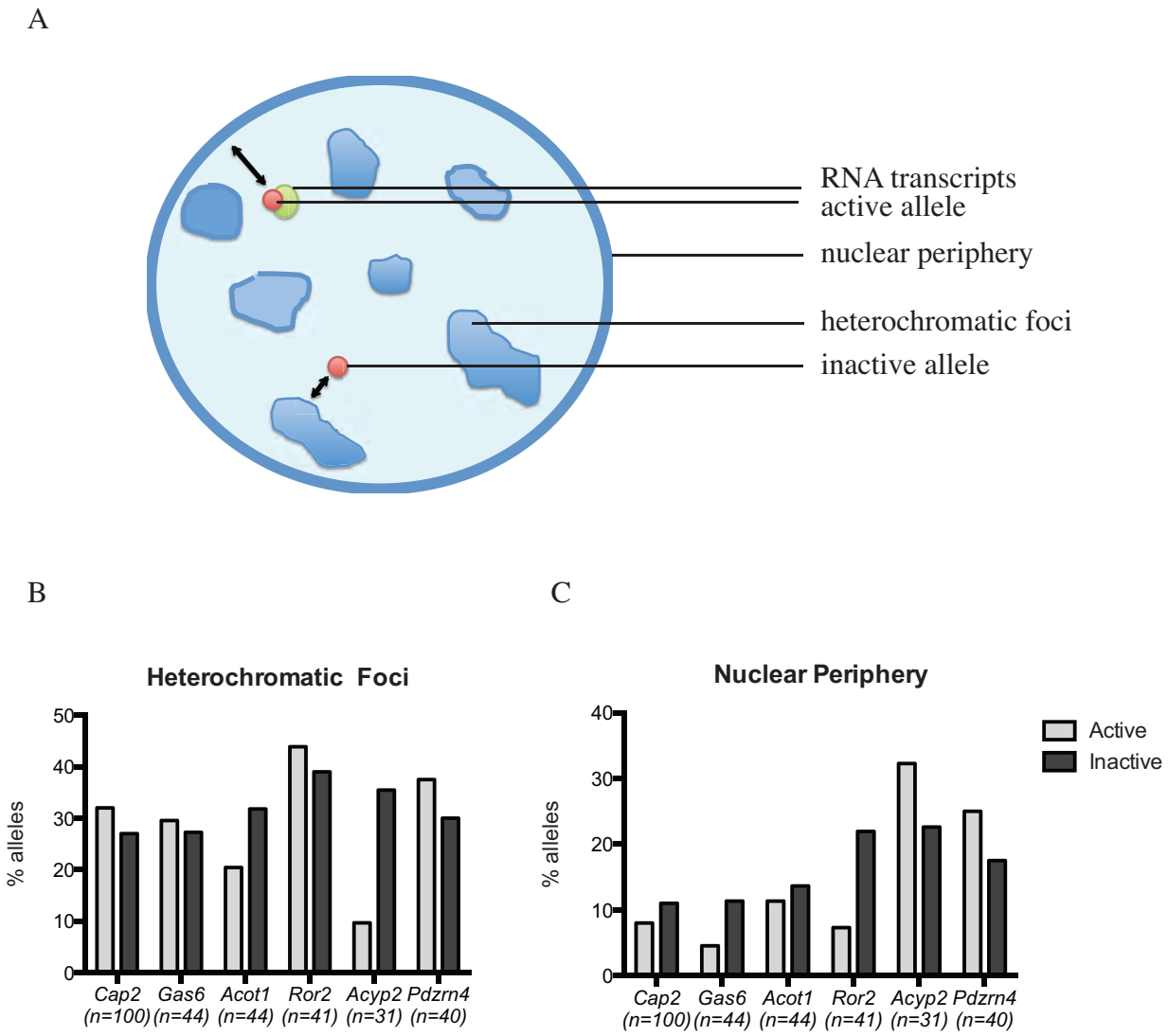


Figure 5.13: 3D nuclear position analysis of active and inactive alleles.
(A) schematic representation of measurements made in 3-D between the active and inactive alleles with the nuclear periphery and heterochromatic foci. (B,C) Bar graph showing the proportion of active (light grey) and inactive (dark grey) alleles associated with either heterochromatic foci (B) or the nuclear periphery (C) for 6 separate monoallelically expressed genes.

To further confirm that the nuclear periphery was not involved in distinguishing the active and inactive alleles, global analysis of the monoallelically expressed genes with respect to Lamin Associated Domains (LADs) defined in NPCs (Peric-Hupkes et al., 2010) was performed. Approximately 10% of monoallelically expressed genes were found within a LAD, however this frequency was similar to what is seen for all assessable genes (figure 5.14A on the next page). To test whether monoallelically expressed genes were positioned at LAD boundaries which hypothetically could become either active or repressed upon differentiation, the distance of the gene to the closest LAD was determined, however there was no difference between monoallelically expressed genes and all assessable genes (figure 5.14B). These results suggest that the nuclear lamina is not a major determinant of monoallelically expressed genes.

Finally, gross chromatin compaction of the active and inactive alleles was determined by measuring both the volume and the intensity of the DNA-FISH signal in the RNA-DNA FISH images. The volume of the alleles was determined in 3-D using the polygon finder tool in Softworx software. There was no statistically significant difference in the volume of the active and inactive alleles of *Gas6* (5.15A on page 139) with approximately 50% of cells showing a greater volume for the active allele, and the remaining 50% showing greater volume for the inactive allele. Nor was there a difference in the signal intensity of the active and inactive alleles (figure 5.15B) with no striking pattern seen in the pairwise comparisons of active and inactive alleles within the same nucleus. Therefore, at the gross level of light microscopy, the active and inactive alleles of *Gas6* have similar overall chromatin compaction. These results, however, are not of sufficient resolution to rule out local chromatin compaction differences, for example at the promotors of these genes. Other methods, such as DNAse I hypersensitivity assays, would be able to provide this resolution, however such types of assays cannot be performed in an allele specific manner.

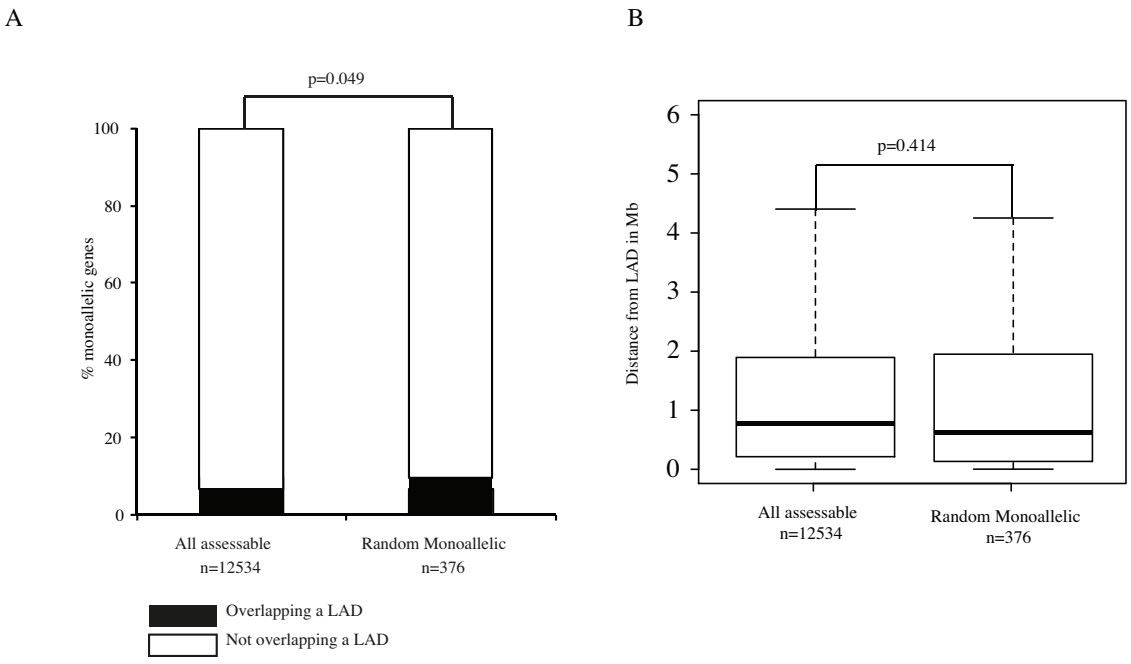
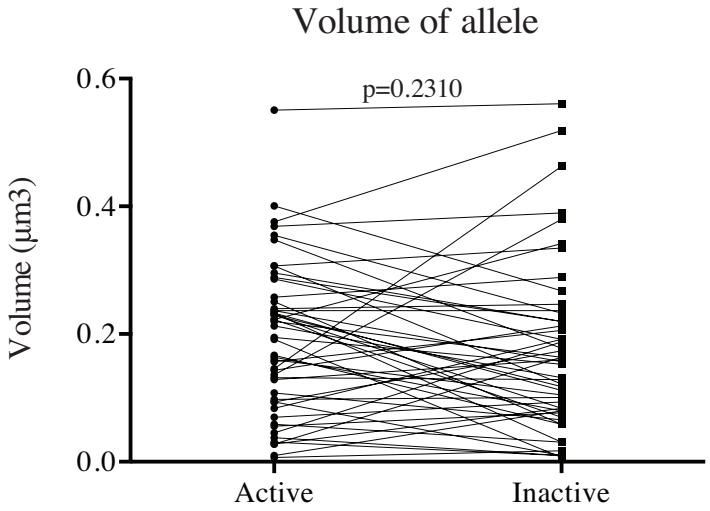


Figure 5.14: Association of monoallelically expressed genes with Lamin Associated Domains. (A) proportion of monoallelically expressed genes and all assessable genes located within a LAD (black) or outside of a LAD (white). (B) minimal distance of all assessable versus monoallelically expressed genes to the closest LAD.

A



B

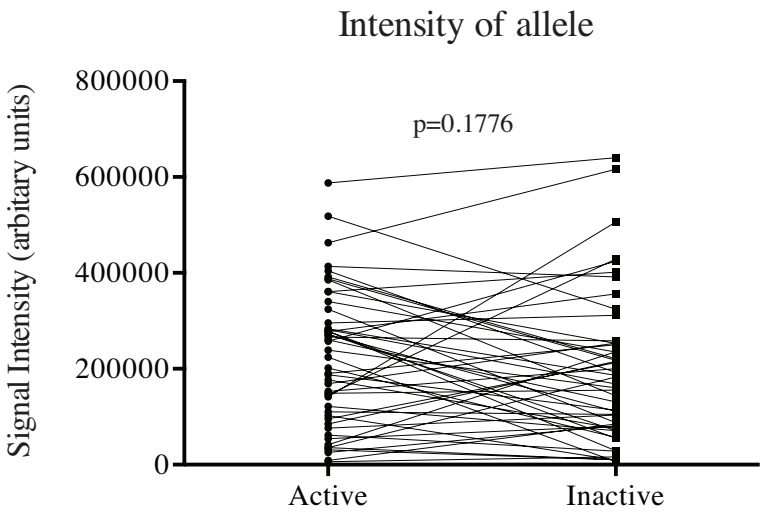


Figure 5.15: Gross chromatin compaction of active and inactive alleles.
(A) comparison of the volume of the active and inactive alleles of *Gas6*. Measurements were performed in 3D using Softworx image analysis software. (B) Signal intensity of the active and inactive alleles. Each dot represents one allele, lines connect active and inactive alleles within the same cell. p-values are not statistically significant (paired two-tailed t-test). Dots represent individual measurements, lines connect measurements made in the same nucleus.

5.4 Epigenetic drug screen

In order to gain some insights into the molecular mechanisms involved in the maintenance of monoallelic expression across cell generations, a small molecule epigenetic drug screen was performed. This utilised a library of 69 small molecule inhibitors (Cayman Chemical 1106) targeting a broad range of enzymes implicated in epigenetic inheritance including methyltransferases, demethylases, histone acetyltransferases, histone deacetylases and acetylated histone binding proteins (see Appendix, page 257). A test trial was first performed by treating one of the NPC clones with 10μM drug for 4 days in 6 well plate format, changing medium and drug every 2 days. Morphological changes were recorded every day, and RNA collected at the end of the 4 days. For each sample, cDNA was synthesised and SNP-PCR performed for two genes which were known to be monoallelically expressed in that particular clone: *Gas6* and *Serpinh1*. Drug treatment for 12 of the 69 compounds resulted in cell death at this concentration, and another 30 showed either slowed growth (n=20), elongated morphology (n=6) or spontaneously formed neurospheres in suspension (n=4). Due to the lack of an internal positive control, it was not possible to definitively conclude that a compound did not lead to activation of the silent allele as the conditions for each compound had not been optimised. Thus a major caveat of the epigenetic drug screen is that we cannot be sure that the compounds are active at the dosage and time points investigated, however any positive results would provide an indication as to a potential mechanism to pursue further. Unfortunately many sequencing traces were not of sufficient quality to be able to assess allelic expression and so not all samples could be analysed in each trial. However, 6 compounds resulted in a small level of expression of the silent allele for at least one of the two genes tested (figure 5.16 on page 142). Interestingly, all 6 compounds resulted in elongation of the cells, and 5 out of 6 of these compounds targeted histone deacetylases: M 344 preferentially inhibits HDAC6 and to a lesser extent HDAC1; EX-527 is a selective inhibitor of SIRT1; Pimelic Diphenylamide 106 is a slow tight-binding inhibitor of class I HDACs; chidamide is a generic HDAC inhibitor; and MS275 preferentially inhibits HDAC1, and to a lesser extent HDAC3, but not HDAC8. The 6th compound resulting in partial activation of the silent allele was decitabine. This compound is an analogue of 5-azacytidine and thus inhibits DNA methyltransferases. As 5-azacytidine itself did not lead to allele activation

nor to an elongated morphology both in this trial and in previous experiments (figure 5.5), it is possible that this sample may have been switched with the adjacent sample which was treated with JGB174, also a HDAC inhibitor.

A second trial was subsequently performed, lowering the dosage to 5μM for those compounds that caused cell death, and increasing to 25μM those which had no noticeable effect on cell morphology or growth rate. Additionally, a different NPC clone was used and SNP-PCR for *Cap2* was performed instead of *Serpinh1*, in addition to *Gas6* which was assayed in both trials. The second trial confirmed partial activation of 3 out of the 6 compounds identified in the first trial (M 344, chidamide and MS275), as well as other compounds including the HDAC inhibitors AGK2, CAY10603, Tenovin-1, CBHA, HNHA and CAY10433, although again the allele activation was minimal. The top 6 candidate compounds from the two trials, all targeting HDACs, were then tested at variable concentrations (5μM, 10μM, 20μM, 25μM and/or 40μM) and time points (2, 4 and 8 days) in order to optimise conditions to achieve maximal reactivation. However, in this third trial, no activation was observed for any of the drugs for any of the conditions tested. New compounds were ordered from an alternative source for 4 separate HDAC inhibitors (M344, pimelic diphenylamide 106, valproic acid and SAHA) and treatment was repeated. Unfortunately, only 1 of the 4 compounds at 1 condition (M-344 for 4 days at 10μM) lead to partial activation of the silent allele. Furthermore, treatment with HDAC inhibitors for 20 or 60 minutes, did not lead to activation of the silent allele of *Pla2g7* for all 4 compounds tested. Due to the small level of activation and lack of reproducibility of results between trials, the small molecule inhibitor screen was no longer pursued. In the future, use of a reporter system, such as a knock-in GFP at one of the two alleles, would provide a useful system for high throughput screening of epigenetic modifiers involved in the maintenance of the inactive allele, either through small molecule inhibitor libraries or through shRNA libraries.

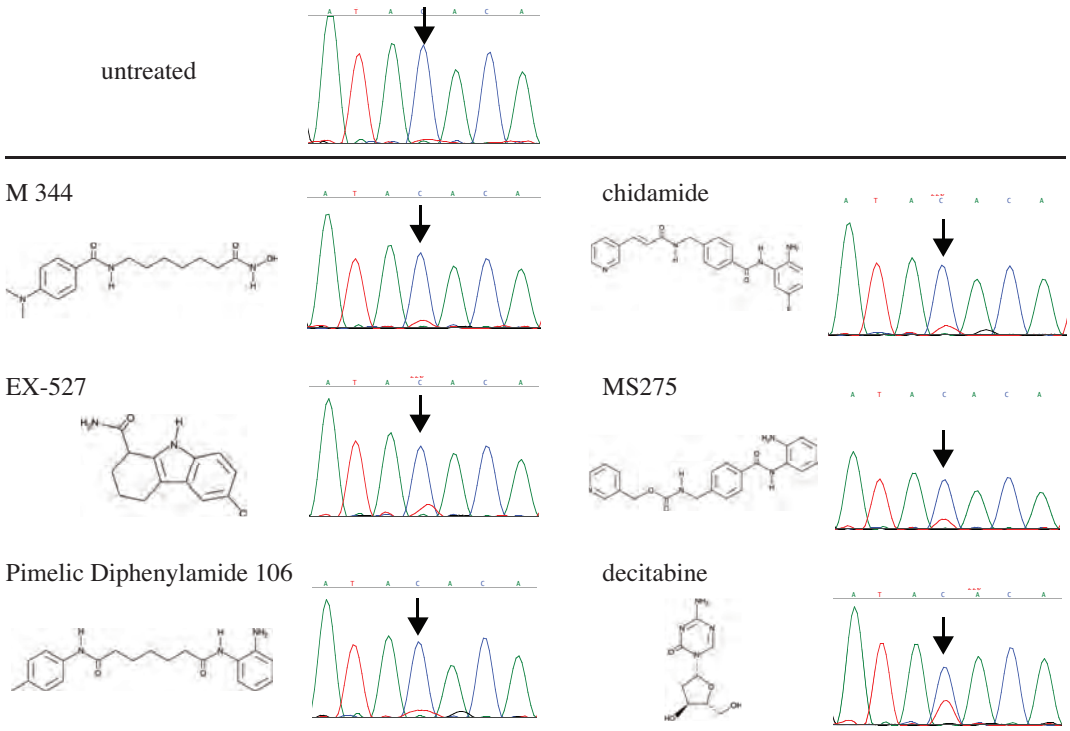


Figure 5.16: Selected results from epigenetic drug screen.
Drug structures and *Gas6* SNP-PCR sequencing traces for 6 compounds from the epigenetic drug screen which led to partial activation of the silent allele (arrows). Untreated sample is shown as a reference. All compounds, except for decitabine, target histone deacetylases.

Chapter 6

Transcriptional Consequences of Monoallelic Gene Expression

One of the potential functional consequences of monoallelic gene expression would be to alter the total transcript levels of the particular gene product. By halving the number of active alleles in the cell, it could be expected that the total transcript levels would also be halved. Alternatively, the cell may respond to the altered gene dosage such that the total mRNA levels approach those of biallelic cells. Previous genome-wide assessments of monoallelic expression have reported an overall reduction in transcript levels of monoallelically versus biallelically expressing cells, however rather than the expected 50%, this reduction was only 30-35% (Gimelbrant et al., 2007; Jeffries et al., 2012; Li et al., 2012b). Therefore it is possible that while others are dosage sensitive with expression levels correlating with the number of active alleles, some genes compensate transcriptionally to match biallelic expression levels.

6.1 Identification of transcriptionally compensating monoallelically expressed genes

First, the expression levels of NPC clones across the range of d-scores was assessed (figure 6.1 on page 145). It is possible for transcripts not classified as monoallelically expressed to have clone(s) with a |d-score| greater than the 0.18 cutoff if they fall into either category C, in which all clones

biased towards same allele, or did not pass the additional filtering in category B. Additionally, transcripts could have a high |d-score| yet not pass the p-value significance threshold. In this way, comparisons between the monoallelically expressed genes and all assessable genes can be made across the entire range of d-scores. At low d-scores, representing equal transcription of the two alleles, expression levels between all assessable transcripts and monoallelically expressed transcripts are similar (figure 6.1). However, at higher d-score values, representing biased or exclusive transcription from only one of the two alleles, transcript levels of the monoallelically expressed genes are greater than that of all assessable genes. Furthermore, while transcript levels for all assessable genes at high d-scores are approximately half of those at low d-scores, the monoallelically expressed genes show a smaller reduction, suggesting that there is not a clear 1:1 correlation between the number of active alleles and transcript levels for the monoallelically expressed genes.

Next, the behaviour of the monoallelically expressed genes was investigated more closely. For each of the monoallelically expressed genes, the clone with the highest and lowest d-scores were selected, and the difference in d-scores, Δd-score=$|d - score_{max}| - |d - score_{min}|$. The majority (531 out of 602 or 88%) of monoallelically expressed transcripts have a Δd-score of 0.3 or greater. Next the ratio of expression for the two clones was calculated. As expected, genes in which there was not a large variation in the extent of monoallelic expression, that is have a Δd-score$<$0.1, had the same level of expression in both the clones (expression ratio = 1, so $\log_2$(expression ratio) = 0). If the level of transcript in the cell followed the number of active alleles, then those genes with |Δd-score|$>$0.4 would be predicted to have an expression ratio of 0.5 (figure 6.2A, red line). However, there was a shift in the distribution of expression ratios away from 1 towards, but not centered around 0.5 (figure 6.2A, blue line). This suggests that there is a range in the transcriptional response of genes to monoallelic expression, with some genes showing dosage sensitivity and others dosage compensation. Importantly, this effect was not influenced by low abundant transcripts, as the same analysis performed on highly expressed transcripts with FPK$>$50 showed similar results (figure 6.2B). Note that the results of genes with Δd-score from 0.1-0.2 should not be over-interpreted as this bin contains only 10 genes.

Next, linear regression analysis was used to compare expression levels across clones with different extents of allelic imbalance. This allowed for the identification of the random monoal-

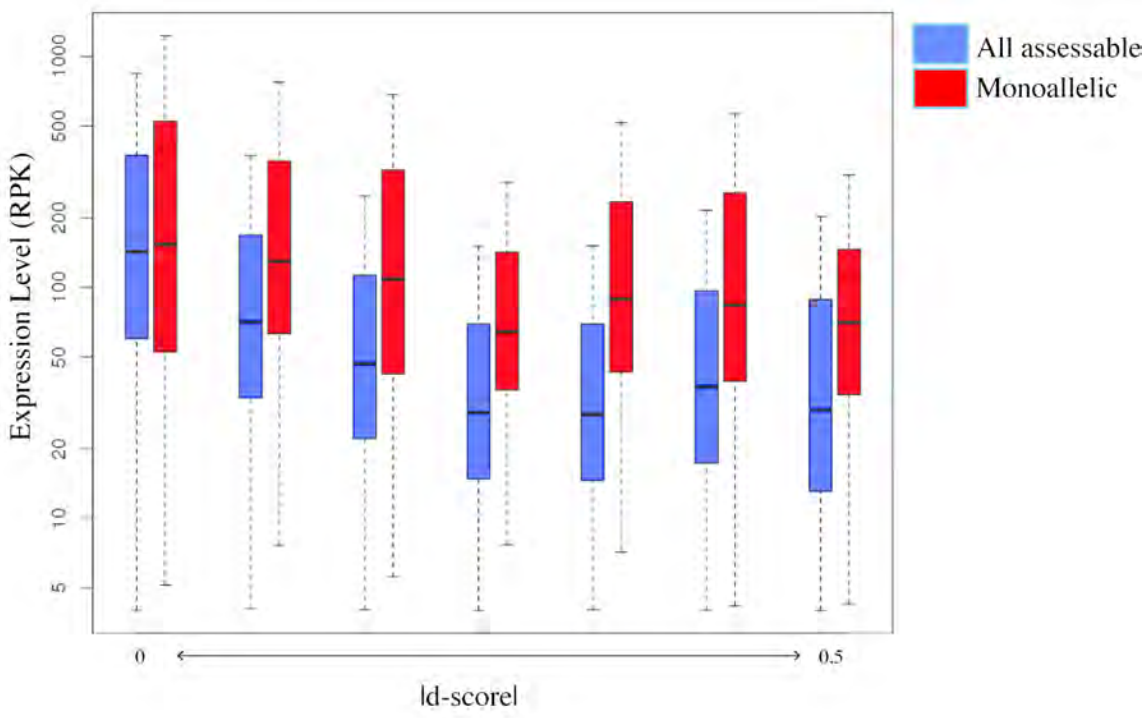


Figure 6.1: Influence of d-score on transcript levels.
Expression levels of transcripts of individual NPC clones placed into different arbitrarily sized bins based on the |d-score| ranging from 0 to 0.5. Blue represents all assessable transcripts, red represents monoallelically expressed transcripts.

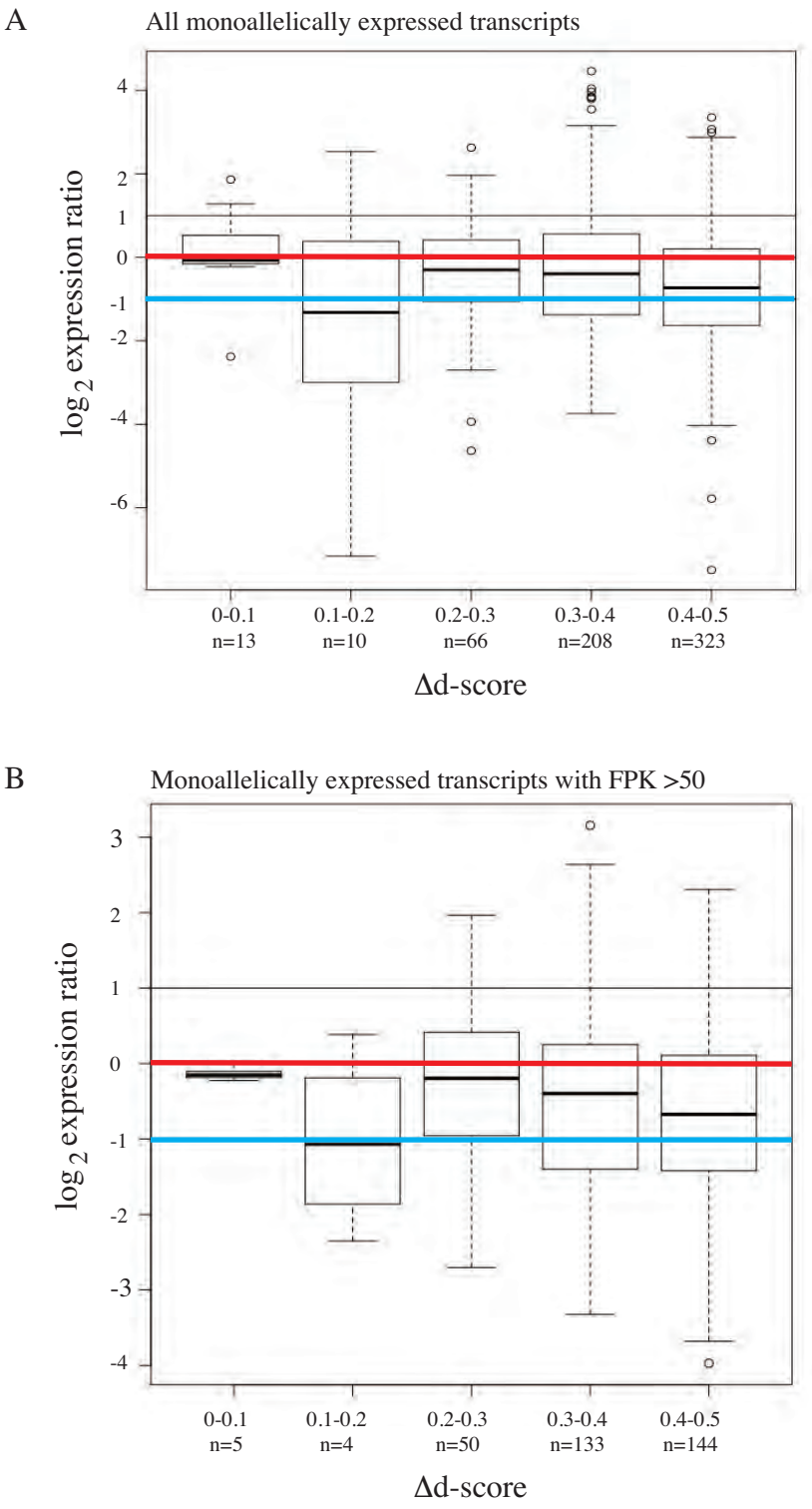


Figure 6.2: Transcript expression ratio between biallelic and monoallelic clones. Expression ratios of monoallelically expressed transcripts between the clone with the largest and smallest |d-score|. Transcripts are binned based on this difference between the largest and smallest |d-score| (Δd-score). An expression ratio of 1 (red line) denotes equal expression between the two clones, where as an expression ratio of 0.5 (blue line) denotes a 50% reduction, predicted if the expression level of the gene follows the allele dosage. (A) all monoallelically expressed transcripts, (B) Highly expressed monoallelically expressed transcripts with FPK>50.

lelically expressed transcripts showing transcriptional compensation versus dosage sensitivity. Through linear regression analysis, estimates of the slope (a) and the y-intercept (b) were calculated and the ratio $\frac{a}{-b}$ calculated. Compensating transcripts in which the expression level is similar across all clones will have a ratio close to 0 (figure 6.3A, red line), while those which are dosage sensitive will have a ratio close to 1 (figure 6.3A, blue line). Ratios were calculated for all 602 monoallelically expressed transcripts in NPCs (figure 6.3B), and arbitrarily defined thresholds were used to classify transcripts as either compensating ($-0.35 < \frac{a}{-b} < 0.35$), dosage sensitive ($0.75 < \frac{a}{-b} < 1.25$), intermediate ($0.35 < \frac{a}{-b} < 0.75$), over-compensating ($\frac{a}{-b} < -0.35$), or over-sensitive ($1.25 < \frac{a}{-b}$) (figure 6.3B). Of the 74 monoallelically expressed transcripts in ESCs, 18 transcripts corresponding to 18 genes were classified as compensating, and another 13 transcripts or 11 genes were classified as dosage-sensitive, corresponding to 24.3% and 17.6% of transcripts respectively (figure 6.4A on page 149). Interestingly, in NPCs, there were only 62 transcripts, or 30 genes, corresponding to 10.3% or 8% of monoallelically expressed transcripts or genes respectively. This is approximately 2 fold less than in the pluripotent ESCs. Instead, there were more transcripts which were dosage sensitive (140 transcripts or 23.3%) or over-sensitive (223 transcripts or 36.9%). Examples of transcripts which are over-compensating (figure 6.4B), compensating ((figure 6.4C), intermediate (figure 6.4D), dosage-sensitive (figure 6.4E) or over-sensitive (figure 6.4F) are shown. Interestingly, the genes that exhibited transcriptional compensation were enriched for DNA binding proteins and transcription factor activity, although the confidence of enrichment was low (p-value 0.037). There was no enrichment for the genes that undergo transcriptional compensation in ESCs, although this could be due to the limited power of gene ontology analysis with low number of genes.

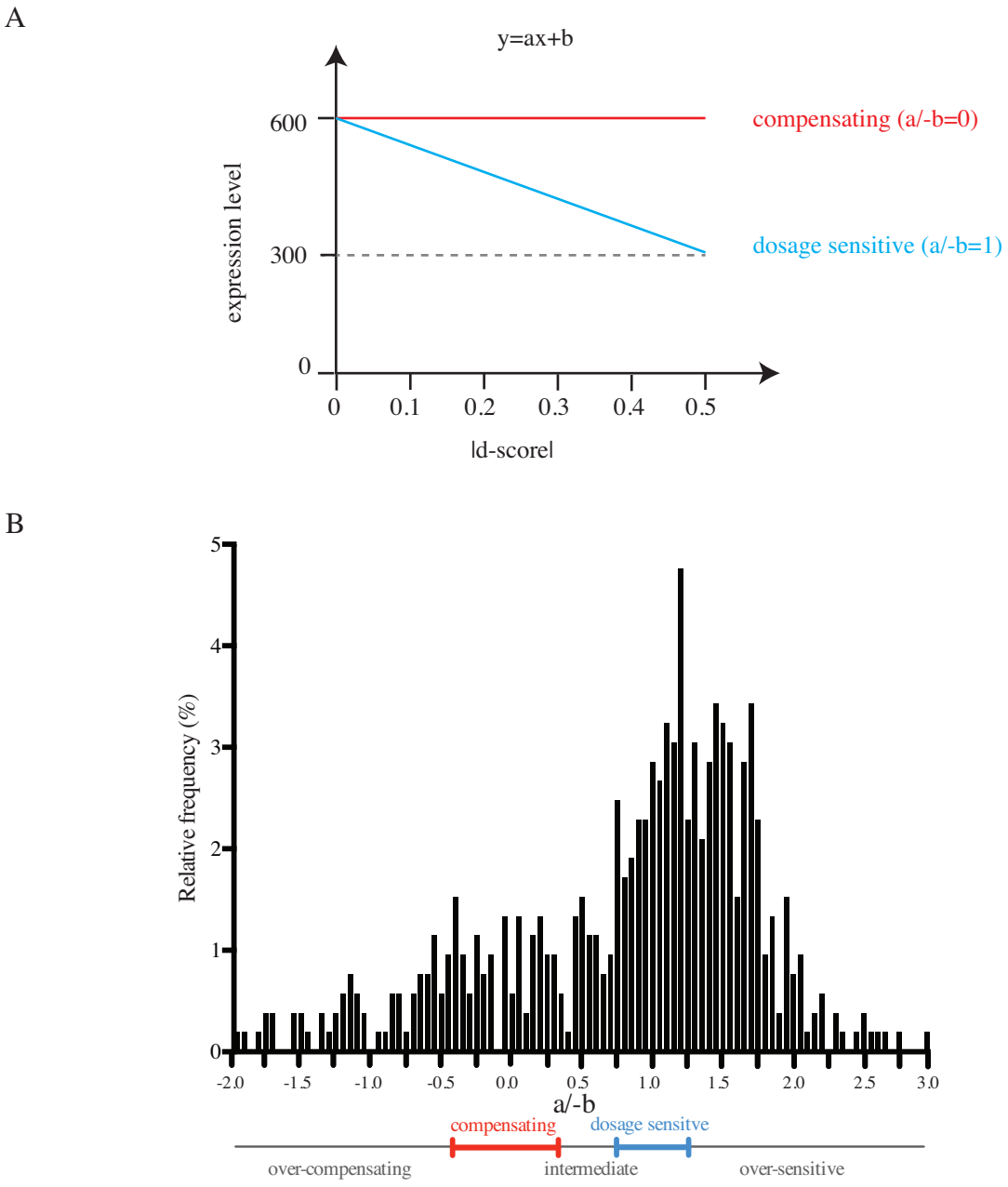


Figure 6.3: Defining transcripts which undergo transcriptional compensation.
Linear regression analysis was performed to identify transcripts that undergo transcriptional compensation (red) or are dosage-sensitive (blue). (A) hypothetical example showing a compensating transcript (red), in which there is equal expression levels across all clones of variable |d-score|, and a dosage-sensitive transcript (blue), in which the monoallelic clone (|d-score|=0.5), has half the expression level as a biallelic clone (|d-score|=0). The ratio of the slope (a) over the y-intercept (-b) was computed and used to classify the transcripts. (B) histogram showing the distribution of NPC monoallelically expressed transcripts across the spectrum of (a/-b) ratios. Those with a ratio between -0.35 and +0.35 were classified as compensating (red), and those with a ratio between 0.75 and 1.25 as dosage-sensitive (blue). Transcripts were also classified as intermediate ($0.35 < a/-b < 0.75$), over-compensating ($a/-b < -0.35$), or over-sensitive ($1.25 < a/-b$).

A

	over-compensate	compensate	intermediate	[illegible]	over-sensitve
ESC number transcripts	27	18	10	15	6
ESC percentage	36.5	24.3	13.5	17.6	8.1
NPC number transcripts	131	62	47	140	223
NPC percentage	21.8	10.3	7.8	23.3	36.9

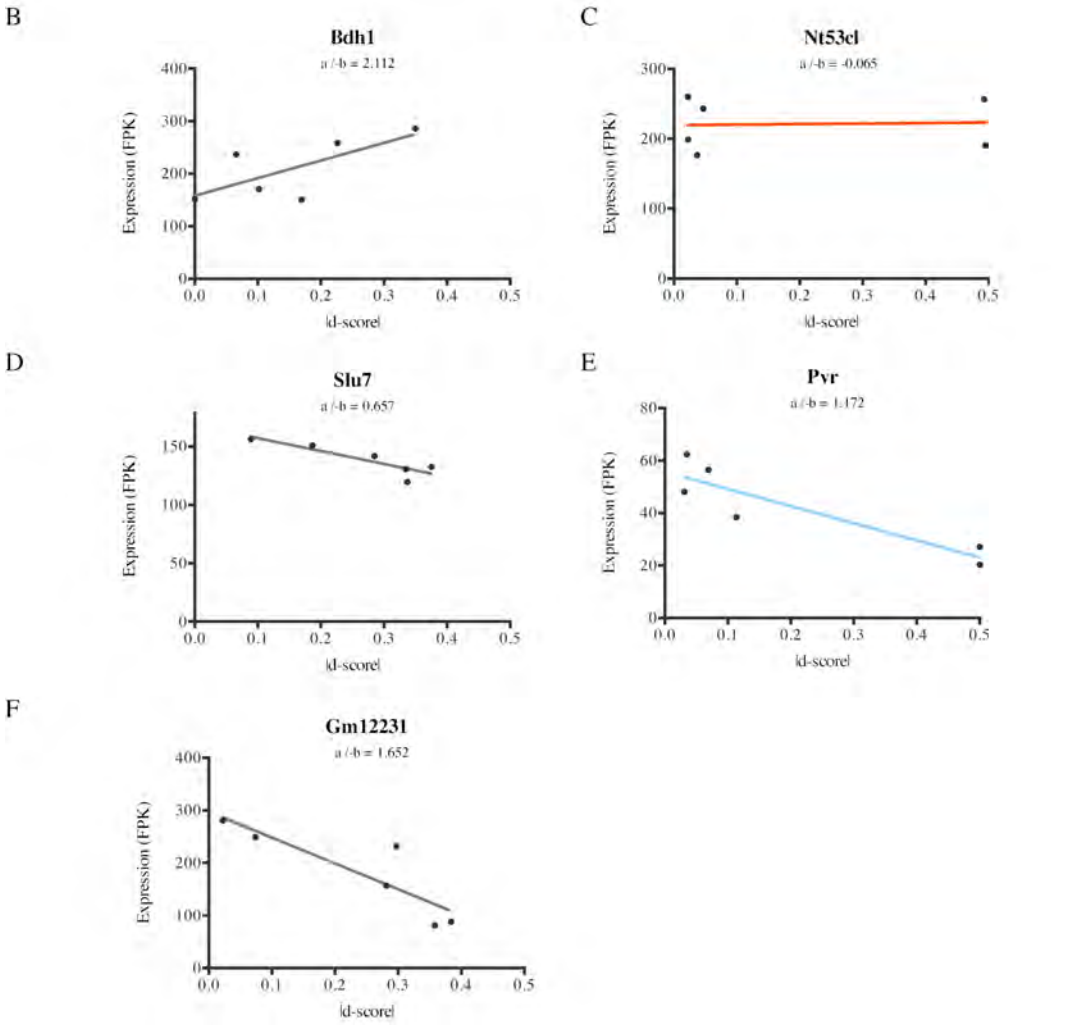


Figure 6.4: Linear regression analysis of transcript expression levels.
(A) summary of results from the linear regression analysis of expression levels versus d-scores. Examples of NPC monoallelically expressed transcripts which are over-compensating (B), compensating (C), intermediate (D), dosage-sensitive (E) and over-sensitive (F) are shown.

6.2 Validation of transcriptional compensators and dosage-sensitive transcripts

One of the major caveats in the approach used to classify transcripts as compensating or dosage-sensitive is that there are only 6 data-points used for the linear regression analysis, as only 6 clones were analysed. Furthermore RNA-seq expression data for only one replicate per clone was available for analysis. Therefore the amount of variation in expression levels of a given transcript is unknown, which makes interpreting a potential 50% reduction in transcript levels across 6 samples difficult. For these reasons, the RNA-seq analysis was independently validated by quantitative RT-PCR for 20 different genes. For each of the 6 clones, 3 biological replicates representing different passages were analysed. All data was normalised to the geometric mean of at least 3 housekeeping genes. Of the 20 genes assessed, 12 were in agreement with the linear regression analysis. This included 10 genes classified by linear regression analysis as transcriptional compensators, of which 6 were validated by Q-RT-PCR (figure 6.5A). The remaining 4 genes classified as compensators by RNA-seq analysis, were either dosage sensitive (*Fkbp7, Ttc4* and *Ptgr1*) or over-compensated (*Pcsk6*) by Q-PCR analysis (figure 6.5B). Another 3 genes were classified as dosage-sensitive by RNA-seq analysis (figure 6.6A), of which one (*Pla2g7*) was validated by Q-PCR, and the other two genes (*Acyp2* and *Pdzrn4*) represent false-negatives as they were transcriptional compensators by Q-PCR analysis. There was another 3 genes (*Rhoj, Acot1* and *Cap2*) which were over-sensitive by both RNA-seq analysis and Q-PCR analysis (figure 6.6B), and four genes which were over-compensating by RNA-seq analysis (figure 6.6C), of which two were compensating by Q-PCR (*Thrsp* and *Gas6*), and two were dosage-sensitive (*Rgs16* and *Serpinh1*). In summary, there was a 60% validation rate of the linear regression analysis by Q-RT-PCR, with 4 false-positive and 2 false-negative transcriptional compensators. Therefore, while there are some caveats of the RNA-seq based linear regression analysis, the fact that some genes exhibit transcriptional compensation is validated. Future experiments looking at a larger number of clones and replicates by high-throughput microfluidic-based Q-PCR technologies, would enable a more accurate assessment of transcriptional compensation of the monoallelically expressed genes.

Next, I examined whether the transcriptional compensation was reflected in the levels of

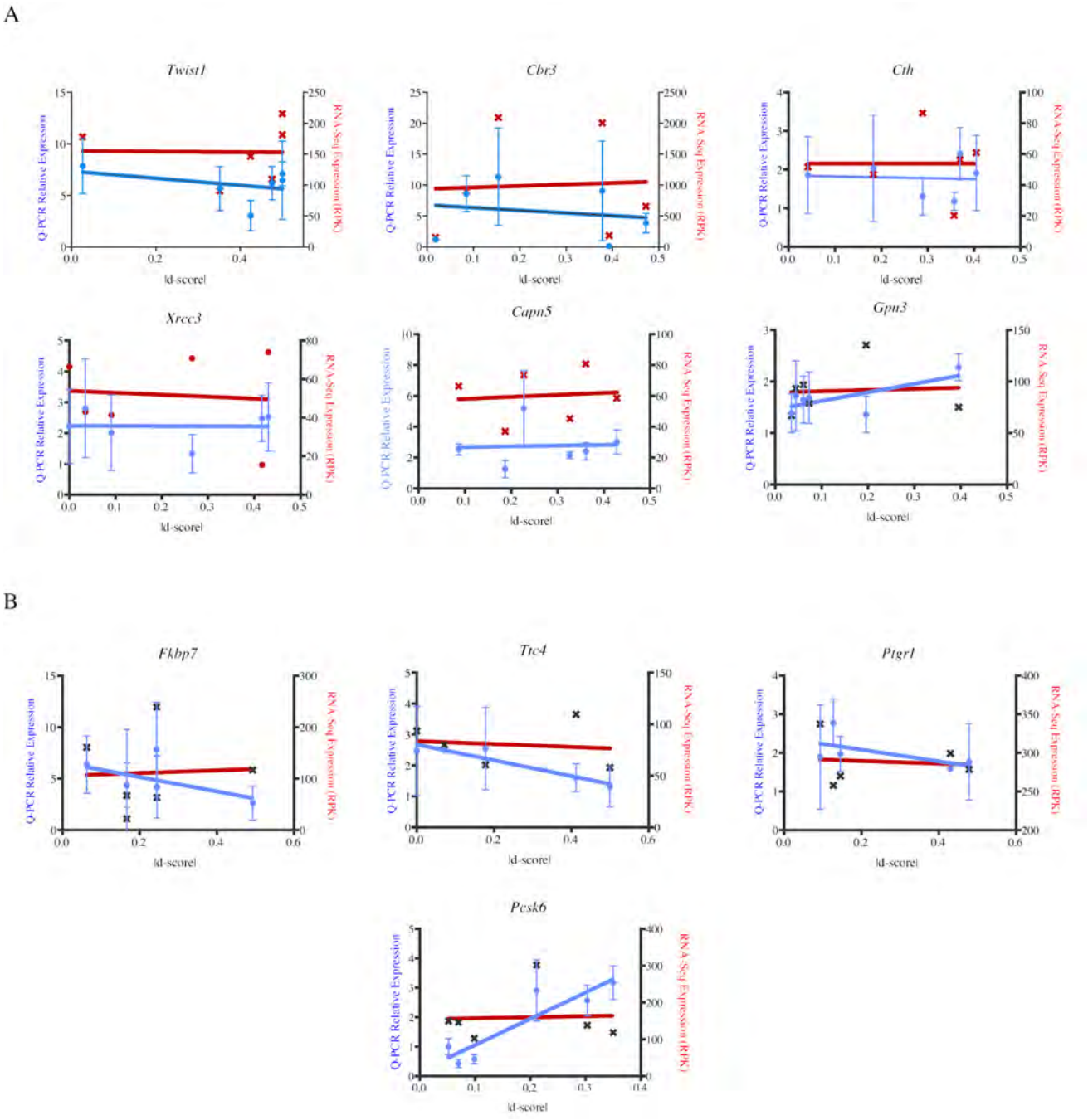


Figure 6.5: Validation of transcriptional compensation by Q-RT-PCR.
Q-RT_PCR validation of RNA-seq linear regression analysis for transcripts classified as transcriptional compensators showing those which (A) validated and (B) did not validate. Blue circles represent Q-RT-PCR data (mean +/- SD of 3 biological replicates), red crosses represent RNA-seq data.

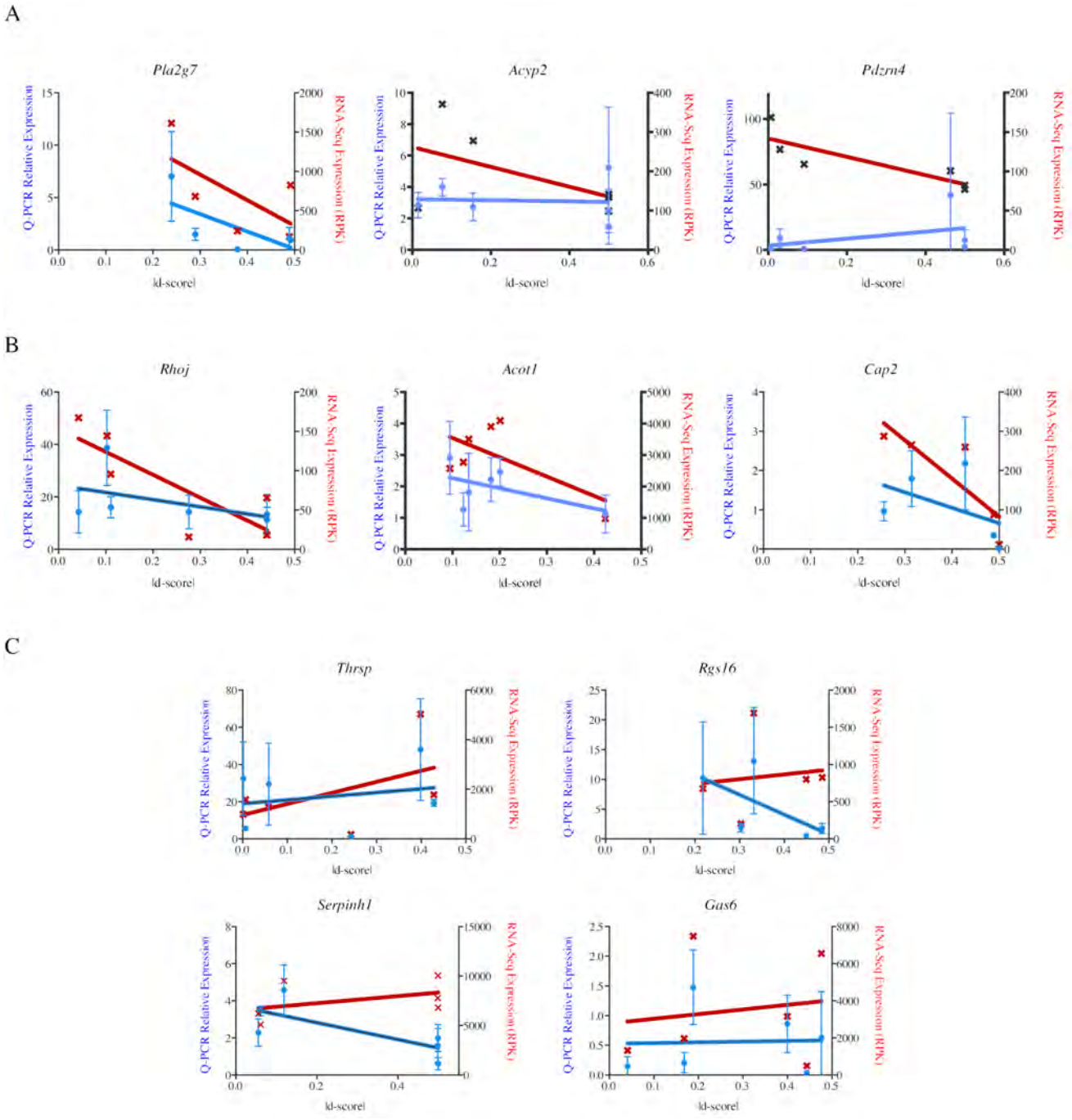


Figure 6.6: Validation of RNA-seq linear regression analysis by Q-RT-PCR.
Q-RT-PCR validation of RNA-seq linear regression analysis for transcripts classified as (A) dosage-sensitive, (B) dosage-oversensitive and (C) over-compensating. Blue circles represent Q-RT-PCR data (mean +/- SD of 3 biological replicates), red crosses represent RNA-seq data.

protein in the cell. Western blotting experiments was performed to compare the levels of monoallelically expressed genes across clones and a non-clonal population of cells (figure 6.7 on the following pageA, B). Quantification of the Western blots was performed by determining the ratio of the band's signal intensity between the gene of interest and a control gene such as Lamin B1 (figure 6.7C, D). Antibodies for two genes *(Cbr3* and *Rgs16)* gave clear and consistent signals in NPC clones. *Cbr3* was classified as a transcriptional compensator by both RNA-seq and Q-PCR analysis (figure 6.5A). This was reflected in the protein levels in the clones as monoallelically expressing, allele-biased and biallelic clones showed similar levels of protein as the non-clonal control sample (figure 6.7A, C). *Rgs16* was classified as over-compensating by RNA-seq linear regression analysis, and over-sensitive by Q-PCR analysis (figure 6.6C). Protein levels in the monoallelically expressing clone were similar to the non-clonal sample, yet higher than in an allele-biased clone (figure 6.7B, D). Therefore it is likely that for *Rgs16*, the exact levels of protein are not critical for the cell, and instead there is a high degree of variability in expression levels both of the mRNA and the protein across samples.

Finally, compensation was confirmed at the level of mRNA transcription in single cells by RNA-FISH analysis. Cells from monoallelic and biallelic clones were mixed together in equal amounts, and seeded onto coverslips for RNA-FISH analysis. This enabled control for any differences in hybridization efficiency or imaging biases introduced by preparing samples independently. For each site of transcription, visualized as an RNA-FISH spot, the z-slice containing the maximum signal intensity was identified. The total signal intensity for the individual allele in that slice was then measured (figure 6.8A on page 156), and the background signal intensity subtracted. The normalized signal intensity between single alleles was then compared between monoallelic and biallelic cells (figure 6.8B). There was a statistically significant 2-fold increase in the signal intensity of alleles in monoallelic cells when compared to biallelic cells for two separate monoallelic candidates (figure 6.8B). *Gas6* was identified as over-compensating by RNA-Seq analysis and compensating by Q-PCR analysis (figure 6.6C), and RNA-FISH analysis confirmed that this observed compensation was due to a two-fold upregulation of the single active allele in monoallelically expressing cells, such that overall transcript levels are comparable to biallelically expressing cells (figure 6.8). The other gene which showed transcriptional upregulation by RNA-FISH analysis was *Cap2*, a gene identified as dosage-sensitive by both RNA-sequencing

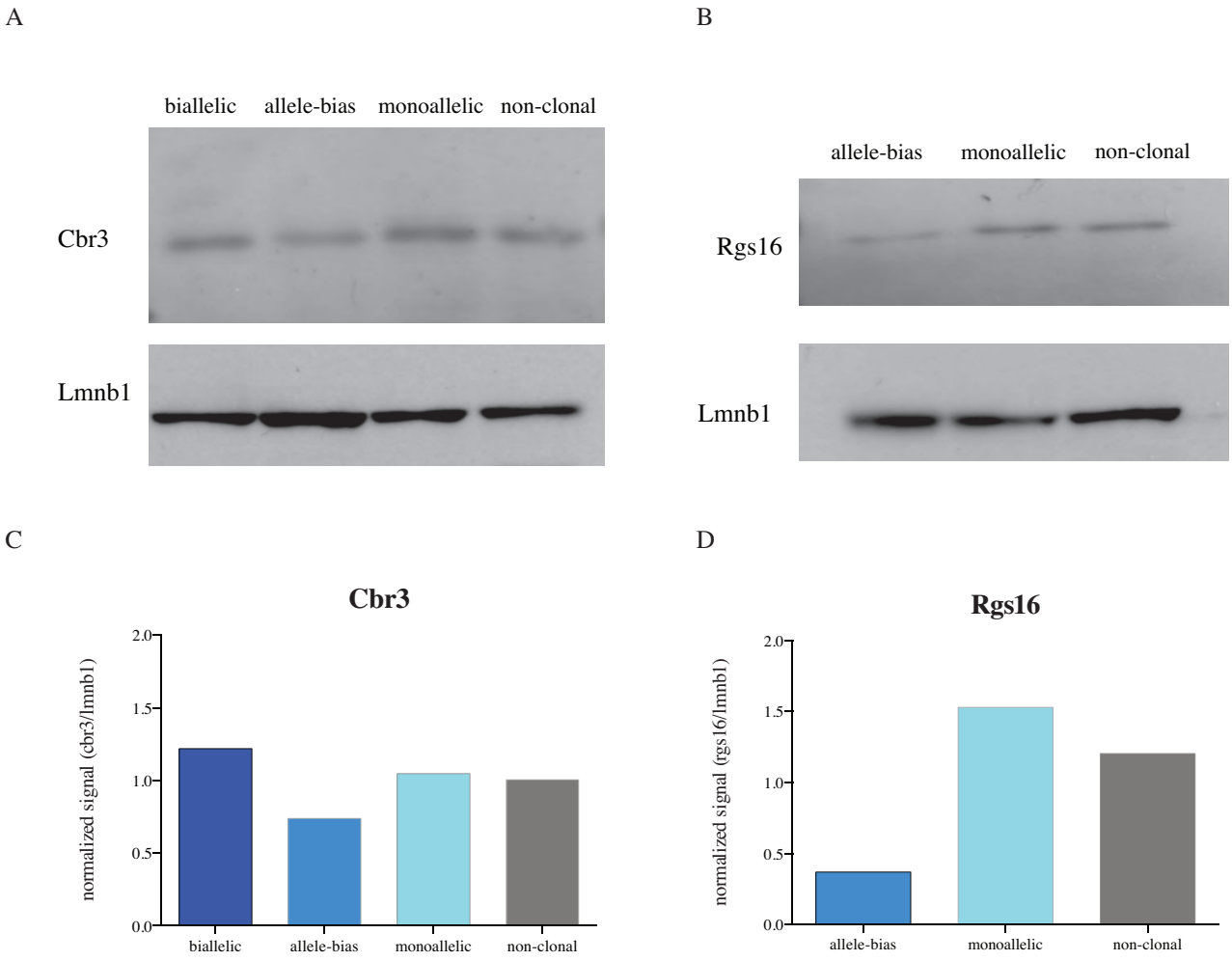


Figure 6.7: Validation of transcriptional compensation by Western blotting. Western blotting analysis for two monoallelically expressed genes Cbr3 (A) and Rgs16 (B) across different NPC clones and non-clonal control. Lamin B1 is used as loading control. Quantification of western blots for Cbr3 (C) and Rgs16 (D) relative to Lamin B1 control is shown.

and Q-PCR analysis. This discrepancy between the two assays could be due to having only 4 clones that were assessable for the RNA-sequencing and Q-PCR analysis, some of which showed high variation in expression levels for this gene (figure 6.6B). In fact, removal of the clone with |d-score| of 0.5 would completely change the results to either compensation (RNA-seq) or over-compensation (Q-PCR), again illustrating the main caveat of the linear regression analysis. Finally, *Ror2*, the third gene assayed by RNA-FISH, showed equal transcription coming from a single allele in both monoallelically and biallelically expressing cells (figure 6.8B). Consistent with these findings, *Ror2* was identified as dosage-sensitive by RNA-sequencing linear regression analysis. In this way the total amount of *Ror2* mRNA in the cell follows the number of equally expressing alleles.

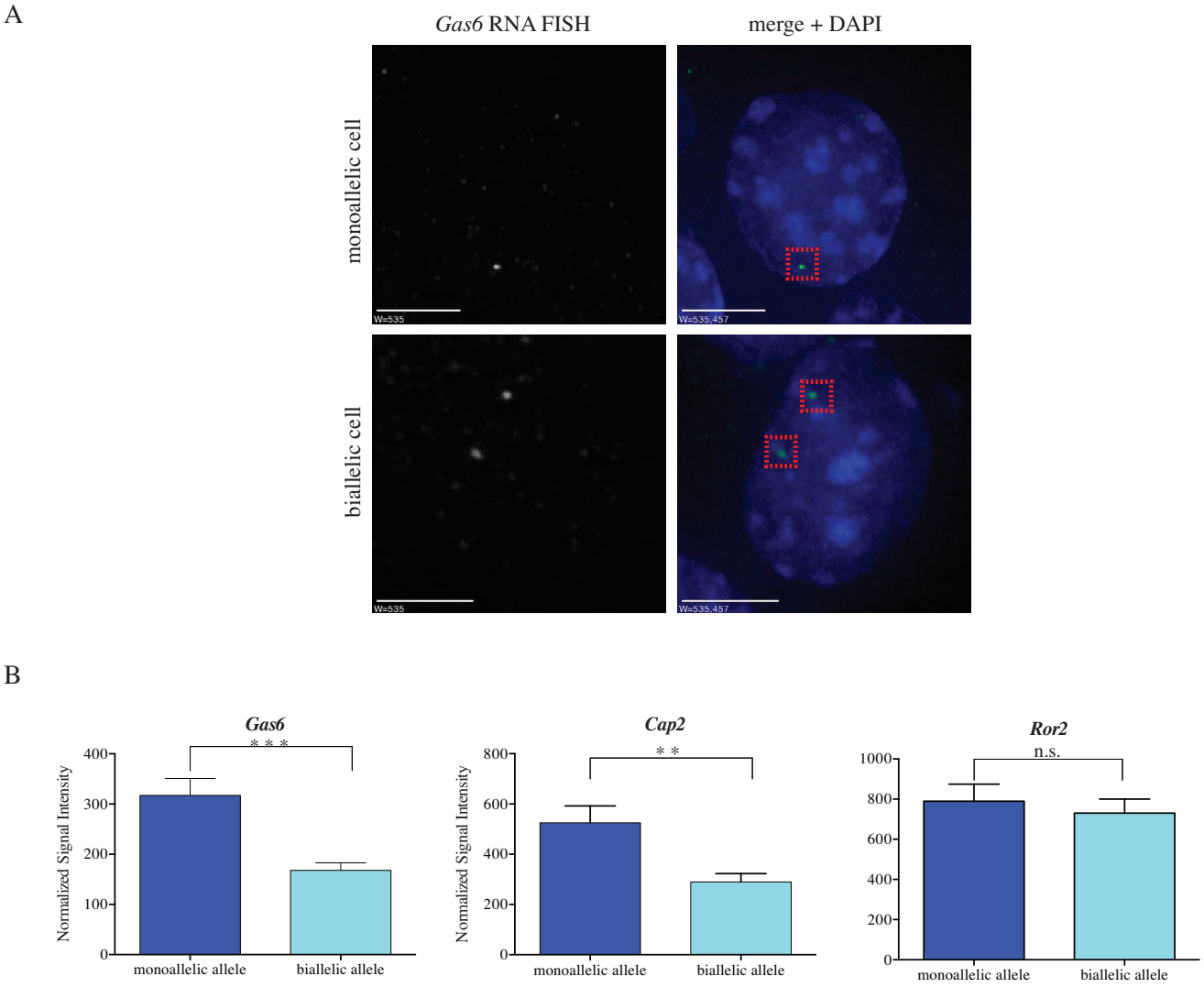


Figure 6.8: Validation of transcriptional compensation by RNA-FISH.
(A) Measurements of transcriptional outputs of single alleles was performed by calculating the maximum signal intensity of the RNA-FISH hybridisation signal in a single slice for monoallelically and biallelically expressing cells. Signal intensity was normalised to the mean of three nucleoplasmic control measurements. Three genes were tested, *Gas6* (B) and Cap2 (C) show a two fold increased output of the single active allele in monoallelically expressing cells compared to a single allele in a biallelically expressing cell, while transcriptional output was similar for alleles of *Ror2* (D). Error bars represent standard error of the mean, n=30. ** $p<0.01$, *** $p<0.001$, Mann-Whitney two-tailed t-test.

Chapter 7

Conclusions and Perspectives

7.1 Summary

Identification of random inheritable monoallelic expression during differentiation of mouse embryonic stem cells (ESCs) to neural progenitor cells (NPCs) was performed through an allele-specific RNA sequencing screen. While just 67 genes were monoallelically expressed in ESCs, this increased 5.6 fold to 376 genes in NPCs, representing 3.0% of the assessable transcriptome, indicating that the establishment of monoallelic expression occurs during early development. There was little to no overlap between monoallelically expressed genes in ESCs and NPCs, supporting cell-type specificity of monoallelic expression. Furthermore, 72% of NPC monoallelically expressed genes switched from biallelic expression in ESCs to monoallelic expression during differentiation. DNA methylation was not sufficient to explain the mitotic inheritance of monoallelically expressed genes, nor was there evidence for differential nuclear positioning of the active versus inactive alleles. However, histone modifications differentially marking the two alleles not only provide an epigenetic signature of monoallelic expression, but could also provide a mechanism through which monoallelic expression could be maintained across cell divisions. Interestingly, overal transcript levels of only a subset of monoallelically expressed genes followed active allele dosage. In contrast, for 8% of monoallelically expressed genes, transcript levels remained similar between monoallelic and biallelic states. For these genes the single active allele is up-regulated to preserve the biallelic levels of the respective mRNA in the cell. This supports a model where stochastic gene regulation during dynamic ESC differentiation results

in monoallelic expression, and for some genes, the cell is able to compensate transcriptionally to maintain the required level of expression of these genes.

7.2 Discussion

I propose that random monoallelic expression exemplifies a stochastic aspect of gene regulation that takes place upon the initiation of specific differentiation programs, that result in global changes in gene expression and chromatin condensation. If the probability of gene activation is less than 1, this would result in a mixed population of cells containing 0, 1, or 2 active alleles. Once established, if not detrimental to the cell, these allele expression patterns could become set epigenetically, and the transcriptional states subsequently maintained across cell divisions and propagated clonally (figure 7.1 on the following page). Probabilistic models of random monoallelic expression have been proposed for other examples of monoallelic expression, including *Albumin* in hepatocytes (Michaelson, 1993), the interleukin genes in T cells (Guo et al., 2005), and Ly49 receptors in NK cells (Held and Raulet, 1997). In all cases, the two alleles are independently regulated with a low activation probability, possibly due to limiting quantities of key activating factors. One consequence of this independent regulation is that it generates both monoallelic and biallelic cells in a mixed population. Indeed, at least one biallelic clone is observed for almost all monoallelically expressed genes, consistent with an independent stochastic regulation model. As monoallelic expression is stochastic, it is likely that genes have different propensities to become monoallelic, depending on the cell type. Consistent with this, screens performed in neuronal cell types (Jeffries et al., 2012; Li et al., 2012b; Lin et al., 2012; Wang et al., 2010) have a higher degree of overlap with our study than those performed in more distant cell types, such as lymphoblasts and fibroblasts (Gimelbrant et al., 2007; Zwemer et al., 2012). The outcome of monoallelic expression for some genes may be unfavorable if the cell requires a specific level of transcript that cannot be accommodated for by the single active allele, thus resulting in cell death. However, for those genes for which either the exact level of transcript is not critical, there are functional redundancies with other genes, or for those that are able to compensate transcriptionally, monoallelic expression represents a viable outcome for the cell.

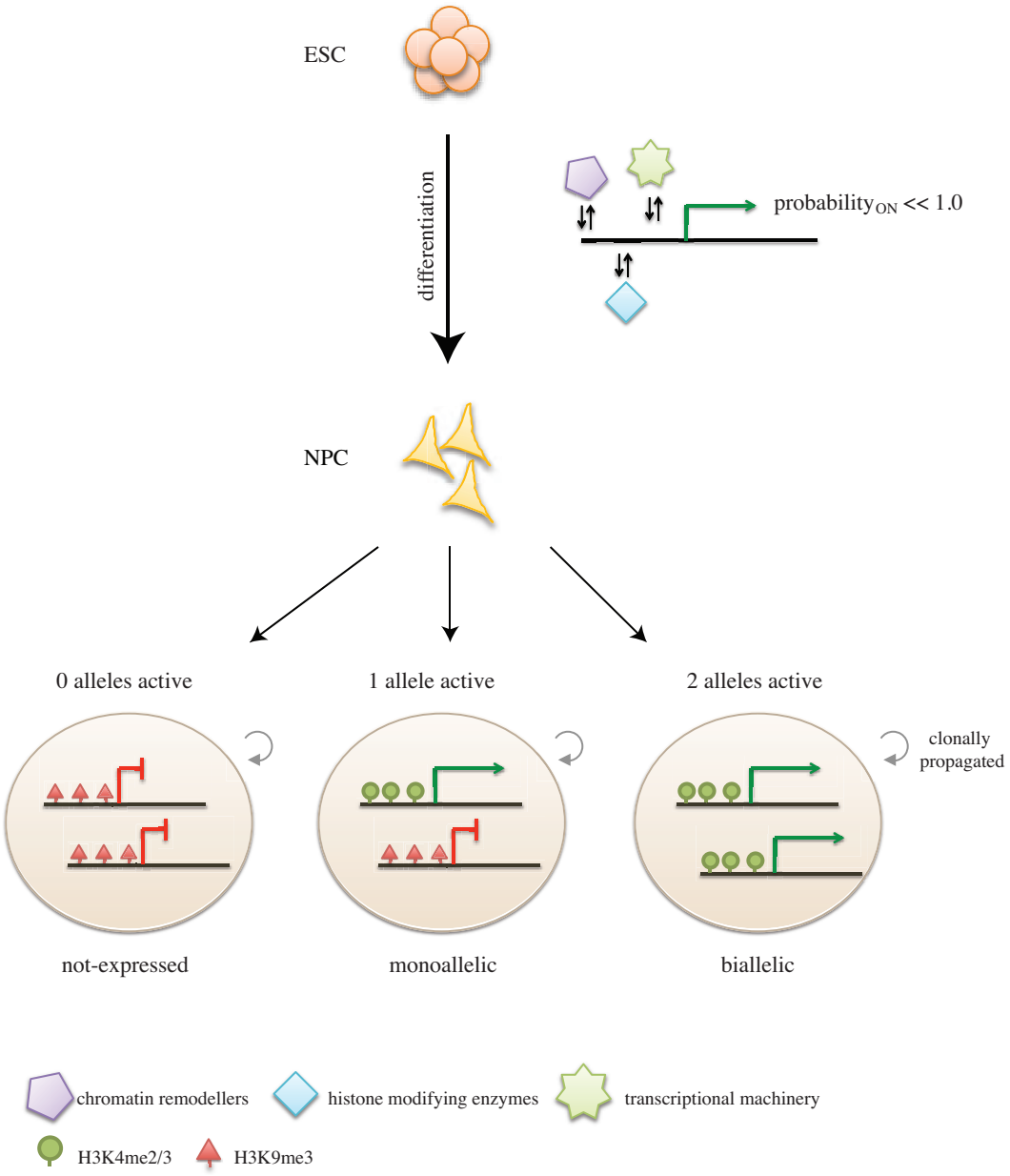


Figure 7.1: Model for random monoallelic gene expression during differentiation. Differentiation of ESCs to NPCs involves dramatic changes to the gene expression program. For those genes in which the probability of gene activation is less than 1, potentially due to limited accessibility or availability of chromatin remodelers (purple pentagons), histone modifying enzymes (blue diamonds) or the transcriptional machinery (green stars), three outcomes exist in the differentiated cell types. Either no alleles are activated (no expression), 1 allele is active and one remains inactive (monoallelic expression) or both alleles are active (biallelic expression). If these states are viable for the cell, either through transcriptional compensation of the monoallelically expressed genes, redundancies with other gene products or due to flexible transcriptional requirements of the cell, these states may become mitotically stable, such that they are propagated across cell divisions in a clonal manner, potentially through the acquisition of either active H3K4me/2 (green circles) or inactive H3K9me3 (red triangles) histone modifications.

The results in this study both support and add significantly to the findings of previously published genome-wide screens (table 7.1 on the next page). By performing an RNA-sequencing based screen, the number of assessable genes is substantially greater than for those screens performed using microarray technologies in which only 1,358 (Zwemer et al., 2012) to 3,939 genes (Gimelbrant et al., 2007) were able to be assessed. This is in contrast to the approximately 14,000 genes that were able to be assessed in this study. Therefore in this way, RNA-sequencing based approaches enable a more complete global picture of random monoallelic gene expression. This study identified a similar frequency of random monoallelic expression to that of other RNA-sequencing based approaches in neural cell types (Li et al., 2012b; Lin et al., 2012), however it was substantially less than that reported using microarray based approaches in human and mouse lymphoblasts and fibroblasts (Gimelbrant et al., 2007; Zwemer et al., 2012). This is likely not due to the technology platform used, as another microarray based approach in neural stem cells reported similar frequencies (Jeffries et al., 2012) to the 3% of assessable genes identified in this study. Instead it likely represents differences in the stringency thresholds and criteria used to define random monoallelic expression. Our approach has been very stringent in order to minimise the number of false-positives, but in this way likely underestimates the number of monoallelically expressed genes. For example, the vast majority of monoallelically expressed genes identified by Gimelbrant *et al.* and Zwemer *et al.* (78% and 91% respectively) were called using only 1 available SNP. As these studies only required 1 biased clone for a gene to be classified as monoallelically expressed, they likely have a higher rate of false positives in their screens (Gimelbrant et al., 2007; Zwemer et al., 2012). In contrast, my screen required adequate coverages (at least 5 reads) per SNP for a transcript to be assessable, and there was on average 6 SNPs per transcript (see section 4.1 on page 81), increasing the robustness of the screen. However, this may have led to a higher rate of false negatives in the screen. An alternative possibility for the observed differences in frequencies, although unlikely, is that there are more monoallelically expressed genes in fibroblasts and lymphoblasts compared to neural progenitor cells. Additional screens across different cell types, both *in vivo* and in cell culture systems, are required to gain further insight into how random autosomal monoallelic gene expression changes both during development and between cell types.

Consistent with previous findings, we see no genomic clustering of the random monoalleli-

Table 7.1: Comparison of features of monoallelically expressed genes identified with previously published screens.

	Gimelbrant et al. 2007 Science	Wang et al. 2007 PLoS ONE	Wang et al. 2007 PLoS ONE	Li et al. 2012 PLoS ONE	Lin et al. 2012 PLoS ONE	Jeffries et al. 2012 Stem Cells	Zwemer et al. 2012 Genome Biology	Eckersley-Maslin et al. 2013
technology platform	microarray	concurent methylated and unmethylated DNA	concurent methylated and unmethylated DNA	RNA-sequencing	RNA-sequencing	microarray	microarray	**RNA-sequencing**
species	human	mouse	mouse	mouse	human	human	mouse	**mouse**
cell type	lymphoblast fibroblast placenta	CNS	brain neural stem cells	neural stem cells	iPS cells and derived neurons	neural stem cells from cerebral cortex, striatum and spinal cord	lymphoblast fibroblast	**ESCs and NPCs**
number genes / percentage of assessed genes	371 genes 5-10%	5 genes 1%	9 genes 1-2%	170 genes 2.4%	314 - 801 genes (includes X-linked genes)	143 - 203 genes ~2%	212 genes >10%	**67 (<0.5%) in ESCs 376 (3.0%) in NPCs**
chromosome coordination	no	no	-	-	-	-	no	**no**
gene ontology	transmembrane proteins	-	-	glutathione receptors, annexins, protocadherins	-	transmembrane glycoproteins	none	**small increase in signalling and glycoproteins**
proportion genes strictly monoallelic	<20%	40%	<40%	35%	low	low	low	**<1%**
DNA methylation	-	assumed	assumed	-	-	correlative evidence for higher methylation levels	-	**occassionally correlated but not essential**
Histone modifications	-	-	-	-	-	correlative evidence for increased H3K27me3, decreased H3K4me3, H3K9ac, H3K9me3	-	**H3K4me2/3 on active, H3K9me3 on inactive**
Expression levels of monoallelic clones	less	-	-	30-35% less	-	weak but significant decrease	-	**10% transcripts compensate, 23% dosage sensitive**

cally expressed genes, nor any coordination in expression biases along the chromosomes (table 7.1). There is an enrichment for signaling proteins and glycoproteins, similar to that observed in previously observed (Gimelbrant et al., 2007; Jeffries et al., 2012; Li et al., 2012b), however this enrichment is small and likely not of great biological significance. Importantly, less than 1% of the random autosomal monoallelically expressed genes identified demonstrated strict monoallelic expression in every single clone assessed. That is, for the vast majority of monoallelically expressed genes, at least 1 clone was biallelically expressed. This is consistent with previous findings and supports a model in which random autosomal monoallelic expression is a consequence of stochastic low probability gene expression. Finally, this study has advanced our understanding of monoallelic expression beyond the published literature by providing substantial advancements into the understanding of both the mitotic inheritance of the allelic imbalance, but also the consequences on expression levels at these genes.

One key finding of this study is that there is significantly less monoallelic expression in pluripotent ESCs compared to lineage committed NPCs, supporting the original hypothesis that the establishment of random autosomal monoallelic expression occurs during early development. ESCs are unique in their developmental plasticity which is reflected in a more open and dynamic chromatin and promiscuous gene expression profiles (reviewed in (Fisher and Fisher, 2011; Mattout and Meshorer, 2010; Meshorer and Misteli, 2006)). The ESC colonies are highly heterogeneous both in terms of transcriptional profiles, but also in developmental potentials (Huang, 2011; Martinez Arias and Brickman, 2011). Their genome is in general transcriptionally hyperactive, with low level expression of lineage-specific genes and normally silenced repetitive elements, which become repressed upon differentiation (Efroni et al., 2008). Within a population of ESCs, cell states range from pure unrestricted pluripotent potential, to a poised state in which the cell has commenced to proceed towards a specific differentiation pathway. This heterogeneity is hypothesised to reflect properties of the pluripotent state (Martinez Arias and Brickman, 2011), rather than being a feature of stem cells in culture, as similar heterogeneities exist in the preimplantation embryo (Guo et al., 2010; Plusa et al., 2008). In this way, these heterogeneities may reflect trapped transition states, that is, intermediates during cell fate decision in which a cell exhibits a mixture between the state of origin and the state of destination (Arias et al., 2013). As the ESCs within the colony cycle between these different

states of developmental potential (Canham et al., 2010), it is therefore not surprising that there is significantly less stably inheritable random monoallelic expression in ESCs compared to more static differentiated cell types. While the initial frequency of random monoallelic expression resulting from stochastic gene expression may be similar to that of differentiated cell types, these allelic imbalances would not be clonally propagated and maintained as efficiently in ESCs given the dynamics in transcriptional profiles and developmental potential. It would be of interest to determine whether the initial frequency of random monoallelic expression, while not necessarily maintained, is similar in ESCs compared to other cell types, an experiment which will become more feasible with the advances made in single-cell RNA sequencing technologies. Furthermore, it will be of interest to study how monoallelic expression changes not only during mouse development, but also across both closely related and distant organisms, and how, given the potentially lowered evolutionary constraint of these genes, monoallelic expression may play a role in speciation, not only of mammals (Keverne, 2009; Renfree et al., 2013) but also for other species (Georgy and Widdicombe, 2002).

Although DNA methylation is perhaps the best understood mechanism by which transcriptional states can be inherited (reviewed in (Smith and Meissner, 2013)), and has been demonstrated to be important for other examples of monoallelic expression, including genomic imprinting (reviewed in (Kelsey and Feil, 2013)), a global role for DNA methylation in maintaining monoallelic expression for the genes identified in this study was not observed. However, it does remain possible that selected genes are regulated in this way. While allele-specific DNA methylation has been previously reported in the mouse CNS (Wang et al., 2007) and used to identify monoallelically expressed genes, including *Thrsp* which was also identified in this screen, it has never been demonstrated that DNA methylation drives monoallelic expression. My observations are consistent with a recent report that DNA methylation is not involved in maintaining active and inactive alleles of the monoallelically expressed *Cubilin* gene in kidney and intestinal cell lines (Aseem et al., 2013). Additionally, a recent genome-wide analysis of the interplay between genetic variation, DNA methylation and gene expression, did not find any evidence for allele-specific DNA methylation driving monoallelic expression in human cells in the absence of DNA sequence variation effects (Gutierrez-Arcelus et al., 2013). Therefore DNA methylation likely does not regulate random autosomal monoallelic expression.

However, I did observe that the active and inactive alleles were associated with different histone modifications: the active allele marked by H3K4me2/3 and the inactive allele by H3K9me3. Histone modifications have been shown to mark other examples of monoallelically expressed genes, consistent with our results. For example X-inactivated genes and imprinted genes are marked by promoter-restricted H3K4me2 (Rougeulle, 2003), and olfactory receptors by H3K9me3 and H4K20me3 (Magklara and Lomvardas, 2013). Interestingly, we did not see evidence for the Polycomb-associated H3K27me3 repressive mark at the promoters of inactive alleles of monoallelically expressed genes, whereas H3K9me3 was present. It remains to be determined whether these histone modifications are actively involved in the inheritance of the transcriptional state or if they are a byproduct of the respective state. Additionally, there may be yet undetected characteristics that distinguish the active and inactive alleles of monoallelically expressed genes that may also play a role in maintaining the difference in transcriptional state across cell divisions. The position and organization of alleles within the three-dimensional nucleus has been linked to transcriptional output (reviewed in (Hübner et al., 2013)), and while small differences in nuclear positioning between the active and inactive alleles has been observed for monoallelically expressed GFAP in astrocytes (Takizawa et al., 2008a), and nuclear organization has been implicated in regulating both olfactory receptor genes (Clowney et al., 2012) and immunoglobulins (Skok et al., 2001), I did not observe any difference in the position of active versus inactive alleles with respect to the nuclear periphery or heterochromatin. Thus, it is unlikely that nuclear organization is critical for maintaining the expression bias of the monoallelically expressed genes examined in this study.

One would predict that decreasing the gene dosage would lead to decreased expression levels of the particular gene. Therefore the observation of transcriptional compensation for some of the monoallelically expressed genes was surprising, albeit not without precedent. Transcriptional compensation has also been observed for examples of aneuploidy in non-mammalian systems, including Drosophila (McAnally and Yampolsky, 2010) and Maize (Guo and Birchler, 1994) in which mRNA levels do not strictly follow the dosage of the gene. Examples of transcriptional compensation can also been seen when comparing heterozygous knockout mice to their wild-type counterparts. For example, two genes we identified as monoallelically expressed in NPCs, *Mks1* (Wheway et al., 2013) and *Bag3* (Homma et al., 2006), both show comparable levels of protein

between wild-type and heterozygous knock-out mice. There are, however, examples of genes in which the heterozygous mice have reduced transcript levels, including the monoallelically expressed genes *Cth* (Kaasik et al., 2007) and *Cstb* (Ishii et al., 2010). Therefore, it is likely that transcriptional upregulation is not only gene specific but also cell type specific. For those genes that do not compensate, the precise level of gene product may not be critical for normal cell viability. Thus, in some cases, the cell is able to tune the transcriptional output of the respective allele in response to either genetic or epigenetic inactivation of the second allele to maintain overall expression levels (figure 7.2 on the next page). This has important implications, especially in the interpretation of copy number variants, as these may not necessarily result in a change of transcript and protein product.

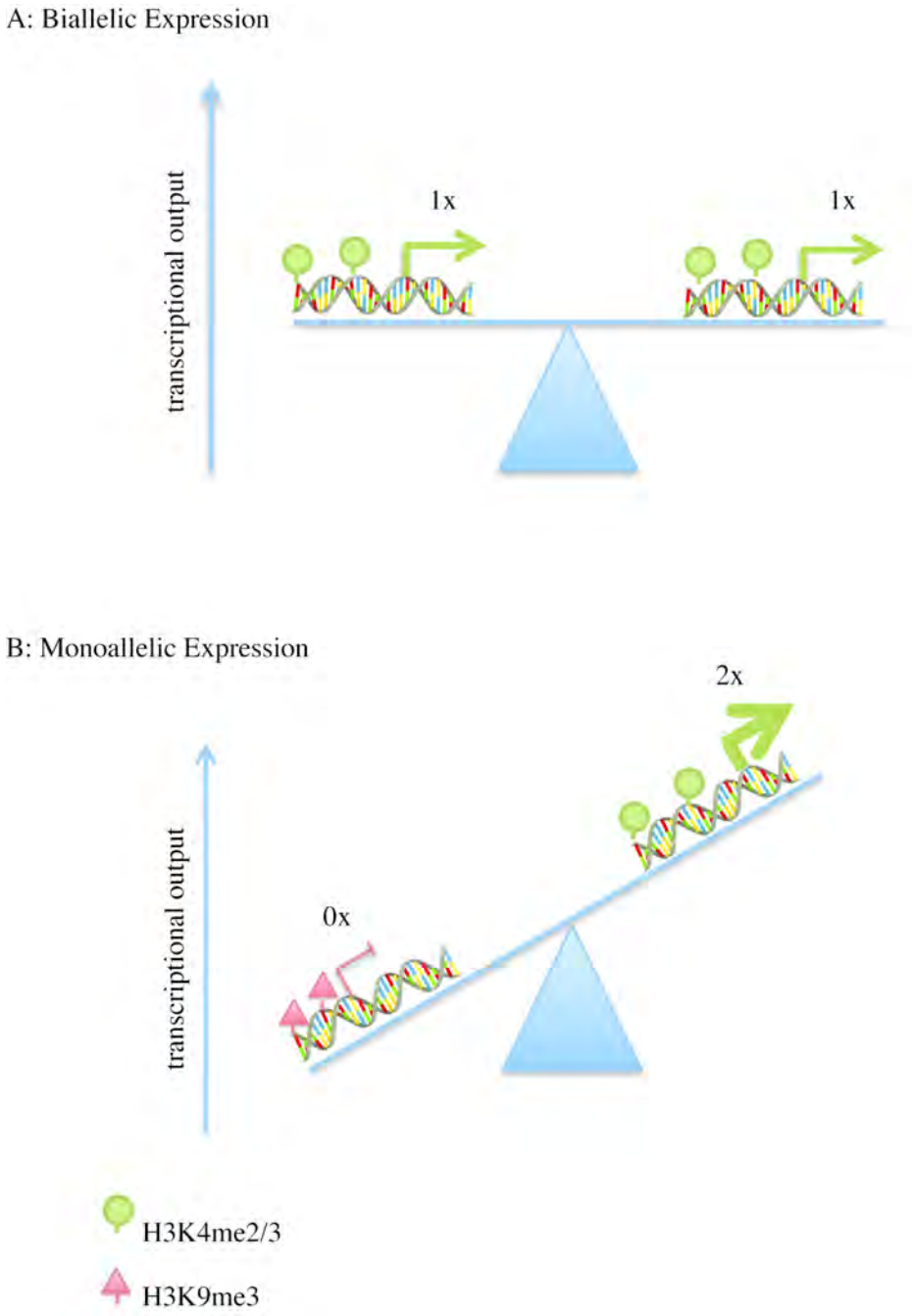


Figure 7.2: Transcriptional compensation of monoallelically expressed genes. Schematic showing the effect of number of active alleles of a transcriptionally compensating gene. (A) In biallelically expressing cells, transcriptional output is similar between the two alleles which are both marked by the active-associated histone modifications H3K4me2/3 (green circles). (B) In contrast, monoallelically expressed cells show expression from only one of the two alleles. This expressing allele has similar levels of H3K4me2/3 however is now expressing at twice the level of a single allele compared to the biallelic cell. In contrast, the inactive allele has lost the H3K4me2/3 modifications and gained the repressive H3K9me3 modification (red triangle).

7.3 Perspectives and future directions

At the outset of this thesis work, there were many unanswered questions regarding the characteristics, inheritance and significance of random autosomal monoallelic gene expression (table 1.1 on page 51). Many of these questions have now been resolved in this study (table 7.2 on page 169). One of the major questions in the field is how monoallelic gene expression is maintained mitotically through cell divisions. I have found that unlike most other examples of classic monoallelic expression, DNA methylation is not involved in the inheritance of the active and inactive alleles, although it may distinguish between the alleles. Nor is there a role for nuclear organisation of these genes, also in contrast to the classical examples of monoallelic expression. Instead H3K4me2/3 and H3K9me3 differentially mark the two alleles and are likely candidates for the maintenance of monoallelic expression. There was no enrichment for non-coding RNAs surrounding these genes, nor did they fall into clusters, further distinguishing them from the classical monoallelically expressed genes. Thus I propose that random autosomal monoallelic expression results from very different processes and for very different reasons from that which occurs at X-chromosome inactivation, imprinted genes, immunoglobulins and olfactory receptors. Rather than being an active process in the cell to either control gene dosage or generate receptor diversity, random autosomal monoallelic expression is a consequence of the stochastic nature of gene expression regulation. Low probability of gene activation may be due to a number of different reasons, such as limited transcription factor accessibility and/or availability, and has been described for a number of different genes such as the interleukins (Paixão et al., 2007). This low probability of activation results in a mixed population of cells, some of which express only from one allele, others are biallelic and some not expressing at all. Indeed single cell RNA-sequencing and Q-PCR analysis is generating a picture of high variability of gene expression levels in an otherwise seemingly homogeneous population of cells (Bengtsson, 2005; Kalisky et al., 2011; Raj and van Oudenaarden, 2009; Shalek et al., 2013). This variation in gene expression levels within single cells can be tolerated in several ways. First, the genes involved may have functional redundancies such that the exact levels of the genes are not critical. Alternatively, and what we see, there may exist tight feedback loops that act to ensure the critical level of the genes are reached. Furthermore, rather than being noise in the system, this

variability in gene expression levels may have important roles during development in cell fate decisions (Zernicka-Goetz and Huang, 2010).

Now that a comprehensive set of genes that are monoallelically expressed in both ESCs and NPCs has been identified, there are many exciting future directions that could now be taken with this project. It would be interesting to determine how extensive and similar monoallelic gene expression is in other developmental contexts. This could be performed either using a similar approach to that in this thesis, that is allele-specific RNA sequencing of single-cell derived clones, or alternatively as single-cell RNA sequencing technology becomes more sensitive, monoallelic expression could be assessed at the single cell level, both in cell culture as well as in tissues. By examining monoallelic expression in neighboring cells one could distinguish inheritable monoallelic expression versus transcriptional pulsing. It would also be interesting to perform time-course experiments during the course of differentiation to follow the kinetics of monoallelic expression. Again, this could be performed by allele-specific RNA-sequencing based approaches, or alternatively the dynamics of a specific monoallelically expressed gene of interest could be studied using a live-cell reporter system. In such a system, knock-in at the endogenous alleles of repeats of the 19 base pair MS2 stem loop sequence into the 3'UTR of the gene would allow visualisation of the site of transcription of that allele in single cells, through expression of the MS2-coat protein fused to a fluorescent reporter (Bertrand et al., 1998; Chubb et al., 2006; Janicki et al., 2004; Lionnet et al., 2011; Yunger et al., 2010). Insertion of different RNA stem-loop motifs, which would be recognised by different proteins, into the second allele would allow direct visualisation of transcription of the two alleles through the cell cycle, across cell divisions and during differentiation.

While I ruled out DNA methylation and nuclear organisation, it is not definite that the histone modifications identified play a causal role in maintaining monoallelic expression. Therefore future experiments are required to determine whether these marks are the cause or consequence of the transcriptional status of the allele. While siRNA mediated knock-down of the relevant histone methyltransferases could be performed, such an approach will have widespread effects on the overall transcriptional program of the cell which will confound any results making a direct causal link hard to establish. Instead targeting of the relevant histone methyltransferases or demethylases to the alleles could be performed by fusing these proteins to sequence specific

Table 7.2: Contribution of this work to understanding the features of random monoallelically expressed genes.
Contributions to the understanding of monoallelic expression in this thesis are highlighted in red.

	X-inactivation	**Genomic Imprinting**	**Immuoglobulins**	**Olfactory Receptors**	**Random Autosomal**
Genomic distribution	X-chromosome	Autosome, clusters	Autosome	Autosome, clusters	Autosome, dispersed
Choice of allele	random	parental	random	random	random
Mitotic inheritance	yes	yes	yes	n.a.	yes
Number of genes	~1,000	~150	7	~1400	100s - 1,000s
DNA methylation	yes	yes	yes	yes	**no**
Histone modifications	yes	yes	yes	yes	**H3K4me2/3 and H3K9me3**
Nuclear organisation	yes	yes	yes	yes	**no**
DNA replication timing	asynchronous	asynchronous	asynchronous	asynchronous	?
Non-coding RNAs	*Xist, Tsix…*	*Airn, Kcnq1ot1…*	no	no	**none annotated**
Function	dosage compensation	parental conflict hypothesis	receptor diversity	receptor diversity	**consequence of stochastic gene expression**
Mechanism	sotchastic	deterministic	deterministic	stochastic	**stochastic**

DNA binding proteins such as TAL effectors (Boch et al., 2009; Moscou and Bogdanove, 2009) or through use of the CRISPR system (Gilbert et al., 2013; Jinek et al., 2012). This will enable direct evidence for these marks in the maintenance of the monoallelic state. Furthermore, establishment of TALE- or CRISPR-based targeting of an activator could enable additional experiments in which the silent allele is induced, and the response to the cell investigated over time. TALE-activators targeting the monoallelically expressed *Gas6* gene were designed and successfully generated during the course of this study, however there were issues in transfection efficiencies in NPCs of these plasmids due to their large size. For these, as well as other technical reasons, a CRISPR based system is now being pursued in order to perform these types of endogenous locus manipulation experiments.

In order to determine whether there are other factors involved in the maintenance of monoallelic expression, a fluorescent-protein knock-in reporter line could be established to enable high-throughput screens to be performed. This system would ideally involve insertion of a fluorescent reporter protein within the 3'UTR of the monoallelically expressed gene at only one of the two alleles. Clones in which this fluorescent reporter is not expressed could be picked and used to screen for expression of the fluorescent protein, against either small molecule inhibitor libraries, such as was used for the pilot screen in this thesis, or through shRNA libraries targeting chromatin modifying enzymes.

Finally, for those genes that exhibit compensation, it will be of interest to determine the mechanisms by which transcriptional compensation maintains the total level of mRNA in the cell. One possibility is that the levels and accessibility of specific transcription factors themselves are critical in determining the frequency of gene transcription at the respective alleles. Alternatively, there may be feedback loops in which the levels of mRNA and/or protein can be sensed in the cell. For example, the tumor suppressor *Ini1* compensates transcriptionally in heterozygous cells (Guidi et al., 2004). This transcriptional compensation was shown to be dependent on the levels of *Ini1* mRNA in the cell, as expression of a transgene led to down-regulation of the endogenous alleles (Guidi et al., 2004). Additionally, *Brn3a*, which is not expressed in NPCs, has been shown to compensate transcriptionally through autoregulation (Trieu et al., 2003). An alternative possibility is that there is transcription factor feedback to regulate the activity of the upregulated allele. Experiments are currently underway to examine the effect of over-expression

of a transgene on the endogenous gene expression levels of those monoallelically expressed genes which exhibit transcriptional compensation.

7.4 Concluding remarks

This study provides the first in depth analysis of random autosomal monoallelic expression during the course of ESC differentiation. There are few monoallelically expressed genes in pluripotent ESCs, likely due to their highly dynamic chromatin and extensive expression profiles. Upon lineage commitment, as cells become more restricted in their expression, there is a random and stochastic activation or repression of alleles, which in some cases results in allelic-imbalance. This cell-type specific monoallelic expression is subsequently maintained epigenetically across cell divisions, likely through histone modifications that differentially mark the two alleles. For the majority of monoallelically expressed genes, the inactivation of one allele results in a decrease in gene product, however some genes are able to upregulate the single active allele to compensate for the allelic-imbalance. In this way, random monoallelic expression illustrates the remarkable plasticity and robustness of gene regulation in mammalian cells.

In conclusion, monoallelic expression represents an interesting consequence of stochastic gene expression for which in some situations, the cell is able to compensate for, demonstrating the remarkable plasticity of the cell in response to initiating developmental programs including changes in gene expression, chromatin structure and chromatin compaction. These findings have very important biological significance, not only when considering how gene regulation is carried out in single cells, but also in relation to heterozygous mutations and copy number variations involved in disease.

Chapter 8

Extended Materials and Methods

8.1 Cell culture

Mouse ESC lines (Table 8.1 on page 174) were grown on irradiated Mouse Embryonic Fibroblast (MEF) feeder cells pre-seeded for at least 24 hours on gelatin coated plates to give a confluent monolayer of cells (approximately 1.5 x 10^6cells per 10cm plate) in MEF medium (DMEM high glucose, 10% FBS, 1x P/S). ESCs were grown in ESC medium (500ml medium contains 425ml Knockout DMEM, 75ml FBS, 5ml NEAA, 5ml PSG, 500µl LIF (ESGRO 10^6units) and 7µl β-mercaptoethanol) and require passaging approximately every 2-3 days before colonies come into contact with each other and start to differentiate. Prior to RNA or protein isolation, ESCs were soaked 1-2 times for 1 hour each on gelatin coated dishes to remove MEF feeder cells which typically adhere faster than ESCs. The non-adherent cells, containing predominantly pluripotent ESCs, can then be collected for subsequent analysis. ESCs cultured in 3i medium (GlobalStem) were passaged at least 6 times in the absence of MEFs prior to experiments being conducted.

ESC differentiation time courses were performed by plating soaked pluripotent ESCs onto gelatin coated plates in differentiation medium (for 500ml medium mix 450ml DMEM-GlutaMax, 50ml FBS, 5ml 100xNEAA, 5ml PSG, 7µl β-mercaptoethanol). This -LIF differentiation protocol differentiates the ESCs into all lineages. For a faster more directed differentiation, 1µl of 1mM retinoic acid (RA) can be added per 10ml differentiation medium. This will select for cells in the neuronal lineages, however all other lineages will still be present, albeit at lower

proportions.

Detailed NPC differentiation protocol can be found in appendix C.1 on page 259. Briefly, NPC differentiation was performed by culturing ESCs in the absence of MEF feeders in 50:50 DMEM/F12:Neurobasal medium (Gibco) supplemented with 1x N2 (Gibco), 1x B27 (Gibco), 40mg/L insulin (Sigma), 25μg/ml BSA Fraction V (Gibco) at $0.5x10^6$ - $2.0x10^6$ cells per 10cm plate for 6 days. Cells were then resuspended in N2 expansion medium (DMEM/F12, 50μg/ml BSA fraction V, 10ng/ml EGF (Preprotech), 10ng/ml FGF (Preprotech), 1μg/ml laminin (Invitrogen), 1x N2) and plated onto uncoated T75 flasks to allow for neurosphere outgrowth. Following 4 days, neurospheres were collected by mild centrifugation and plated onto gelatin-coated plates in N2 expansion medium. Following 2-3 passages, cells represent a homogeneous population of NPCs, which can be seeded to expand single-cell derived clones. 0.5x to 1x B27 supplement can be added to medium to aid in single-cell colony expansion or to help rescue a poorly growing culture.

Cell lines can be routinely frozen and stored in liquid nitrogen by resuspending cells in freezing medium (regular medium supplemented with 10% FBS, 10% DMSO), slow cooling at -80°C for 1-3 days, then transferring to liquid nitrogen. Thawing cells should be performed quickly by placing frozen vials in 37°C water bath and transferring immediately to fresh medium to minimise cell death.

Table 8.1: Cell lines used in this work

Name	Cell Type	Species	Genetic Background	Source	Notes
AB2.2	Embryonic Stem Cell	mouse	129/SvEvBrd-Hprt^{b-m2}	A. Mills, CSHL	
Bl6 x CAST	Embryonic Stem Cell	mouse	C57Bl6/J x CAST/EiJ	C. Vakoc	
CSH1	Primary Embryonic Stem Cell	mouse	129 x C57Bl/6	derived by S. Kim, CSHL	
lmna$^{-/-}$	Embryonic Stem Cell	mouse		C. Stewart, A*STAR	LmnA/C knockout (Sullivan et al., 1999)
GM12878	lymphoblast	human	GM12878-16XiMat	Coriell	one of cell lines used in (Gimelbrant et al., 2007)
MEF	Mouse Embryonic Fibroblast	mouse	CF-1	GlobalStem, Rockville, MD, USA	Irradiated for ESC culture
MK6	Primary Embryonic Stem Cell	mouse	C57B4	derived by S. Kim, CSHL	
NE4C	Neural Stem Cell	mouse	C57Bl x Sv129	ATCC CRL-2925	cell line is triploid
nGFP2 iPS	Induced Pluripotent Stem Cell	mouse	ROSA26-rtTA/pgk-puro; Nanog-GFPiresPuro	R. Jaenisch, Whitehead Institute, MIT	contain doxycycline inducible Yamanaka factors and Nanog-GFP reporter (Wernig et al., 2008)
R1	Embryonic Stem Cell	mouse	129X1 x 129S1	S. Kim, CSHL	
v6.5	Embryonic Stem Cell	mouse	C57Bl/6 x 129/Sv	S. Kim, CSHL	
WB6.1	Primary Embryonic Stem Cell	mouse	C57Bl6 Cbrd/Cbrd/Cr	derived by S. Kim, CSHL	
Ts65Dn (5-1c)	Euploid Embryonic Stem Cell	mouse	female, Ts65Dn x C57Bl/6EiJ/C3H/HeSnJ	U. Hochgeschwender, Duke University	euploid control for Ts65Dn (7-5) line (Reeves et al., 1995)
Ts65Dn (7-5)	Trisomic Embryonic Stem Cell	mouse	female, Ts65Dn x C57Bl/6EiJ/C3H/HeSnJ	U. Hochgeschwender, Duke University	trisomic for region between *Mrpl39* and *Znf295* (Reeves et al., 1995)
Ts1Cje (line 2)	Trisomic Embryonic Stem Cell	mouse	male, B6EiC3Sn-Ts(16C-tel)1Cje x C57Bl6/Ei/C3H/HeSnJ	U. Hochgeschwender, Duke University	trisomic for region between Sod1 and Mx1. Sod1 inserted copy contains neomycin resistance cassette (Sago et al., 1998)
Ts1Cje (line 9)	Euploid Embryonic Stem Cell	mouse	male, B6EiC3Sn-Ts(16C-tel)1Cje x C57Bl6/Ei/C3H/HeSn	U. Hochgeschwender, Duke University	euploid control for Ts1Cje (line 2) (Sago et al., 1998)
ZHBTc4	Embryonic Stem Cell	mouse		A. Smith, University of Cambridge	*Pou5f1*$^{+/-}$tetracycline regulated transactivator tTA and tTA-responsive Oct3/4 transgene (Niwa et al., 2000)

8.2 RNA isolation and cDNA synthesis

Total RNA was isolated using Trizol reagent (Ambion). Cells were resuspended in 1ml Trizol reagent and either stored at -80°C or processed immediately. 0.2ml of chloroform was added per 1ml Trizol used. Samples were shaken vigorously, incubated for 2-3 minutes at room temperature then centrifuged at 12,000g for 15 minutes at 4°C. Following centrifugation, the upper aqueous phase containing RNA was transferred to a new tube, and 0.5ml isopropanol added, gently mixed, incubated at room temperature for 10 minutes, then centrifuged at 12,000g for 10 minutes at 4°C. Following centrifugation, the RNA forms a small white pellet which was washed once in 1ml 75% ethanol, air dried for 10 minutes and resuspended in 10-30µl water. RNA concentration was measured using nanodrop, only samples with OD260/280 and OD260/230 ratios above 1.6 were used for subsequent experiments.

cDNA synthesis was performed using 1µg total RNA. RNA was first treated with DNAse I (Invitrogen) to remove any potential contaminating genomic DNA for 15 minutes at 25°C. DNAse I enzyme was then inactivated by adding 1µl EDTA and heating to 70°C for 10 minutes. Following DNAse I treatment, reverse transcriptase reaction was performed using TaqMan RT reagents from Applied Biosystems (#N808-024), using either oligodT or random hexamer primers and a reaction time of 30 minutes at 48°C.

8.3 RNA-sequencing library preparation

Two protocols were utilised to generate libraries from 4-10µg total RNA isolated using Trizol as in section 8.2 for deep sequencing analysis. The first is a 5-day protocol utilised by ENCODE and was used to generate all libraries in the initial RNA-sequencing screen, as well as libraries analysed in the Lamin A/C manuscript (Chapter 3). Subsequent libraries were generated using Illumina TruSeq mRNA library preparation kit, and typically take 2-3 days to generate. Prior to library construction, the quality of RNA was verified by running an Agilent RNA nano ChIP. Only samples with RNA Integrity Number (RIN) > 9.0 were used for library preparation. Detailed ENCODE RNA-seq protocol can be found in the appendix C.2 on page 262. Illumina TruSeq libraries were prepared as per manufacturers instructions. Paired-end libraries prepared with the ENCODE RNA-seq protocol were sequenced on the Illumina genome analyser (GA) IIx

platform. Paired-end libraries generated using the Illumina TruSeq mRNA library preparation kit were pooled and sequenced on the Illumina HiSeq platform.

8.4 Allele-specific RNA-sequencing analysis

8.4.1 C57Bl/6J and CAST/EiJ transcriptome and read mapping

The C57Bl/6J transcriptome was collected from ENSEMBL v59. The transcriptome of Mus castaneus (CAST/EiJ) was created by substituting the 423,197 exonic single nucleotide variants (SNV) between C57BL/6J and CAST/EiJ (Keane et al., 2011). In total 82.8% of transcripts have at least 1 SNV between CAST/EiJ and C57BL/6J. RNA-seq libraries with ~80 million reads per clonal sample were mapped to both the C57BL/6J and the CAST/EiJ transcriptomes using BWA (Li and Durbin, 2010) with default parameters, resulting in ~50 million mapped reads per sample.

8.4.2 Definition of expressed transcripts and genes

Overall, transcript expression levels were determined by calculating the union of raw reads mapped to both the C57BL/6J and the CAST/EiJ transcriptome files and were normalized by library size using DESeq (Anders and Huber, 2010). Expression levels are reported as the number of Normalized Reads Per Kilobase (NRPK), by dividing the normalized read count by the transcript length. In each clone, a transcript was defined as not expressed when its expression level was below a threshold of 5.8 NRPK. This threshold is equal to the average of the expression of the 20^{th} percentile of expression values across all clones. The set of assessable transcripts used as a background for all subsequent analysis was defined as the transcripts that were expressed in at least two clones. The set of assessable genes used as a background was defined as the genes containing at least one transcript in the set of assessable transcripts.

8.4.3 Identification of monoallelically expressed transcripts within each clone

To identify monoallelically expressed transcripts, we exploited the SNVs between C57BL/6J and CAST/EiJ present in each transcript. At each SNV position we first determined the total read count at the respective SNV position in both the C57BL/6J and the CAST/EiJ transcriptomes.

Reads that mapped equally well to both transcriptomes were only counted once. Subsequent analysis was performed on the 72,754 transcripts (82.8% of all transcripts) that contained at least one exonic SNV between C57BL/6J and CAST/EiJ. A read was then assigned to the C57BL/6J (or CAST/EiJ) allele if it contained the nucleotide variant associated with C57BL/6J (or CAST/EiJ) at the particular SNV position. Transcripts in each clone were classified into one of six categories defined below based on both a p-value generated from a binomial test (see below), and a d-score representing the ratio of expression between the two alleles and hence the strength of allelic imbalance (see below):

Monoallelic: |d-score|$\geq$0.40, p-value $\leq 10^{-8}$

Allele-biased: 0.18$\leq$|d-score|$<$0.40, p-value $\leq 10^{-8}$

Biallelic: |d-score|$<$0.18 and/or p-value$>10^{-8}$

Not-expressed: expression is lower than 5.8 normalized reads per kilobase (NRPK)

Non-assessable: No SNV in the transcript with at least 5 reads coverage or the transcript is within a genomic region filtered for aneuploidy

Other: all remaining transcripts P-value calculation

For each transcript, we assessed whether there was evidence of allele specific expression by considering the set of k SNVs between C57BL/6J and CAST/EiJ. For $SNV_i (i = 1, \ldots, k)$ let X_i denote the number of reads mapped to C57BL/6J allele, n_i denote the number of reads mapped across both alleles, and p be the probability that a read carries the C57BL/6J allele. Subsequently, we assume that X_i $B_i(n_i, p)$ where $B_i(N, p)$ denotes a binomial distribution with parameters N and p. For each transcript, we then assess the following hypotheses: $H_0 : p = 1/2$ $H_1 : p \neq 1/2$ by calculating a likelihood-ratio test statistic before computing a p-value to test whether there is evidence to reject the null hypothesis by comparing the likelihood-ratio to the quantiles of a chi-squared distribution with one degree of freedom.

d-score calculation: To calculate a d-score for each transcript, we first d_i as $d_i = \frac{X_i}{n_i} - \frac{1}{2}$ where X_i and n_i are as above. The d-score, d, of a transcript is calculated as the weighted mean of each d_i of the transcript: $d = \sum_{(i=1\ldots k)} w_i d_i$ where $w_i = n_i/N$ with N being the total

number of reads that map over all positions i in the transcript. This gives less weight to SNV positions with fewer mapped reads, since they will have more sampling noise and are thus less representative of the overall d-score of the transcript. The d-score takes values between -0.5 to 0.5, where negative values correspond to a bias toward the CAST/EiJ allele and positive values to the C57BL/6J.

8.4.4 Classification of monoallelically expressed transcripts across clones

Based on the classification of transcripts into the 6 categories described in the previous section, we further assigned transcripts into one of three classes of monoallelically expressed genes. Transcripts assigned to the not-expressed, non-assessable clones and other classes for a given clone were ignored in the classification process.

ClassA: transcripts where at least one clone showed allele specific expression (either monoallelic or allele-bias, defined above) of the BC57Bl/6J allele and a different clone showed allele specific expression of the CAST/EiJ allele. Some clones may also be biallelic. Transcripts in this class represent high confident random monoallelic genes.

ClassB: transcripts with at least one biallelic clone and at least one clone showing allele specific expression (either monoallelic or allele-bias) towards either the C57Bl/6J or CAST/EiJ allele. All clones showing allele-specific expression have the bias in the same direction. Class B transcripts were further filtered using more stringent parameters to select those in which there was strong evidence that the second allele was transcribed in at least one clone. Filtered class B transcripts contain at least one clone with stringent allele specific expression (p-value$<10^{-10}$, |d-score|>0.35) and one clone with stringent biallelic expression (|d-score|<0.10).

ClassC: transcripts in which allele specific expression was observed toward the same allele for all clones and no biallelic clones were observed. This class contains non-random monoallelically expressed genes, including imprinted genes or those biased due to strong cis genetic effects. Class C transcripts were not included in subsequent analysis.

Not all transcripts were necessarily assigned to one of the three classes above. Genes were assigned to a class if at least one corresponding transcript was in the class. Consequently, genes could be present in more than one class. The final set of random monoallelically expressed transcripts includes class A and filtered class B transcripts.

8.4.5 Identification of genes showing transcriptional upregulation

Linear regression analysis was used to compare expression levels across clones with different extents of allelic imbalance. For each gene the model assumes a linear relation between its expression level in a given clone and the degree of allelic imbalance (defined by the |d-score|) in the same clone. In other words, a transcript expressed in a monoallelic clone (|d-score|=0.5) would have half of the expression of the same transcript expressed in a biallelic clone (|d-score|=0). In a given clone, the relation between expression level and d-score is: $y_c = -Exp_{Biall} \times |d|_c + Exp_{Biall}$ where y_c is the expression level of the transcript in clone c, $|d|_c$is the d-score of clone c for this transcript and Exp_{Biall} is the expression of this transcript in the biallelic state. To test whether this relationship holds for monoallelically expressed genes we regressed y_c upon $|d|_c$ and obtained maximum likelihood estimates of the slope a and the intercept b. We then calculated the ratio $r = |\frac{a}{-b}|$ If r is close to 0, then the expression level is similar when the transcript is both biallelically or monoallelically expressed. Genes with a transcript with a ratio $r < 0.35$ were defined as upregulated and investigated further.

8.5 Genomic DNA isolation

Genomic DNA was isolated using Trizol reagent (Ambion). Cells were resuspended in 1ml Trizol reagent and either stored at -80°C or processed immediately. 0.2ml of chloroform was added per 1ml Trizol used. Samples were shaken vigorously, incubated for 2-3 minutes at room temperature then centrifuged at 12,000g for 15 minutes at 4°C. Following centrifugation, the white interphase can be seen containing genomic DNA. The upper aqueous phase containing RNA is completely removed, and 0.3ml 100% ethanol added to the DNA interphase sample. Samples were mixed by inversion, incubated at room temperature for 2-3 minutes, then centrifuged at 2,000g for 5 minutes at 4°C. Following centrifugation, the pellet is washed twice in 0.1M sodium

citrate in 10% ethanol by incubating at room temperature for 30 minutes with periodic mixing followed by centrifugation at 2,000rpm for 5 minutes at 4°C. The pellet is then air dried for 15 minutes at room temperature, and resuspended in 300-600µl 8mM NaOH then centrifuged at 12,000g for 10 minutes at 4°C to remove insoluble material including membranes. The DNA solution is transferred to a new vial, and the high pH is adjusted by adding HEPES such that it is lowered to pH7-8. Genomic DNA is subsequently purified by phenol chloroform extraction. Equal volumes of phenol is added to the sample. Samples are vigorously mixed, centrifuged and the upper phase transferred to a new vial. Next the DNA is purified by one round of phenol:chloroform:isoamylalcohol extraction and one round of chloroform:isoamylalcohol extraction. DNA is precipitated by adding 0.1 volumes of 3M NaOAc and 1.1 volumes of isopropanol, incubating at -80°C for 30-60 minutes, centrifuging at 12,000g for 30 minutes at 4°C, washing once in 75% ethanol, air drying the pellet, and resuspending in water. Sample concentration was measured using the nanodrop.

8.6 Mouse diversity SNP arrays

To control for any variation in DNA copy number that may result in identification of false-positive random monoallelically expressed genes, copy number was assessed using SNP arrays. Mouse Diversity Arrays (Affymetrix) containing probes for over 600,000 SNPs, were run using 5µg purified genomic DNA per sample by The Jackson Laboratory, Bar Harbor, Main, USA. The library MouseDivGeno from JAX lab was used to normalise the signal from raw data in CEL format, keeping only the SNPs present between C57BL/6J and CAST/EiJ mouse strains for analysis. For each SNP we calculated a BAF value, $BAF = \frac{B}{(A+B)}$, where A and B are the signal value for the A and B alleles respectively. Each chromosome was split into bins of 1Mb in size with an average of 85.5 SNPs per bin. Average BAF values from the ESC SNP array were used to determine the background for each chromosome. Bins with a distribution of BAF values that differed from the background were defined as aneuploid (Kolmogorov-Smirnov test, p-value$<10^{-8}$). A total of 16 aneuploid regions were identified across the 6 NPC clones, and a subset were confirmed by karyotype analysis (Molecular Cytogenetics Core Facility, Memorial Sloan Kettering Cancer Center, NY, USA). Transcripts within aneuploid regions were defined

as non-assessable and removed from further analysis.

8.7 SNP-PCR

Exonic SNPs were identified in the Mouse Phenome Database hosted at The Jackson Laboratory (http://www.phenome.jax.org/). Primers spanning the SNP were designed using NCBI's online primer design tool (http://www.ncbi.nlm.nih.gov/) ensuring that SNPS were positioned at least 20 base pairs from both primers (table 8.3 on the following page). SNPs were verified using either genomic DNA or cDNA from non-clonal samples. PCR was performed using Phusion polymerase (Finnzymes) which has proofreading capabilities. Products were either gel purified or reactions were cleaned up using PCR purification spin columns (Qiagen). Samples were sent to Molecular Cloning Laboratory (MCLAB, South San Francisco, CA, USA), or Genscript (Piscataway, NJ, USA) for Sanger sequencing with either forward or reverse primers used for PCR reaction. Sequencing traces were analysed using MacVector software (MacVector Inc.) or 4Peaks (Mekentosj).

Table 8.3: List of primers used for SNP detection from cDNA.

Gene	SNP ID	Forward	Reverse	Product size (bp)
Anxa6	rs26961422	TGGCCGCATACAAAGACGCC	TGCTGTTTGCTGCGGTTTCCCA	394
Bai2	rs27506945	GTCACAGGTGGTGCGGAGCC	GCCCCGCATGGGAAACGTGA	388
Cap2	rs29711297, rs37350673	GTGGCACCGTCCGTGGAAGC	AGGTGGCAGCGTCGTTCATCT	402
Cbr3	chr16:93690862, chr16:93690751	GCCAGACTCGGCTTACGGGG	AGCTGAGAGTGTGGACACAGCG	424
Fam111a	rs30355943	ACCAACACCAGGGACCAAAGACCA	GCTGCAACCTGAAAAAGCTTGCCC	641
Fkbp7	rs27948038	GTTGTTCCGATGCTCAGGGA	TTGCCGGTCATTATCCGTGT	437
Gas6	rs29938009	ACAGATGGCTGGGGAGGCCG	ACGGGGGTCCCGCTGAACAT	418
Klhl22	rs32332429	ACCGGCTAGCAGACCTCTTTGAC	TGCAGCCGCTGAAGGACCTC	307
Mcts2	rs215420528	GTTTCCAGGTATCGAGCCGT	CTGCTGGTGTGGCAGGATAA	193
Npl	rs32582098	CTGAGGAGCTGCTGGACGGGA	CCGGGGTGGGCCCATTGGAA	383
Pla2g7	rs51295335, rs45816585	ACCAGGTGGCTGCAAAAGTGGA	TGCGTCAAAGGGTGACCCAGG	432
Retsat	rs37330726, rs6203726	TCAGCTGGGAGAGCGTGTGGT	TGCATGAGGATGCAGCCGAGC	438
Rgs16	rs32553675	CCGCGCTTCCTCAAGTCACCA	GGGACGGGCGACTCCACCTC	158
Rhoj	rs29197294	TGGAGCGCAGTGCTACTTGGA	CACAGCAGCCGTGACACCCT	134
Scd2	rs36387300	CCGGCCTCCTGCTGATGTGC	ATCCGTAGAGGTGGGCGGCA	159
Sema3c	rs237113365	TGTCCCACCAGCAGTTAGAC	GTGGTCTTTGCTCCCCACAT	80
Serpinh1	rs13459732	CCTTGGGCAGGGAGATGGCG	TGGTGGAGATGCCCCTGGCT	162
Sgsm1	rs33590970	GTGAGGCGGACAAGCCCAGT	AGGCCAGGGCCTCATCATCCAG	375
Tgds	rs49904275	CGCAAGCTGGACGACCGACA	TCATGGATCAGGGCTTCAAAGGAGA	142
Tubb2a	rs36722013	ACAACCAGCACCATGCGCGA	GCTTTCAGACTCCTTCCTCACCACA	389

8.8 Quantitative RT-PCR

Quantitative RT-PCR was performed using 2µl of cDNA (see section 8.2 on page 175) using primers amplifying a maximum of 300bp (table 8.8 on page 193) using SYBR green reagents (Applied Biosciences). At least 3 biological replicates for each sample were used in each experiment, and values normalised to the geometric mean of at least 3 separate housekeeping genes. Data was analysed using Excel (Microsoft) and graphing performed using Prism (GraphPad).

Table 8.4: Primers used for quantitative RT-PCR for mouse genes.

Gene	Forward	Reverse
18S	GGGCCCGAAGCGTTTACTTT	CGCCGGTCCAAGAATTTCAC
Acan	GCACAGCGGGCTTGCCTACA	CAGCCAGCCTGCATCGCACT
Actin-beta	GCAGCTCCTTCGTTGCCGGT	TACAGCCCGGGGAGCATCGT
Acyp2	CTGGTTGGCTGGGTGAAGAA	AACCTTTTCTTCGGGGCCTT
Airn	CTCATTGGGAGCCTGTGATT	TCTGTGCAGCTCATTTCACC
Acot1	CATCACCTTTGGAGGGGAGC	TGTACCTTTCCCCAACCTCC
Adamts1	CCGTGCCCGAGGAGGAGTCA	AGGCCACTGGTTCCGGGTGT
Anxa6	CCGCCGTGACCTGACCAAGC	GTCCAACCCCACCCCTTGCG
Anxa6 (intron)	ACCACTGCCCCTTTAGCATC	ATTCACCAGCATCCACCCTG
App	GGTGGCCCACGCAGGATCAC	CGGCGTTGCCATCAGTGGGT
Blbp	TCATGTGCAGAAGTGGGATGGCA	AGCTGGCTAACTCTGGGACTCCA
Brachyury (T)	TGGCGGGAGGATGAGCTCGC	CAGTTCGCGTTCGGTGGGGT
Cap2	TGGCTCCAAGCTGGTCACCG	CGGGCGTCGGGGCTATGGTA
Cap2 (intron)	TCCCTCGTGATCCCCATTCT	TGACTAACTGCCGGGGTAGA
Capn5	ACACGTCAGAGGAATGGCAG	GGATGCTCAGGTAGGACGTG
Cbr3	GTCCCTCTGACATGTCGTCC	CGTTAAGTCCCCCGTACTCC
Cbr3 (intron)	CCTTGGAACTTTCAGCTGCG	TGATGAGCTTGACAAGGCGT
Cdx2	GCGGCTGCTACGGCGAACTT	GCTGTAGCCCATAGCGGCGG
Chaf1b	GCCAAATCCACAGCCCAATAGC	CCATCACCAGCCTTCGAGCTTTG
Cpne8	ACCCTCACCCCTTCCCGTCG	TCCACTCGTGTGGCGGGGAT
Cth	CTTTGCATCGGGTCTTGCTG	CAGATGCCACCCTCCTGAAG
CycloB1	GACAGACAGCCGGGACAAGC	GGGGATTGACAGGACCCACA
Dscam	AGCATGGAAGGCTCACAGGCA	TGAGGGACGGATGGGCAAGTGA
Dio3	GCCTCGTGCCTCGTGCTGTT	ATTGAGCACCAACGGGCGGG
Dlk1	AGCTGGCGGTCAATATCATC	AGCTCTAAGGAACCCCGGTA
Ebf1	ATCCAGCCCCACCATGGCCT	TCGCTTGCAGGCTGTTCCCG
Emerin	AACACCGCTTCCAGGTGCCG	GTCCACGGCGGCTGAATCCAA
Eomes	GGGAGGCCGTGGCGCTTATC	ACCCCGGAGTCGCTGGAGTC
Fgf5	GCTGTGTCTCAGGGGATTGTAGGA	GCCACGTACCACTCTCGGCC
Fkbp7	ACTTGGCTAAAGACGGCTCC	TGGAATCTTGCCTTCTGCGT
Foxd3	GGTGCGGCGTGGAGAAACCT	GCTCCGCGAACTCTTGCGGT
Gabp1	GGGGCGTTTGTGCCTCGTGA	CCCTTCCCGTGGAATCGGCG
Gapdh1	GGTGGTGAAGCAGGCATCTG	CGGCATCGAAGGTGGAAGAG
Gas6	CGTGAGGCGGCGCAGTTTCT	CCTCCTCCACGCACTCCCGT
Gas6 (intron)	GTCACCCACACCATGTTTGC	GCCAGGAATGAAGCTGACCT
Gata4	TGACAGTTCCGCACACCCGC	GGCCACCTGCTTCGTAGGCG
Gfap	AGGGCCAAAGCCTCAAGGAGGA	TGGTGCTTTTGCCCCCTCGGATC
Gtl2	AGCCTCGCGGTCTCTTCGGT	TGTGTCCAGGGTCCCACGCA
Gpc5	GCCGGATGCTTGCCGTGTCT	CCGCGGCTTCACCTGCAAGA
Gpn3	ACTACAGCATGGTCCGCTTC	CCCGTACTGAATGGCGAAGT
H19	GGCAGCAGTGGGCAGGTGAG	GTCGCCAATGGTCCCCCAGC
Igf2bp1	CCCGCTCCTAGAGCCCATGTGA	CACCCCAAGTCTTTCCTAAGCGC
Jam2	CGGGCGAAGTGCTGGGAGAC	TTGGGGGCTCCTCGCCATCT
Kcnq5	AGCGGCTCTCGAGGCAGTCA	GAAAGCAGCCTCCCCCGCAG

Table 8.4 *continued from page 184.*

Gene	Forward	Reverse
Lamin A	GCCTTCGCACCGCTCTCATCA	TGCGTGAGCGCAGGTTGTACT
Lamin B1	AGCGCGCCAAGCTCCAGATC	GTAGCCAGCGCCGCATCCTT
Lamin A/C	GGCTGTGGGAGCAGCCTTCG	GGCTGCCACTCACACGGTGG
Lamin B receptor	GTGTCAAGTGGCTGTGCCCGT	CTGTGGCACTGTCCGCCCTT
Me1	TGCCGCCCTGCATCATCAGC	TCAGGGCCTCTGACTCGCCG
Mrpl39	TGTGGACGGACAACCGTGGGA	TCCGGTGCTCTGACCAGGCT
Myo1e	CTCTATGGCCCGCCCACCCT	GGCCGGCTGCTCGTTTGTCT
Nanog	CCTCGCCATCACACTGACATGAGT	GCAGGTCTTCAGAGGAAGGGCG
Necdin	CTGGTGCAGAAGGCGCACGA	GCTCCTCTGGGCTGAGGGCT
Nespas	CGGCGTCGGCAGGAGCTATG	ACAGCCCTGGGGGAGCAACT
Nestin	CATGAGGGCAGCCATGCCCC	CAGTCCCAGCTTCCGGCTGC
Nurim	TGTACCGCGGCTGTGCTTCC	GGGAGCCCTGAGTCAACCCGA
Oct4	GACCGCCCCAATGCCGTGAA	TGGTCTGGCTGAACACCTTTCCA
Pabpc1	GCCGGGCTGAACTTCTCGTA	CGCACTCGCTCTCCTCCTCT
Pcsk6	ACAGGAAGTGTGTGAGCGAG	GATAGAACCCACGGCGACAA
P4ha3	ATGCTACGTCGCCCAGCAGC	ACCATCCCCCTCTCCGCTCC
Pla2g7	CAACCGGAGGAGGGGAGGGTC	GAGCGCAGCAATGCCACCAGA
Ptgr1	CCCTACGGACGGTAACTTCG	CTGTCGCCCTCCTTCAGTTT
Ptk2	TCGCTGGTGCTCCTGAGCCT	TGGCAACGCCCGTTCACCTT
Rcan	CGACACTCGGTGAACATTTG	GGTTTGGAGAACATCAACC
Rgs16	CACGAGACCCGAGAACTGAC	GACTTGAGGAAGCGCGGATA
Rhoj	GGCCACTCTCTTACCCCAAC	GAGGCATGCAGTCCTTCAGT
Serpinh1	CAGTGGAGTTGCTGAGGGAG	AGCTAAGTTCCAAGGCGACC
Snrpn	GGCAATGCAGCAACCAAGCTGT	CTGGACCTCGACCTCGGGCA
Sox1	CAGTGGGAGTGCCAACGCGA	CAAGGCAGAGAAGAGAGTGTTTGCA
Sox2	CAAGGCAGAGAAGAGAGTGTTTGCA	GCCGCCGCGATTGTTGTGAT
Sox7	ACTGGAGTGTCCCGCCCTGG	GCGTTGTGCAGGTCCGGGTT
Sox17	TGCATTCTGGACCCGCTACTGTT	GCCTCTCAGAAGCAGGGACGC
Thrsp	GTGCAGCTGAGTGGGCCTGG	AGCGTCGTTCTCAGCCTCGC
Ttc4	GACTGAAGCGCCGAAGTTG	CCATGGAAGCGTCCTCTGTA
Twist	GCAGAGCGACGAGCTGGACT	GGTGGGGAGCTCCGCTGCTA
Xrcc3	CTTGGCATCACCTGGGCTAA	CCACTGACTGTGTAGCAGCA

8.9 Bisulfite analysis

Bisulfite conversion was performed on 500ng of purified genomic DNA using reagents from the EZ DNA Methylation-Gold Kit (Zymo Research, D5005). The bisulfite conversion reaction was performed at 64°C for 2.5 hours, samples were subsequently purified using spin columns with an addition desulphonation step for 20 minutes at room temperature between washes. Samples were eluted from the columns in 10µl H_2O and stored at -20°C. PCR amplification of bisulfite treated DNA was performed using OneTaq Hot Start DNA polymerase (NEB, M0481). Using a hot start polymerase is crucial to amplify bisulfite treated DNA. Primers were designed to amplify regions within 500bp of the transcription start site using MethPrimer (Li and Dahiya, 2002) online tool (http://www.urogene.org/methprimer/). Where possible, primers were designed to include an A-T SNP between the C57Bl/6J and CAST/EiJ mouse strains. Only A-T SNPs were possible due to loss of information occurring at cytosines during bisulfite conversion, and consequently guanines during the PCR amplification steps. Primer sequences are listed in table 8.5. PCR was performed using an initial denaturation at 95°C for 5 minutes, followed by 45 cycles of 95°C for 30 seconds, 52°C for 30 seconds and extension between 68°C and 72°C for 60 seconds, then a final extension phase at 72°C for 60 seconds. 4µl of gel-purified PCR products were cloned into pCR4-TOPO-TA vector using the TOPO-TA cloning kit (Life Technologies) and transformed into competent Mach1 *E. coli* (Life Technologies). Single colonies were picked and expanded in LB overnight. DNA purification was either performed using the Qiagen miniprep kit reagents, or rolling circle amplification was performed directly on bacterial cultures prior to Sanger sequencing by Molecular Cloning Laboratory (MCLAB, South San Francisco, CA, USA). Sanger sequencing was performed using the M13 F (-20) universal primer and sequencing traces analysed manually for the SNP allele information, and using the BiQ_Analyzer software for bisulfite analysis (Bock et al., 2005).

Table 8.5: Primer sequences used for bisulfite analysis.

Gene	SNP	Forward	Reverse	Product Size (bp)	CpGs
Anxa6	n.a.	GTTAGGGAGGTTTTAGATTAGGG	CAAATAAAAAATAACAAATTTCCCC	258	20
Cacng5	n.a.	TTGGTTATTTAGTTAAGGAGGTTAG	CTTCTTCTAATCCAATATTCCAAC	283	24
Cbr3	rs4220182	GAGGGTGTTGGGTAGTGAGTTT	ACTAAAAATTCCAAAAAAAACCTCC	201	20
Cpped1	n.a.	GAAGAGTTAAGTAGTATTTGAGGAAAA	AAACAATACTAACAACACCCAAAAC	237	16
Fkbp7	n.a.	TAGTTTTTTGTTAGGTTTTTAAAGG	TATCTAAACTACACCACCAAACCAA	282	14
Gas6	chr8:13494073	TTTTTTTTGTTGTATATTTTTTTTA	ACATCCCAAACACACAAACATTAC	233	15
Npl	n.a.	TATTTAGGTTGTTGTGTATTGGAGA	AACTATAACCACACCCAAAAAAAA	196	15
Rgs16	n.a.	GGGGATAAAAGTTTTAGTAGTTAT	ATCCTCTTACCTCTCCAAACAAATAT	203	15
Rhoj	n.a.	GTGTGATTTTGGAAGATTTTTTTT	TCCTACTTTATCCCCTACTAACTAACC	289	7
Serpinh1	n.a.	GATTTTTTTTAAATTATTTTGGTTTT	ACCTATCTACCTTCAATTCACTCTC	210	8

8.10 Western blotting

Protein extracts were made by resuspending cell pellets in 1x protein sample buffer, 10% β-mercaptoethanol (100ml of 3x protein sample buffer contains 15ml 1M Tris-HCl pH6.8, 45g sucrose, 6ml 0.1M EDTA pH7.4, 45ml 20% SDS, 0.1g Bromophenol Blue), boiling for 10 minutes at 100°C and filtering through a needle and syringe. 10μl of protein extract was loaded onto 8-15% SDS-PAGE gels and run for 2 hours at 90V until the gel front reached the end of the gel. Transfer to nitrocellulose membranes was performed overnight at 4°C at 15V. Membranes were blocked in 5% skim milk, 0.1% Triton X-100, 1xPBS for at least 3 hours at room temperature with gentle rocking. Primary antibodies were incubated with the membranes in 5% skim milk powder, 0.1% Triton X-100, 1xPBS for 2-4 hours at room temperature at concentrations listed in table 8.6. Membranes were washed at least 3 times for 10 minutes each in 0.1% Triton X-100 prior and subsequent to secondary antibody incubation. Secondary antibodies conjugated to HRP were diluted 1:10,000 in 5% skim milk, 0.1% Triton X-100, 1xPBS and incubated with membrane for 30-60 minutes at room temperature with gentle rocking. Membranes were subject to ECL (Perkin-Elmer) exposure for 1 minute the immediately exposed to film.

Table 8.6: Antibodies used for Western Blotting

Antigen	Source	Species	Working Dilution	Size (kDa)
Anxa6	abcam ab114811	rabbit	1:400	76
Cbr3	Santa-Cruz sc-70218	rabbit	1:30	31
Emerin	Leica Microsystems NCL-Emerin	mouse	1:1,000	34
Histone H3	abcam ab1791	rabbit	1:10,000	15
Lamin A (323-11)	R. Goldman, Northwestern University	rabbit	1:2,000	75
Lamin A/C (266)	R. Goldman, Northwestern University	rabbit	1:2,000	72,75
Lamin A/C	Active Motif 39287	mouse	1:1,000	72,75
Lamin A/C (N18)	Santa-Cruz sc6215	goat	1:50	72,75
Lamin B1	abcam ab16048	rabbit	1:2,000	75
NFKB p105/p50	abcam 133264	rabbit	1:5,000	50, 105
Oct3/4	Santa-Cruz sc9081	rabbit	1:2,000	45
Rgs16	ab119424	mouse	1:100	23
YY1	active motif 39071	rabbit	1:1,000	63

8.11 Chromatin immunoprecipitation

Cells were fixed in 1% molecular biology grade formaldehyde (Calbiochem) at 1x10^6 cells/ml for 5 minutes before being quenched with 2M glycine. Cells were lysed on ice for 10 minutes (10mM Tris-HCl, 10mM NaCl, 0.5% NP-40, pH8.0) in the presence of protease inhibitors (cOmplete Mini cocktail tablets, Roche and 1mM PMSF), and nuclei snap frozen and stored at -80°C. For histone ChIP, nuclei were sonicated using Bioruptor (Diagenode) for 20 cycles of 30 seconds each in dilution buffer (50mM Tris-HCl, 2mM EDTA, 0.2% SDS, 134mM NaCl, 0.88% Triton X-100, 0.088% Na-deoxycholate, pH8.0) in an ice slurry with water exchanged every 4 cycles. Chromatin size was confirmed by agarose gel electrophoresis for each preparation. 2mg antibody (table 8.7 on the next page) were pre-incubated with 40µl Dynabeads anti-mouse or anti-rabbit magnetic beads (Invitrogen) per reaction for 4 hours at 4 degrees centigrade. Chromatin was incubated with bead-antibody mix for 16-20 hours at 4 degrees centigrade. Bound chromatin was eluted with 10mM Tris-HCl, 1mM EDTA, 1% SDS, pH8.0 at 65°C. Reverse crosslinking was performed for 16 hours at 65°C in 10mM Tris-HCl, 5mM EDTA, 0.5% SDS, 300mM NaCl, pH8.0. Following treatment with 50µg/ml RNase A (Invitrogen), and 250µg/ml proteinase K (Viagen), DNA was purified using Qiaquick PCR purification kit (Qiagen). For full protocol see appendix on page 271.

For RNA polymerase II ChIP, frozen nuclei were resuspended in sonication buffer (50mM Tris-HCl, pH7.5, 140mM NaCl, 1mM EDTA, 1mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS) for 16 cycles of 30 seconds each using Bioruptor (Diagenode). Antibody was bound to 40µl anti-mouse magnetic beads (Invitrogen) at 4°C for 4 hours. Chromatin was incubated with bead-antibody mix for 16-20 hours at 4°C. Chromatin bound to the beads was washed three times in sonication buffer, once in sonication buffer containing 500mM NaCl, once in LiCl wash (20mM Tris pH8.0, 1mM EDTA, 250mM LiCl, 0.5% NP40, 0.5% Na-deoxycholate) and once in TE. Bound chromatin was eluted with 50mM Tris-HCl, pH8.0, 10mM EDTA, 1% SDS. Reverse crosslinking was performed for 16 hours at 65 degrees centigrade in 10mM Tris-HCl, 5mM EDTA, 0.5% SDS, 300mM NaCl, pH8.0. Following treatment with 50µg/ml RNase A (Invitrogen), and 250µg/ml proteinase K (Viagen), DNA was purified using Qiaquick PCR purification kit (Qiagen). For full protocol see appendix on page 276.

Table 8.7: Antibodies used for chromatin immunoprecipitation.

Antigen	Source	Species
H3K4me2	Millipore 07-030	rabbit
H3K4me3	Abcam ab8580	rabbit
H3K9me3	Abcam ab8898	rabbit
H4K16ac	Millipore 07-329	rabbit
panH4ac	Millipore 06-866	rabbit
H3K27me3	Active Motif 39155	rabbit
H4K20me1	Abcam ab9051	rabbit
H4K20me3	Millipore 07-463	rabbit
H3K36me3	Abcam ab9050	rabbit
RNApII (8WG16)	Abcam ab817	mouse

Quantitative PCR was performed using 2µl ChIP DNA per sample with 2 technical replicates per sample. Standard curves were performed using serial 1:10 dilutions of input DNA. Primers were designed to contain an informative SNP and were located within 1kb of the transcription start site for promotors, or within exonic sequences, and are listed in table 8.8 on the following page. Analysis was performed in Excel (Microsoft) and graphs generated using Prism6 (GraphPad).

ChIP-seq libraries were performed in a similar manner to ChIP-PCR with a few changes. Cells were fixed and nuclei lysed as above. Sonication was performed for 24 cycles of 30 seconds each on highest setting. 10^7 cells were used per IP, with 4mg antibody pre-bound to 80µl beads. Washes and elution were performed as above. Purified immunoprecipitated DNA was run on a high sensitive DNA chip to verify pulldown and sufficient DNA within 200-600 base pairs. At least 10ng of DNA in the correct size range is required to proceed with library preparation. Libraries were generated using reagents from the Illumina mRNA TruSeq preparation kit, starting at the end-repair step, followed by A-tailing and adapter ligation, as per kit protocol. Prior to PCR amplification, samples were run on a 2% agarose gel for 2 hours at 90V and DNA at 200bp gel purified. PCR was performed for 15 cycles, amplified libraries were purified, verified using the bioanalyser, pooled and sent for single-end 36 base pair sequencing using Illumina HiSeq platform. At least 20 million reads per sample is required for proper analysis. 3 biological replicates per sample were generated, as well as input controls. For full detailed protocol see appendix C.5 on page 281.

Table 8.8: Primers used for ChIP-qPCR.

Gene	Region	SNP	Forward	Reverse
Anxa6	Promoter	chr11:54847121	CCACCCCGAGTGCTTCTTTT	TCTATTTCTGGGTGGCAGCG
Cap2	Gene Body	rs37375663	GGCATCCACAGCGTCAGCGT	GGGGAGGGCTCCTCCTTTTTATCC
Cbr3	Promoter	rs49806072	ATTGCGTGCCCACACTTAGA	ACGTGGGAGAGAAGGCTCAT
Cbr3	Gene body	rs245732913	TGTGACACACTTACCGAGGTG	GTAAGGACTGTCACCCCCAG
Gapdh1	Promoter	chr6:125115595	CCAGCATCCCTAGACCCGTA	CGCCCTTGAGCTAGGACTG
Nanog	Promoter	n.a.	ACAATGTCCATGGTGGACCC	AGCCTTCCCACAGAAAGAGC
Npl	Gene Body	rs48544131	ACTGAATCTTGGTGGGCCCTTGACT	ACCATCCACTGCAAGCCACCC
Rhoj	Promoter	rs32388653	GCCACACTATGTGCCTCTCTCTCT	AGACTCCCCAAAACAACCCCAATGA
Rhoj	Gene Body	rs29149379	TGCTGTCAACAATCCAGTCCCCA	TGGACTCATGTTGATTGGTGGAGCC
Ror2	Gene Body	rs29515585	TGAAATGACGGAGCGAATGC	CGCAAGGAGTTTGATGGTCAC
Serpinh1	Promoter	chr7:106501835	CAAACCACCCTGGCTCCCGC	TGCCCTGGCTCCCAAGGTCT
Serpinh1	Gene Body	rs31820242	ACTAGGCCACCTATTGAAGTCCCCC	TGGCTACAGCCTACTTCAGGACACC
Tubb2a	Promoter	rs38475982	ACACTCTGGTTCTGCTCCCACACT	GAGCGAGGCCACACCTGTCC
Active control*	chr10:79073450-79074050	n.a.	ACAGTAGCTCAGGCCTTCCC	GATACTGAGCCCCTTCCCTC
Inactive control*	chr14:78504700-78505300	n.a.	GCGCGTGGTCGTTTACAT	GTGCTGGCCTTCGATGTT

*Control primers are obtained from Mikkelsen et al. 2007 *Nature*.

8.12 Electroporations and transfections

Overexpression experiments performed in ESCs were conducted by electroporation using 25μg of phenol-chloroform purified plasmid DNA. For stable cell lines, plasmid DNA must be linearised to allow genomic integration. ESCs were fed 2-4 hours prior to harvesting, resuspended at 1.1 x10^7 cells/ml, and 0.9ml cells used per electroporation. Cells and DNA were mixed together, placed into electroporation cuvettes and electroporated using Biorad GenePulser at 250V, 500μF. Time constants should range between 5.6 and 8.0. Following electroporation, cells were plated onto drug-resistant MEF feeder cells and if necessary, appropriate drug selection commenced after 24 hours.

The best method for transfecting NPCs is through calcium phosphate transfections, although Fugene 6 transfection reagent (Roche) also works. Optimal results are made with cesium chloride purified DNA, although endo-free purification (Qiagen) can also be performed. For one well of a 6 well plate at 60-70% confluency, 4μg plasmid DNA is mixed with 12.4μl of 2M $CaCl_2$ (reagents from CalPhos Mammalian Transfection Kit, Clontech) and made up to 100μl with sterile H_2O. To this mix, 100μl of 2x HBS is slowly added while gently vortexing. The combined mix is incubated at room temperature for 30 minutes, then added dropwise to the cells. Cells are incubated for 2-3 hours, then washed 1x for 15 minutes at 37°C in 5% $CO_2$with acidified medium pre-incubated at 10% CO_2 for at least 15 minutes. Medium is then replaced and cells returned to the incubator and assessed the following day.

8.13 Immunofluorescence staining

Immunofluorescence of adherent cells was performed by growing cells on glass coverslips which were then fixed for 15-20 minutes at room temperature in 4% freshly prepared formaldehyde in 1x PBS. Cells were subsequently permeabilised in ice-cold 0.1% Triton X-100 in 1x PBS for 5 minutes on ice, and washed at least 3 times in 1xPBS. Coverslips could be stored at 4°C for up to 2 days in 1xPBS, however best results are obtained with freshly prepared cells. Cells were blocked in 3% BSA for 1 hour at room temperature, then incubated with primary antibodies (table 8.9 on page 196) for either 1 hour at room temperature or overnight at 4°C in 1% BSA, 1x PBS. Cells were washed 3 times in 1xPBS, then incubated with appropriate secondary antibodies

conjugated to Alexa-488, Alexa-594 or Alexa-647 (Invitrogen) diluted 1:1,000 in 1% BSA, 1x PBS for 30-60 minutes at room temperature. Cells were washed again 3 times in 1x PBS, counterstained with DAPI (1:10,000) and mounted in antifade containing 10% glycerol and 1 mg/ml p-Phenylenedi- amine (Sigma).

Blastocysts were fixed in freshly prepared 4% formaldehyde for 30 min at room-temperature, permeabilised in 0.25% Triton X-100 for 15 min at room-temperature, then blocked in 10% FBS with 0.1% Triton X-100 for 1–3 h at room-temperature. Cells were incubated with primary antibodies containing 1% BSA overnight at 4°C. Secondary incubations and DAPI counterstaining was performed as per adherent cells.

Table 8.9: Antibodies used for immunofluorescence staining.

Antigen	Source	Species	Working Dilution	Localisation
Cdx2	BioGenex AM392	mouse	1:200	nuclear
Emerin	Leica Microsystems NCL-Emerin	mouse	1:50	nuclear membrane
Lamin A (323-11)	R. Goldman, Northwestern University	rabbit	1:200, overnight 4^{o}C	nuclear membrane
Lamin A/C (266)	R. Goldman, Northwestern University	rabbit	1:200, overnight 4^{o}C	nuclear membrane
Lamin A/C	Active Motif 39287	mouse	1:200, overnight 4^{o}C	nuclear membrane
Lamin B1	abcam ab16048	rabbit	1:400-1:500	nuclear membrane
Nanog	eBioscience 14-5761	rat	1:200 1hr or 1:400 overnight 4^{o}C	nuclear
Nestin	G. Enikolopov, CSHL	chicken	1:2,000	intermediate filament
Oct3/4	Santa Cruz sc9081	rabbit	1:500	nuclear
SSEA-1	Chemicon MAB4301	mouse IgM	1:100	cell surface

8.14 RNA-DNA Fluorescence *In Situ* Hybridisation

Probes for both RNA and DNA FISH were generated from BAC or fosmid DNA (see Appendix C.6 on page 287 for detailed protocol) by Nick Translation (Abbott Molecular Inc.) with red or green fluorescently conjugated d-UTP nucleotides (Enzo Life Sciences) for 10 hours at 15 degrees. Probe size was verified by agarose gel electrophoresis to be 50-400nt. Probes were mixed with competitor DNA, lyophilized and resuspended in 50% deionized formamide, 2x SSC and 10% dextran sulfate. Cells were grown on coverslips, fixed in freshly prepared 4% formaldehyde for 15 minutes and permeabilized in 0.1% Triton for 5 minutes on ice in the presence of 5mM vanadyl ribonucleoside complex (NEB). Full protocol for RNA-DNA FISH is present in Appendix C.7 on page 289.

Briefly, RNA FISH was performed by hybridizing prepared denatured probes on coverslips overnight at 40 degrees centigrade. DNA FISH required prior treatment with 0.1mg/ml RNase A (Invitrogen) for 1 hour at 37 degrees centigrade, followed by heat denaturation in 70% formamide, 2x SSC for 5 minutes at 80 degrees centigrade, and hybridization with denatured probe overnight at 37 degrees centigrade. Following hybridization, cells were washed in 2xSSC/50% formamide, 2xSSX then 1xSSX at 37 degrees, counterstained with DAPI and mounted in antifade containing 10% glycerol and 1mg/ml p-Phenylenediamine (Sigma). RNA-DNA FISH was performed sequentially with separate images taken for both RNA-FISH and DNA-FISH.

Table 8.10: BAC and fosmids used for FISH probes.

Target	ID	Chr	Start	Stop	Size (kb)	Gene specific	Entire gene
Abcg1	RP24-108E21	17	31164093	31311286	147	no	yes
Acan	RP24-308C23	7	86129664	86297666	168	no	
Acot1	WI1-2795P12	12	85334811	85372727	38	yes	yes
Acyp2	RP23-405O2	11	30375678	30562420	187	yes	yes
Adarb1	CH29-574B03	10	76740613	76906667	166		yes
Anxa6	WI1-0992J09	11	54797616	54836918	39	yes	no
App	RP23-424A20	16	84924341	85121213	197		no
Bag3	WI1-0675I05	7	135665920	135703160	37	yes	yes
Cap2	RP23-105K14	13	46579015	46764554	186	yes	yes
Cbr3	WI1-1329N21	16	93671074	93708521	37	yes	yes
Chaf1b	RP23-290C6	16	93880797	94040198	159	no	yes
Col18a1	WI1-0204H24	10	76532609	76578562	46	yes	no
Col4a1	WI1-1996G17	8	11235581	11272563	37	yes	no
Cpne8	RP24-180J16	15	90333195	90490125	157		no
Cpped1	WI1-2354D17	16	11868258	11909617	41	yes	no
Crip1/2	RP23-313C4	12	114259856	114431973	170	no	
Dscam	RP23-46N2	16	96812588	97069668	257	no	no
En2	WI1-2728F04	5	28477913	28517563	40	yes	yes
Enf1	RP24-140K14	11	44486728	44728495	242		no
Ets2	RP24-372P3	16	95909494	96049560	140	yes	yes
Evc	WI1-2255C20	5	37688862	37727984	39	yes	no
Fam111a	WI1-0083C22	19	12632378	12670899	39	yes	no
Farp1	RP23-91P8	14	121441409	121675164	234	yes	no
Gas2	WI1-663M1	7	59134904	59175563	41	yes	no
Gas6	RP24-203O11	8	13381191	13539259	158	no	yes
Gas6	WI1-0153N19	8	13474615	13517192	43	yes	no
Gpc5	RP24-144L17	14	115509075	115741582	233		no
Gucy1b3	WI1-1386D05	3	81838650	81879656	41	yes	no
HoxD cluster	RP24-398B4	2	74484615	74609972	126	no	
Igf2bp1	RP23-383D10	11	95758900	95942277	186	no	
Kcne1	RP23-204P24	16	92345613	92544377	199	no	yes
Kcnq5	RP23-209O20	1	21444460	21719510	275		no
Me1	RP23-304B9	9	86473026	86670829	198	no	
Mrpl39	RP23-201L12	16	84602500	84758422	156	yes	yes
Myh9	RP24-127K7	15	77590338	77749554	159	no	yes
Myo1e	RP24-400C15	9	70078194	70202876	125		no
Npl	RP24-146M18	1	155336674	155494008	157	no	yes
Npl	WI1-2769H07	1	155350845	155392555	42	yes	no
Oca2	RP23-290P10	7	63521994	63692267	170	yes	
P4ha3	RP24-343C20	7	107360776	107518336	158	no	yes
Pdzrn4	RP23-322L18	15	92356026	92579506	223	yes	no
Pla2g6	WI1-2275J10	15	79130927	79169069	38	yes	no
Rfx4	RP24-342J16	10	84201615	84357496	156	yes	no
Rgs16	RP23-91E22	1	155546928	155679601	133	no	yes

Table 8.10, *continued from page 198.*

Target	ID	Chr	Start	Stop	Size (kb)	Gene specific	Entire gene
Rhoj	RP23-67B14	12	76390736	76504993	114	yes	yes
Ror2	RP23-280O5	13	53207914	53383775	176	yes	yes
Serpinh1	RP23-86E4	7	106411402	106565864	154	no	yes
Serpinh1	WI1-1903D04	7	106486433	106524944	39	yes	yes
Skap2	RP23-261L24	6	51765049	5196949	204	yes	
Slc25a30	WI1-2250C21	14	76162005	76200862	39	yes	yes
Stmn2	RP23-245F10	3	8432176	8623501	191	no	
Thrsp	RP23-63G13	7	104542422	104760155	218	no	yes
Trim12	RP23-325P21	7	111427313	111495378	68	no	yes
Tubb2a	RP24-307M23	13	34048502	34192157	144	no	yes
Twist1	WI1-1685D03	12	34619710	34657218	38	yes	yes

8.15 Microscopy

Immunofluorescence imaging of single sections for images in Chapter 3 was performed using a Zeiss LSM710 laser scanning confocal micro- scope using a 63 × 1.4N.A. oil-immersion objective using 405 nm, 488 nm, 594 nm and 647 nm lasers. Images represent single 0.1 μm sections through the center of ESC colonies or individual blastocysts. No post- acquisition image processing was performed. Figures show representative single z-sections.

All other fixed cell microscopy was performed using an Applied Precision DeltaVision Core wide-field fluorescence microscope system (GE Healthcare, Issaquah, WA) equipped with a PlanApo 60x 1.40 numerical aperture objective lens (Olympus America). Image stacks were taken at 0.2nm intervals throughout the entire cell and deconvolved using Applied Precision SoftWoRx software version 4.2.1. with default parameters. Separate projections of RNA-FISH/DAPI and DNA-FISH/DAPI images were overlayed in Photoshop using heterochromatin foci as a guide. Image analysis was performed manually using Applied Precision SoftWoRx software.

Live-cell imaging was performed using an Applied Precision DeltaVision Core wide-field fluorescence microscope system equipped with PlanApo 60x 1.40 numerical aperture objective lens (Olympus America). For live cell imaging, cells were seeded onto MatTek glass-bottomed microwells. During imaging cells were cultured in DMEM medium without phenol red (Invitrogen), supplemented with 30% FBS (Clontech) and 1000U/ml LIF (Millipore) in an environmental chamber equilibrated at 37°C and 5% CO2. Cells were monitored using wide-field images to verify cell viability. Image stacks (25 slices x 0.8μm, 2x2 binning) were taken at 5 minute intervals over 4 hours. Ultraviolet wavelengths were filtered out using a sharp cutoff (long-pass) filer (˜420nm) to minimize phototoxicity. Blastocyst images were deconvolved using Applied Precision SoftWoRx software version 4.2.1. with default parameters.

Super-resolution microscopy was performed using an Applied Precision OMX (Optical Microscope eXperimental) system equipped with 405, 457, 488, 514 and 593 nm solid-state lasers, a UPlanS Apochromat 100x/1.4 NA objective lens (Olympus) and 4 EMCCD cameras (Cascade II 512, Photometrics). Images were collected in three-dimensional structured illumination mode and reconstructed using SoftWoRx 4.5.0 software (Applied Precision).

Bibliography

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., and Bazett-Jones, D. P. (2010). Global Chromatin Architecture Reflects Pluripotency and Lineage Commitment in the Early Mouse Embryo. *PLoS ONE*, 5(5):e10531.

Allfrey, V. G. and Mirsky, A. E. (1964). Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. *Science*, 144(3618):559.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106.

Arias, A. M., Nichols, J., and Schröter, C. (2013). A molecular basis for developmental plasticity in early mammalian embryos. *Development*, 140(17):3499–3510.

Aseem, O., Barth, J. L., Klatt, S. C., Smith, B. T., and Argraves, W. S. (2013). Cubilin expression is monoallelic and epigenetically augmented via PPARs. *BMC Genomics*, 14:405–423.

Auclair, G. and Weber, M. (2012). Mechanisms of DNA methylation and demethylation in mammals. *Biochimie*, 94(11):2202–2211.

Ayyanathan, K., Lechner, M. S., Bell, P., Maul, G. G., Schultz, D. C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F. J. (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes & Development*, 17(15):1855–1869.

Bacher, C. P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., and Heard,

E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nature Publishing Group*, 8(3):293–299.

Barlow, D. P. (2011). Genomic imprinting: a mammalian epigenetic discovery model. *Annual review of genetics*, 45:379–403.

Barr, M. L. and Bertram, E. G. (1949). A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature*, 163(4148):676.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–837.

Barth, T. K. and Imhof, A. (2010). Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences*, 35(11):618–626.

Bartolomei, M. S. and Ferguson-Smith, A. C. (2011). Mammalian genomic imprinting. *Cold Spring Harbor Perspectives in Biology*, 3(7):a002592.

Barton, S. C., Surani, M. A., and Norris, M. L. (1984). Role of paternal and maternal genomes in mouse development. *Nature*, 311(5984):374–376.

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(7):894–900.

Bengtsson, M. (2005). Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Research*, 15(10):1388–1392.

Bernardi, R. and Pandolfi, P. P. (2007). Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nature Reviews Molecular Cell Biology*, 8(12):1006–1016.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., and Lander,

E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2):315–326.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., and Long, R. M. (1998). Localization of ASH1 mRNA particles in living yeast. *Molecular Cell*, 2(4):437–445.

Bestor, T. H. and Ingram, V. M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 80(18):5559–5563.

Bickmore, W. A. and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. *Cell*, 152(6):1270–1284.

Bird, A. P., Taggart, M. H., and Gehring, C. A. (1981). Methylated and unmethylated ribosomal RNA genes in the mouse. *Journal of molecular biology*, 152(1):1–17.

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., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C. J., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu1, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Dutta, A., Guigó, R., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Flicek, P., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermüller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao,

X., Srinivasan, K. G., Sung, W.-K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Dermitzakis, E. T., Margulies, E. H., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C.-L., Ruan, Y., Snyder, M., Birney, E., Struhl, K., Gerstein, M., Antonarakis, S. E., Gingeras, T. R., Brown, J. B., Flicek, P., Fu, Y., Keefe, D., Birney, E., Denoeud, F., Gerstein, M., Green, E. D., Kapranov, P., Karaöz, U., Myers, R. M., Noble, W. S., Reymond, A., Rozowsky, J., Struhl, K., Siepel, A., Stamatoyannopoulos, J. A., Taylor, C. M., Taylor, J., Thurman, R. E., Tullius, T. D., Washietl, S., Zheng, D., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Collins, F. S., Margulies, E. H., Cooper, G. M., Asimenos, G., Thomas, D. J., Dewey, C. N., Siepel, A., Birney, E., Keefe, D., Hou, M., Taylor, J., Nikolaev, S., Montoya-Burgos, J. I., Löytynoja, A., Whelan, S., Pardi, F., Massingham, T., Brown, J. B., Huang, H., Zhang, N. R., Bickel, P., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., Sequencing Program, N. C., Human Genome Sequencing Center, B. C. o. M., Genome Sequencing Center, W. U., Institute, B., Oakland Research Institute, C. H., Gerstein, M., Antonarakis, S. E., Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Pachter, L., Green, E. D., Sidow, A., Weng, Z., Trinklein, N. D., Fu, Y., Zhang, Z. D., Karaöz, U., Barrera, L., Stuart, R., Zheng, D., Ghosh, S., Flicek, P., King, D. C., Taylor, J., Ameur, A., Enroth, S., Bieda, M. C., Koch, C. M., Hirsch, H. A., Wei, C.-L., Cheng, J., Kim, J., Bhinge, A. A., Giresi, P. G., Jiang, N., Liu, J., Yao, F., Sung, W.-K., Chiu, K. P., Vega, V. B., Lee, C. W. H., Ng, P., Shahab, A., Sekinger, E. A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Clelland, G. K., Wilcox, S., Dillon, S. C., Andrews, R. M., Fowler, J. C., Couttet, P., James, K. D., Lefebvre, G. C., Bruce, A. W., Dovey, O. M., Ellis, P. D., Dhami, P., Langford, C. F., Carter, N. P., Vetrie, D., Kapranov, P., Nix, D. A., Bell, I., Patel, S., Rozowsky, J., Euskirchen, G., Hartman, S., Lian, J., Wu, J., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Hoon Kim, T., Wang,

K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Zhang, X., Xu, M., Haidar, J. N. S., Yu, Y., Birney, E., Weissman, S., Ruan, Y., Lieb, J. D., Iyer, V. R., Green, R. D., Gingeras, T. R., Wadelius, C., Dunham, I., Struhl, K., Hardison, R. C., Gerstein, M., Farnham, P. J., Myers, R. M., Ren, B., Snyder, M., Thomas, D. J., Rosenbloom, K., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Haussler, D., Kent, W. J., Dermitzakis, E. T., Armengol, L., Bird, C. P., Clark, T. G., Cooper, G. M., de Bakker, P. I. W., Kern, A. D., Lopez-Bigas, N., Martin, J. D., Stranger, B. E., Thomas, D. J., Woodroffe, A., Batzoglou, S., Davydov, E., Dimas, A., Eyras, E., Hallgrímsdóttir, I. B., Hardison, R. C., Huppert, J., Sidow, A., Taylor, J., Trumbower, H., Zody, M. C., Guigó, R., Mullikin, J. C., Abecasis, G. R., Estivill, X., Birney, E., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V. V. B., Maskeri, B., McDowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B., and de Jong, P. J. (2007). Identification and analysis of functional elements in 1human genome by the ENCODE pilot project. *Nature*, 447(7146):799–816.

Bix, M. and Locksley, R. M. (1998). Independent and Epigenetic Regulation of the Interleukin-4 Alleles in CD4+ T Cells. *Science*.

Black, J. C., Van Rechem, C., and Whetstine, J. R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Molecular Cell*, 48(4):491–507.

Blau, H. M., Chiu, C. P., and Webster, C. (1983). Cytoplasmic activation of human nuclear genes in stable heterocaryons. *Cell*, 32(4):1171–1180.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959):1509–1512.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). BiQ Ana-

lyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics (Oxford, England)*, 21(21):4067–4068.

Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, 317(5845):1760–1764.

Boyle, S., Rodesch, M. J., Halvensleben, H. A., Jeddeloh, J. A., and Bickmore, W. A. (2011). Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Research*, 19(7):901–909.

Branco, M. R., Ficz, G., and Reik, W. (2011). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nature Reviews Genetics*, 13(1):7–13.

Brannan, C. I., Dees, E. C., Ingram, R. S., and Tilghman, S. M. (1990). The product of the H19 gene may function as an RNA.

Briggs, S. D., Xiao, T., Sun, Z.-W., Caldwell, J. A., Shabanowitz, J., Hunt, D. F., Allis, C. D., and Strahl, B. D. (2002). Gene silencing: trans-histone regulatory pathway in chromatin. *Nature*, 418(6897):498.

Brown, C. J., Lafreniere, R. G., Powers, V. E., Sebastio, G., Ballabio, A., Pettigrew, A. L., Ledbetter, D. H., Levy, E., Craig, I. W., and Willard, H. F. (1991). Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature*, 349(6304):82–84.

Calado, D. P., Paixão, T., Holmberg, D., and Haury, M. (2006). Stochastic monoallelic expression of IL-10 in T cells. *Journal of immunology (Baltimore, Md. : 1950)*, 177(8):5358–5364.

Canham, M. A., Sharov, A. A., Ko, M. S. H., and Brickman, J. M. (2010). Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biology*, 8(5):e1000379.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298(5595):1039–1043.

Cattanach, B. M. (1986). Parental origin effects in mice. *Journal of embryology and experimental morphology*, 97 Suppl:137–150.

Cedar, H. and Bergman, Y. (2011). Epigenetics of haematopoietic cell development. *Nature reviews. Immunology*, 11(7):478–488.

Chaumeil, J., Micsinai, M., Ntziachristos, P., Deriano, L., Wang, J. M.-H., Ji, Y., Nora, E. P., Rodesch, M. J., Jeddeloh, J. A., Aifantis, I., Kluger, Y., Schatz, D. G., and Skok, J. A. (2013). Higher-Order Looping and Nuclear Organization of Tcra Facilitate Targeted RAG Cleavage and Regulated Rearrangement in Recombination Centers. *Cell reports*.

Chen, W. V. and Maniatis, T. (2013). Clustered protocadherins. *Development*, 140(16):3297–3302.

Chess, A. (2012a). Mechanisms and consequences of widespread random monoallelic expression. *Nature Reviews Genetics*, 13(6):421–428.

Chess, A. (2012b). Random and Non-Random Monoallelic Expression. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 38(1):55–61.

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell*, 78(5):823–834.

Chow, J. C. and Heard, E. (2010). Nuclear Organization and Dosage Compensation. *Cold Spring Harbor Perspectives in Biology*, 2(11):a000604–a000604.

Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G., and Li, B. F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science*, 277(5334):1996–2000.

Chubb, J. R., Trcek, T., Shenoy, S. M., and Singer, R. H. (2006). Transcriptional Pulsing of a Developmental Gene. *Current Biology*, 16(10):1018–1025.

Cihák, A. (1974). Biological effects of 5-azacytidine in eukaryotes. *Oncology*, 30(5):405–422.

Cisse, I. I., Izeddin, I., Causse, S. Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2013). Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science*.

Clowney, E. J., Legros, M. A., Mosley, C. P., Clowney, F. G., Markenskoff-Papadimitriou, E. C., Myllys, M., Barnea, G., Larabell, C. A., and Lomvardas, S. (2012). Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell*, 151(4):724–737.

Clowney, E. J., Magklara, A., Colquitt, B. M., Pathak, N., Lane, R. P., and Lomvardas, S. (2011). High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Research*, 21(8):1249–1259.

Coffinier, C., Jung, H.-J., Nobumori, C., Chang, S., Tu, Y., Barnes, R. H., Yoshinaga, Y., de Jong, P. J., Vergnes, L., Reue, K., Fong, L. G., and Young, S. G. (2011). Deficiencies in lamin B1 and lamin B2 cause neurodevelopmental defects and distinct nuclear shape abnormalities in neurons. *Molecular biology of the cell*, 22(23):4683–4693.

Conconi, A., Widmer, R. M., Koller, T., and Sogo, J. M. (1989). Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. *Cell*, 57(5):753–761.

Conrad, T. and Akhtar, A. (2011). Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. *Nature Reviews Genetics*, 13(2):123–134.

Constantinescu, D., Gray, H. L., Sammak, P. J., Schatten, G. P., and Csoka, A. B. (2006). Lamin A/C Expression Is a Marker of Mouse and Human Embryonic Stem Cell Differentiation. *Stem Cells*, 24(1):177–185.

Conti, L., Pollard, S. M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.-L., Cattaneo, E., and Smith, A. (2005). Niche-Independent Symmetrical Self-Renewal of a Mammalian Tissue Stem Cell. *PLoS Biology*, 3(9):e283.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., Abramowitz, L. K., Bartolomei, M. S., Rambow, F., Bassi, M. R., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A. J., Matsumoto, Y., Kobi, D.,

Davidson, I., Alberti, C., Larue, L., and Bellacosa, A. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, 146(1):67–79.

Cremer, T. and Cremer, C. (2006). Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories. *European journal of histochemistry : EJH*, 50(3):161–176.

Cremer, T., Cremer, C., Baumann, H., Luedtke, E. K., Sperling, K., Teuber, V., and Zorn, C. (1982). Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Human Genetics*, 60(1):46–56.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., and Imhof, A. (2002). Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. *Cell*.

D'Angelo, M. A., Gomez-Cavazos, J. S., Mei, A., Lackner, D. H., and Hetzer, M. W. (2012). A Change in Nuclear Pore Complex Composition Regulates Cell Differentiation. *Developmental Cell*.

Davis, T. L., Trasler, J. M., Moss, S. B., Yang, G. J., and Bartolomei, M. S. (1999). Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics*, 58(1):18–28.

Davis, T. L., Yang, G. J., McCarrey, J. R., and Bartolomei, M. S. (2000). The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Human Molecular Genetics*, 9(19):2885–2894.

de Laat, W. and Dekker, J. (2012). 3C-based technologies to study the shape of the genome. *Methods*, 58(3):189–191.

de Wit, E. and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes & Development*, 26(1):11–24.

Deaton, A. M. and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, 25(10):1010–1022.

Dechat, T., Adam, S. A., Taimen, P., Shimi, T., and Goldman, R. D. (2010). Nuclear Lamins. *Cold Spring Harbor Perspectives in Biology*, 2(11):a000547–a000547.

Dechat, T., Pfleghaar, K., Sengupta, K., Shimi, T., Shumaker, D. K., Solimando, L., and Goldman, R. D. (2008). Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. *Genes & Development*, 22(7):832–853.

Dechat, T., Shimi, T., Adam, S. A., Rusinol, A. E., Andres, D. A., Spielmann, H. P., Sinensky, M. S., and Goldman, R. D. (2007). Alterations in mitosis and cell cycle progression caused by a mutant lamin A known to accelerate human aging. *Proceedings of the National Academy of Sciences of the United States of America*, 104(12):4955–4960.

DeChiara, T. M., Robertson, E. J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*, 64(4):849–859.

Denholtz, M. and Plath, K. (2012). Pluripotency in 3D: genome organization in pluripotent cells. *Current Opinion in Cell Biology*.

DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genetics*, 8(3):e1002600.

Dimitrova, D. S. and Gilbert, D. M. (1999). The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Molecular Cell*, 4(6):983–993.

Dion, M. F., Altschuler, S. J., Wu, L. F., and Rando, O. J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. *Proceedings of the National Academy of Sciences of the United States of America*, 102(15):5501–5506.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*.

Donley, N., Stoffregen, E. P., Smith, L., Montagna, C., and Thayer, M. J. (2013). Asynchronous Replication, Mono-Allelic Expression, and Long Range Cis-Effects of ASAR6. *PLoS Genetics*, 9(4):e1003423.

Duester, G. (2008). Retinoic acid synthesis and signaling during early organogenesis. *Cell*, 134(6):921–931.

Duret, L., Chureau, C., Samain, S., Weissenbach, J., and Avner, P. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science*, 312(5780):1653–1655.

Dutta, D., Ensminger, A. W., Zucker, J. P., and Chess, A. (2009). Asynchronous Replication and Autosome-Pair Non-Equivalence in Human Embryonic Stem Cells. *PLoS ONE*, 4(3):e4970.

Eckersley-Maslin, M., Bergmann, J. H., Lazar, Z., and Spector, D. L. (2013). Lamin A/C is Expressed in Pluripotent Mouse Embryonic Stem Cells. *Nucleus*, 4(1).

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D. J., Dash, C., Bazett-Jones, D. P., Le Grice, S., McKay, R. D. G., Buetow, K. H., Gingeras, T. R., Misteli, T., and Meshorer, E. (2008). Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell*, 2(5):437–447.

Eggan, K., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., Axel, R., and Jaenisch, R. (2004). Mice cloned from olfactory sensory neurons. *Nature*, 428(6978):44–49.

Eltsov, M., Maclellan, K. M., Maeshima, K., Frangakis, A. S., and Dubochet, J. (2008). Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proceedings of the National Academy of Sciences*, 105(50):19732–19737.

ENCODE Project Consortium, Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74.

Ensminger, A. W. and Chess, A. (2004). Coordinated replication timing of monoallelically expressed genes along human autosomes. *Human Molecular Genetics*, 13(6):651–658.

Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., Koide, T., Kitsukawa, T., Hamada, S., and Yagi, T. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-α gene cluster in single neurons. *Nature Genetics*, 37(2):171–176.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819):154–156.

Faddah, D. A., Wang, H., Cheng, A. W., Katz, Y., Buganim, Y., and Jaenisch, R. (2013). Single-Cell Analysis Reveals that Expression of Nanog Is Biallelic and Equally Variable as that of Other Pluripotency Factors in Mouse ESCs. *Cell Stem Cell*, 13(1):23–29.

Farago, M., Rosenbluh, C., Tevlin, M., Fraenkel, S., Schlesinger, S., Masika, H., Gouzman, M., Teng, G., Schatz, D., Rais, Y., Hanna, J. H., Mildner, A., Jung, S., Mostoslavsky, G., Cedar, H., and Bergman, Y. (2012). Clonal allelic predetermination of immunoglobulin-κ rearrangement. *Nature*.

Feng, S., Jacobsen, S. E., and Reik, W. (2010). Epigenetic Reprogramming in Plant and Animal Development. *Science*, 330(6004):622–627.

Ficz, G., Hore, T. A., Santos, F., Lee, H. J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., Cook, S. J., Andrews, S., Branco, M. R., and Reik, W. (2013). FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. *Cell Stem Cell*.

Filipczyk, A., Gkatzis, K., Fu, J., Hoppe, P. S., Lickert, H., Anastassiadis, K., and Schroeder, T. (2013). Biallelic expression of nanog protein in mouse embryonic stem cells. *Cell Stem Cell*, 13(1):12–13.

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(7071):1116–1122.

Fisher, C. L. and Fisher, A. G. (2011). Chromatin states in pluripotent, differentiated, and reprogrammed cells. *Current Opinion in Genetics & Development*, 21(2):140–146.

Fleischmann, A., Shykind, B. M., Sosulski, D. L., Franks, K. M., Glinka, M. E., Mei, D. F., Sun, Y., Kirkland, J., Mendelsohn, M., Albers, M. W., and Axel, R. (2008). Mice with a "monoclonal nose": perturbations in an olfactory map impair odor discrimination. *Neuron*, 60(6):1068–1081.

Flicek, P., Ahmed, I., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., García-Girón, C., Gordon, L., Hourlier, T., Hunt, S., Juettemann, T., Kähäri, A. K., Keenan, S., Komorowska, M., Kulesha, E., Longden, I., Maurel, T., McLaren, W. M., Muffato, M., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H. S., Ritchie, G. R. S., Ruffier, M., Schuster, M., Sheppard, D., Sobral, D., Taylor, K., Thormann, A., Trevanion, S., White, S., Wilder, S. P., Aken, B. L., Birney, E., Cunningham, F., Dunham, I., Harrow, J., Herrero, J., Hubbard, T. J. P., Johnson, N., Kinsella, R., Parker, A., Spudich, G., Yates, A., Zadissa, A., and Searle, S. M. J. (2013). Ensembl 2013. *Nucleic Acids Research*, 41(Database issue):D48–55.

Fuss, S. H., Omura, M., and Mombaerts, P. (2007). Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell*, 130(2):373–384.

Fussner, E., Strauss, M., Djuric, U., Li, R., Ahmed, K., Hart, M., Ellis, J., and Bazett-Jones, D. P. (2012). Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. *EMBO reports*, 13(11):992–996.

Georgy, S. T. and Widdicombe, J. G. (2002). The pyrophysiology and sexuality of dragons.

Giambra, V., Volpi, S., Emelyanov, A. V., Pflugh, D., Bothwell, A. L. M., Norio, P., Fan, Y., Ju, Z., Skoultchi, A. I., Hardy, R. R., Frezza, D., and Birshtein, B. K. (2008). Pax5 and linker histone H1 coordinate DNA methylation and histone modifications in the 3' regulatory region of the immunoglobulin heavy chain locus. *Molecular and Cellular Biology*, 28(19):6123–6133.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, 154(2):442–451.

Gimelbrant, A., Ensminger, A. W., Qi, P., Zucker, J., and Chess, A. (2004). Monoallelic Expression and Asynchronous Replication of p120 Catenin in Mouse and Human Cells. *Journal of Biological Chemistry*, 280(2):1354–1359.

Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., and Chess, A. (2007). Widespread Monoallelic Expression on Human Autosomes. *Science*, 318(5853):1136–1140.

Grant, J., Mahadevaiah, S. K., Khil, P., Sangrithi, M. N., Royo, H., Duckworth, J., McCarrey, J. R., Vandeberg, J. L., Renfree, M. B., Taylor, W., Elgar, G., Camerini-Otero, R. D., Gilchrist, M. J., and Turner, J. M. A. (2012). Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature*.

Graumann, J., Hubner, N. C., Kim, J. B., Ko, K., Moser, M., Kumar, C., Cox, J., Schöler, H., and Mann, M. (2008). Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Molecular & cellular proteomics : MCP*, 7(4):672–683.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G. P., Haig, D., and Dulac, C. (2010). High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Mouse Brain. *Science*, 329(5992):643–648.

Greil, F., Moorman, C., and van Steensel, B. (2006). DamID: mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase. *Methods in enzymology*, 410:342–359.

Grigoryev, S. A. and Woodcock, C. L. (2012). Chromatin organization - the 30 nm fiber. *Experimental Cell Research*, 318(12):1448–1455.

Gruenbaum, Y., Cedar, H., and Razin, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature*, 295(5850):620–622.

Gruenbaum, Y., Stein, R., Cedar, H., and Razin, A. (1981). Methylation of CpG sequences in eukaryotic DNA. *FEBS letters*, 124(1):67–71.

Grummt, I. (2007). Different epigenetic layers engage in complex crosstalk to define the epigenetic state of mammalian rRNA genes. *Human Molecular Genetics*, 16 Spec No 1:R21–7.

Guan, X. Y., Trent, J. M., and Meltzer, P. S. (1993). Generation of band-specific painting probes from a single microdissected chromosome. *Human Molecular Genetics*, 2(8):1117–1121.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., Eussen, B. H., de Klein, A., Wessels, L., de Laat, W., and van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453(7197):948–951.

Guibert, S., Forné, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Research*, 22(4):633–641.

Guidi, C. J., Veal, T. M., Jones, S. N., and Imbalzano, A. N. (2004). Transcriptional compensation for loss of an allele of the Ini1 tumor suppressor. *The Journal of biological chemistry*, 279(6):4180–4185.

Gundry, R. L., Tchernyshyov, I., Sheng, S., Tarasova, Y., Raginski, K., Boheler, K. R., and Van Eyk, J. E. (2010). Expanding the mouse embryonic stem cell proteome: Combining three proteomic approaches. *PROTEOMICS*, 10(14):2728–2732.

Guo, G., Huss, M., Tong, G. Q., Wang, C., Sun, L. L., Clarke, N. D., and Robson, P. (2010). Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. *Developmental Cell*, 18(4):675–685.

Guo, L., Hu-Li, J., and Paul, W. E. (2005). Probabilistic Regulation in TH2 Cells Accounts for Monoallelic Expression of IL-4 and IL-13. *Immunity*, 23(1):89–99.

Guo, M. and Birchler, J. A. (1994). Trans-acting dosage effects on the expression of model gene systems in maize aneuploids. *Science*, 266(5193):1999–2002.

Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K. E., Li, W., Myers, R. M., Maniatis, T., and Wu, Q. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. *Proceedings of the National Academy of Sciences*, 109(51):21081–21086.

Gurdon, J. B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *Journal of embryology and experimental morphology*, 10:622–640.

Gutierrez-Arcelus, M., Lappalainen, T., Montgomery, S. B., Buil, A., Ongen, H., Yurovsky, A., Bryois, J., Giger, T., Romano, L., Planchon, A., Falconnet, E., Bielser, D., Gagnebin, M., Padioleau, I., Borel, C., Letourneau, A., Makrythanasis, P., Guipponi, M., Gehrig, C., Antonarakis, S. E., and Dermitzakis, E. T. (2013). Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife*, 2:e00523.

Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H. D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N. C., Benedetti, R., Altucci, L., Jansen, J. H., Walter, J., Gut, I. G., Marks, H., and Stunnenberg, H. G. (2013). Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. *Cell Stem Cell*.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M. A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of development*, 117(1-2):15–23.

Hansen, R. S., Thomas, S., Sandstrom, R., Canfield, T. K., Thurman, R. E., Weaver, M., Dorschner, M. O., Gartler, S. M., and Stamatoyannopoulos, J. A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences*, 107(1):139–144.

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., and Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, 405(6785):486–489.

Heard, E., Clerc, P., and Avner, P. (1997). X-chromosome inactivation in mammals. *Annual review of genetics*, 31:571–610.

Held, W. and Kunz, B. (1998). An allele-specific, stochastic gene expression process controls the expression of multiple Ly49 family genes and generates a diverse, MHC-specific NK cell receptor repertoire. *European journal of immunology*, 28(8):2407–2416.

Held, W., Kunz, B., Ioannidis, V., and Lowin-Kropf, B. (1999). Mono-allelic Ly49 NK cell receptor expression. *Seminars in immunology*, 11(5):349–355.

Held, W. and Raulet, D. H. (1997). Expression of the Ly49A gene in murine natural killer cell clones is predominantly but not exclusively mono-allelic. *European journal of immunology*, 27(11):2876–2884.

Held, W., Roland, J., and Raulet, D. H. (1995). Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells. *Nature*, 376(6538):355–358.

Henikoff, S. and Shilatifard, A. (2011). Histone modification: cause or cog? *Trends in genetics : TIG*, 27(10):389–396.

Hesslein, D. G. T., Pflugh, D. L., Chowdhury, D., Bothwell, A. L. M., Sen, R., and Schatz, D. G. (2003). Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. *Genes & Development*, 17(1):37–42.

Hirano, K., Kaneko, R., Izawa, T., Kawaguchi, M., Kitsukawa, T., and Yagi, T. (2012). Single-neuron diversity generated by Protocadherin-β cluster in mouse central and peripheral nervous systems. *Frontiers in molecular neuroscience*, 5:90.

Hiratani, I. and Gilbert, D. M. (2009). Replication timing as an epigenetic mark. *Epigenetics : official journal of the DNA Methylation Society*, 4(2):93–97.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T. M., Schübeler, D., and Gilbert, D. M. (2008). Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. *PLoS Biology*, 6(10):e245.

Hirota, T., Lipp, J. J., Toh, B.-H., and Peters, J.-M. (2005). Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature*, 438(7071):1176–1180.

Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A., and Burakoff, S. J. (1998). Monoallelic Expression of the Interleukin-2 Locus. *Science*, 279(5359):2118–2121.

Holliday, R. and Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science*, 187(4173):226–232.

Homma, S., Iwasaki, M., Shelton, G. D., Engvall, E., Reed, J. C., and Takayama, S. (2006). BAG3 deficiency results in fulminant myopathy and early lethality. *The American journal of pathology*, 169(3):761–773.

Howell, C. Y., Bestor, T. H., Ding, F., Latham, K. E., Mertineit, C., Trasler, J. M., and Chaillet, J. R. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell*, 104(6):829–838.

Hozumi, N. and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences of the United States of America*, 73(10):3628–3632.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1):44–57.

Huang, S. (2011). Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 366(1575):2247–2259.

Hübner, M. R., Eckersley-Maslin, M., and Spector, D. L. (2013). Chromatin organization and transcriptional regulation. *Current Opinion in Genetics & Development*, 23(2):89–95.

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 Genetics*, 6(9):e1001134.

Ishii, I., Akahoshi, N., Yamada, H., Nakano, S., Izumi, T., and Suematsu, M. (2010). Cystathionine gamma-Lyase-deficient mice require dietary cysteine to protect against acute lethal myopathy and oxidative injury. *Journal of Biological Chemistry*, 285(34):26358–26368.

Jackson, D. A., Hassan, A. B., Errington, R. J., and Cook, P. R. (1993). Visualization of focal sites of transcription within human nuclei. *THE EMBO JOURNA L*, 12(3):1059–1065.

Jahn, D., Schramm, S., Schnölzer, M., Heilmann, C. J., de Koster, C. G., Schütz, W., Benavente, R., and Alsheimer, M. (2012). A truncated lamin A in the Lmna (-/-) mouse line: Implications for the understanding of laminopathies. *Nucleus*, 3(5).

Janicki, S. M., Tsukamoto, T., Salghetti, S. E., Tansey, W. P., Sachidanandam, R., Prasanth, K. V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R. H., and Spector, D. L. (2004). From silencing to gene expression: real-time analysis in single cells. *Cell*, 116(5):683–698.

Jeffries, A. R., Perfect, L. W., Ledderose, J., Schalkwyk, L. C., Bray, N. J., Mill, J., and Price, J. (2012). Stochastic Choice of Allelic Expression in Human Neural Stem Cells. *Stem Cells*, 30(9):1938–1947.

Jeltsch, A. (2006). On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics : official journal of the DNA Methylation Society*, 1(2):63–66.

Jiang, J., Jing, Y., Cost, G. J., Chiang, J.-C., Kolpa, H. J., Cotton, A. M., Carone, D. M., Carone, B. R., Shivak, D. A., Guschin, D. Y., Pearl, J. R., Rebar, E. J., Byron, M., Gregory, P. D., Brown, C. J., Urnov, F. D., Hall, L. L., and Lawrence, J. B. (2013). Translating dosage compensation to trisomy 21. *Nature*, 500(7462):296–300.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816–821.

Jost, J. P. (1993). Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. *Proceedings of the National Academy of Sciences of the United States of America*, 90(10):4684–4688.

Jost, J. P., Siegmann, M., Sun, L., and Leung, R. (1995). Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase. *The Journal of biological chemistry*, 270(17):9734–9739.

Kaasik, A., Kuum, M., Aonurm, A., Kalda, A., Vaarmann, A., and Zharkovsky, A. (2007). Seizures, ataxia, and neuronal loss in cystatin B heterozygous mice. *Epilepsia*, 48(4):752–757.

Kalisky, T., Blainey, P., and Quake, S. R. (2011). Genomic analysis at the single-cell level. *Annual review of genetics*, 45:431–445.

Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., and Yagi, T. (2006). Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *The Journal of biological chemistry*, 281(41):30551–30560.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Human Molecular Genetics*, 16(19):2272–2280.

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., Furlotte, N. A., Eskin, E., Nellåker, C., Whitley, H., Cleak, J., Janowitz, D., Hernandez-Pliego, P., Edwards, A., Belgard, T. G., Oliver, P. L., McIntyre, R. E., Bhomra, A., Nicod, J., Gan, X., Yuan, W., van der Weyden, L., Steward, C. A., Bala, S., Stalker, J., Mott, R., Durbin, R., Jackson, I. J., Czechanski, A., Guerra-Assunção, J. A., Donahue, L. R., Reinholdt, L. G., Payseur, B. A., Ponting, C. P., Birney, E., Flint, J., and Adams, D. J. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, 477(7364):289–294.

Kehayova, P., Monahan, K., Chen, W., and Maniatis, T. (2011). Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proceedings of the National Academy of Sciences*, 108(41):17195–17200.

Kelly, B. L. and Locksley, R. M. (2000). Coordinate regulation of the IL-4, IL-13, and IL-5 cytokine cluster in Th2 clones revealed by allelic expression patterns. *Journal of immunology (Baltimore, Md. : 1950)*, 165(6):2982–2986.

Kelsey, G. and Feil, R. (2013). New insights into establishment and maintenance of DNA methylation imprints in mammals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1609):20110336.

Kerppola, T. K. (2009). Polycomb group complexes–many combinations, many functions. *Trends in Cell Biology*, 19(12):692–704.

Keverne, B. (2009). Monoallelic gene expression and mammalian evolution. *BioEssays*, 31(12):1318–1326.

Khan, M., Vaes, E., and Mombaerts, P. (2011). Regulation of the probability of mouse odorant receptor gene choice. *Cell*, 147(4):907–921.

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 128(6):1231–1245.

Kim, Y., Sharov, A. A., McDole, K., Cheng, M., Hao, H., Fan, C. M., Gaiano, N., Ko, M. S. H., and Zheng, Y. (2011). Mouse B-Type Lamins Are Required for Proper Organogenesis But Not by Embryonic Stem Cells. *Science*, 334(6063):1706–1710.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S. S., Janssen, H., Amendola, M., Nolen, L. D., Bickmore, W. A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. *Cell*, 153(1):178–192.

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D. J., Nicholls, R. D., and Cedar, H. (1993). Allele-specific replication timing of imprinted gene regions. *Nature*, 364(6436):459–463.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biology*, 2(7):E171.

Kosak, S. T. (2002). Subnuclear Compartmentalization of Immunoglobulin Loci During Lymphocyte Development. *Science*, 296(5565):158–162.

Krueger, C., King, M. R., Krueger, F., Miguel R Brancoa, b., Osborne, C. S., Niakan, K. K., Higgins, M. J., and Reik, W. (2012). Pairing of Homologous Regions in the Mouse Genome Is Associated with Transcription but Not Imprinting Status. *PLoS ONE*.

Kubota, A., Kubota, S., Lohwasser, S., Mager, D. L., and Takei, F. (1999). Diversity of NK cell receptor repertoire in adult and neonatal mice. *Journal of immunology (Baltimore, Md. : 1950)*, 163(1):212–216.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & ...*.

Lamond, A. I. and Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nature Reviews Molecular Cell Biology*, 4(8):605–612.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *genesis*, 35(2):88–93.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics*, 13(4):1095–1107.

Latos, P. A., Pauler, F. M., Koerner, M. V., Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., Stricker, S. H., Klement, R. M., Warczok, K. E., Aumayr, K., Pasierbek, P., and Barlow, D. P. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*, 338(6113):1469–1472.

Latos, P. A., Stricker, S. H., Steenpass, L., Pauler, F. M., Huang, R., Senergin, B. H., Regha, K., Koerner, M. V., Warczok, K. E., Unger, C., and Barlow, D. P. (2009). An in vitro ES cell imprinting model shows that imprinted expression of the Igf2r gene arises from an allele-specific expression bias. *Development*, 136(3):437–448.

Lawrence, R. J. and Pikaard, C. S. (2004). Chromatin turn ons and turn offs of ribosomal RNA genes. *Cell Cycle*, 3(7):880–883.

Lee, J. T. and Bartolomei, M. S. (2013). X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. *Cell*, 152(6):1308–1323.

Lee, J. T., Davidow, L. S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genetics*, 21(4):400–404.

Lee, J. T., Strauss, W. M., Dausman, J. A., and Jaenisch, R. (1996). A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell*, 86(1):83–94.

Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A. H. F. M. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *CURBIO*, 13(14):1192–1200.

Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 26(5):589–595.

Li, J., Ishii, T., Feinstein, P., and Mombaerts, P. (2004). Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. *Nature*, 428(6981):393–399.

Li, J., Santoro, R., Koberna, K., and Grummt, I. (2005). The chromatin remodeling complex NoRC controls replication timing of rRNA genes. *THE EMBO JOURNA L*, 24(1):120–127.

Li, L.-C. and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics (Oxford, England)*, 18(11):1427–1431.

Li, M., Liu, G.-H., and Izpisua Belmonte, J. C. (2012a). Navigating the epigenetic landscape of pluripotent stem cells. *Nature Reviews Molecular Cell Biology*, 13(8):524–535.

Li, S. M., Valo, Z., Wang, J., Gao, H., Bowers, C. W., and Singer-Sam, J. (2012b). Transcriptome-Wide Survey of Mouse CNS-Derived Cells Reveals Monoallelic Expression within Novel Gene Families. *PLoS ONE*, 7(2):e31751.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., and Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293.

Lin, M., Hrabovsky, A., Pedrosa, E., Wang, T., Zheng, D., and Lachman, H. M. (2012). Allele-biased expression in differentiating human neurons: implications for neuropsychiatric disorders. *PLoS ONE*, 7(8):e44017.

Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A. L., Chao, J. A., Park, H. Y., de Turris, V., Lopez-Jones, M., and Singer, R. H. (2011). A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nature Methods*, pages 1–9.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., and Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271):315–322.

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell*, 126(2):403–413.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648):251–260.

Luk, E., Ranjan, A., Fitzgerald, P. C., Mizuguchi, G., Huang, Y., Wei, D., and Wu, C. (2010). Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. *Cell*, 143(5):725–736.

Lyle, R., Watanabe, D., Vruchte, D. t., Lerchner, W., Smrzka, O. W., Wutz, A., Schageman, J., Hahner, L., Davies, C., and Barlow, D. P. (2000). The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nature Genetics*, 25(1):19–21.

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature*, 190:372–373.

Lyons, D. B., Allen, W. E., Goh, T., Tsai, L., Barnea, G., and Lomvardas, S. (2013). An epigenetic trap stabilizes singular olfactory receptor expression. *Cell*, 154(2):325–336.

Magklara, A. and Lomvardas, S. (2013). Stochastic gene expression in mammals: lessons from olfaction. *Trends in Cell Biology*.

Magklara, A., Yen, A., Colquitt, B. M., Clowney, E. J., Allen, W., Markenscoff-Papadimitriou, E., Evans, Z. A., Kheradpour, P., Mountoufaris, G., Carey, C., Barnea, G., Kellis, M., and Lomvardas, S. (2011). An Epigenetic Signature for Monoallelic Olfactory Receptor Expression. *Cell*, 145(4):555–570.

Mao, Y. S., Zhang, B., and Spector, D. L. (2011). Biogenesis and function of nuclear bodies. *Trends in Genetics*, pages 1–12.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes & Development*, 11(2):156–166.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A., and Stunnenberg, H. G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*, 149(3):590–604.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 78(12):7634–7638.

Martin, G. R., Epstein, C. J., Travis, B., Tucker, G., Yatziv, S., Martin, D. W., Clift, S., and Cohen, S. (1978). X-chromosome inactivation during differentiation of female teratocarcinoma stem cells in vitro. *Nature*, 271(5643):329–333.

Martinez Arias, A. and Brickman, J. M. (2011). Gene expression heterogeneities in embryonic stem cell populations: origin and function. *Current Opinion in Cell Biology*, 23(6):650–656.

Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., Hupé, P., Barillot, E., Belmont, A. S., and Heard, E. (2011). Live-Cell Chromosome Dynamics and Outcome of X Chromosome Pairing Events during ES Cell Differentiation. *Cell*, 145(3):447–458.

Mattout, A. and Meshorer, E. (2010). Chromatin plasticity and genome organization in pluripotent embryonic stem cells. *Current Opinion in Cell Biology*, 22(3):334–341.

Mayer, C., Neubert, M., and Grummt, I. (2008). The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. *EMBO reports*, 9(8):774–780.

Mayer, C., Schmitz, K.-M., Li, J., Grummt, I., and Santoro, R. (2006). Intergenic transcripts regulate the epigenetic state of rRNA genes. *Molecular Cell*, 22(3):351–361.

McAnally, A. A. and Yampolsky, L. Y. (2010). Widespread transcriptional autosomal dosage compensation in Drosophila correlates with gene expression level. *Genome biology and evolution*, 2:44–52.

McGrath, J. and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, 37(1):179–183.

McMahon, A. and Monk, M. (1983). X-chromosome activity in female mouse embryos heterozygous for Pgk-1 and Searle's translocation, T(X; 16) 16H. *Genetical Research*, 41(01):69–83.

McNairn, A. J. and Gilbert, D. M. (2003). Epigenomic replication: Linking epigenetics to DNA replication. *BioEssays*, 25(7):647–656.

McStay, B. and Grummt, I. (2008). The epigenetics of rRNA genes: from molecular to chromosome biology. *Annual review of cell and developmental biology*, 24:131–157.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., and Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*.

Meister, P., Mango, S. E., and Gasser, S. M. (2011). Locking the genome: nuclear organization and cell fate. *Current Opinion in Genetics & Development*, pages 1–8.

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

Meshorer, E. and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. *Nature Reviews Molecular Cell Biology*, 7(7):540–546.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P. J., Brown, D. T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental Cell*, 22(1):233–234.

Meyer, B. J. (2010). Targeting X chromosomes for repression. *Current Opinion in Genetics & Development*, 20(2):179–189.

Michaelson, J. (1987). CELL SELECTION IN DEVELOPMENT. *Biological Reviews*, 62(2):115–139.

Michaelson, J. (1991). The significance of cell death. *... : The Molecular Basis of Cell Death*.

Michaelson, J. (1993). Cellular selection in the genesis of multicellular organization. *Laboratory investigation; a journal of technical methods and pathology*, 69(2):136–151.

Migeon, B. R., Kazi, E., Haisley-Royster, C., Hu, J., Reeves, R., Call, L., Lawler, A., Moore, C. S., Morrison, H., and Jeppesen, P. (1999). Human X inactivation center induces random X chromosome inactivation in male transgenic mice. *Genomics*, 59(2):113–121.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., RUSS, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E. S., and Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153):553–560.

Miyanari, Y. and Torres-Padilla, M.-E. (2012). Control of ground-state pluripotency by allelic regulation of Nanog. *Nature*, 483(7390):470–473.

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M., and Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular Cell*, 30(6):755–766.

Moir, R. D., Montag-Lowy, M., and Goldman, R. D. (1994). Dynamic properties of nuclear lamins: lamin B is associated with sites of DNA replication. *The Journal of Cell Biology*, 125(6):1201–1212.

Monk, M. and Harper, M. I. (1979). Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature*, 281(5729):311–313.

Moscou, M. J. and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, 326(5959):1501.

Moses, M. J. (1956). Studies on nuclei using correlated cytochemical, light, and electron microscope techniques. *The Journal of biophysical and biochemical cytology*, 2(4 Suppl):397–406.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B. E., Chess, A., Cedar, H., and Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. *Nature*, 414(6860):221–225.

Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E., and Simon, J. A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*, 111(2):197–208.

Murray, K. (1964). The occurrence of epsilon-N-methyl lysine in histones. *Biochemistry*, 3:10–15.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., Tanaka, S., Shiota, K., and Nakano, T. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature cell biology*, 9(1):64–71.

Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Längst, G. (2010). Initial genomics of the human nucleolus. *PLoS Genetics*, 6(3):e1000889.

Nguyen, M. Q., Zhou, Z., Marks, C. A., Ryba, N. J. P., and Belluscio, L. (2007). Prominent roles for odorant receptor coding sequences in allelic exclusion. *Cell*, 131(5):1009–1017.

Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A. S., Imamoto, N., Ishikawa, T., and Maeshima, K. (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO Journal*, 31(7):1644–1653.

Nishizumi, H., Kumasaka, K., Inoue, N., Nakashima, A., and Sakano, H. (2007). Deletion of the core-H region in mice abolishes the expression of three proximal odorant receptor genes in cis. *Proceedings of the National Academy of Sciences*, 104(50):20067–20072.

Niwa, H., Miyazaki, J., and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*, 24(4):372–376.

Nizami, Z., Deryusheva, S., and Gall, J. G. (2010). The Cajal body and histone locus body. *Cold Spring Harbor Perspectives in Biology*, 2(7):a000653.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011). The Dynamic Architecture of Hox Gene Clusters. *Science*, 334(6053):222–225.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., and Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*.

Nutt, S. L., Vambrie, S., Steinlein, P., Kozmik, Z., Rolink, A., Weith, A., and Busslinger, M. (1999). Independent regulation of the two Pax5 alleles during B-cell development. *Nature Genetics*, 21(4):390–395.

Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science*, 248(4962):1517–1523.

Ohno, S., Kaplan, W. D., and Kinosita, R. (1959). Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus. *Experimental Cell Research*, 18:415–418.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, 19(3):219–220.

O'Keefe, R. T., Henderson, S. C., and Spector, D. L. (1992). Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *The Journal of Cell Biology*, 116(5):1095–1110.

Olins, A. L. and Olins, D. E. (1974). Spheroid chromatin units (v bodies). *Science*, 183(4122):330–332.

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C. D., Cheng, X., and Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, 448(7154):714–717.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO reports*, 10(11):1235–1241.

Paixão, T., Carvalho, T. P., Calado, D. P., and Carneiro, J. (2007). Quantitative insights into stochastic monoallelic expression of cytokine genes. *Immunology and Cell Biology*, 85(4):315–322.

Pajerowski, J. D., Dahl, K. N., Zhong, F. L., Sammak, P. J., and Discher, D. E. (2007). Physical plasticity of the nucleus in stem cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(40):15619–15624.

Papantonis, A. and Cook, P. R. (2010). Genome architecture and the role of transcription. *Current Opinion in Cell Biology*, 22(3):271–276.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research*, 37(18):e123.

Pasque, V., Radzisheuskaya, A., Gillich, A., Halley-Stott, R. P., Panamarova, M., Zernicka-Goetz, M., Surani, M. A., and Silva, J. C. R. (2012). Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. *Journal of Cell Science*.

Pauler, F. M., Barlow, D. P., and Hudson, Q. J. (2012). Mechanisms of long range silencing by imprinted macro non-coding RNAs. *Current Opinion in Genetics & Development*, 22(3):283–289.

Payer, B., Saitou, M., Barton, S. C., Thresher, R., Dixon, J. P. C., Zahn, D., Colledge, W. H., Carlton, M. B. L., Nakano, T., and Surani, M. A. (2003). Stella is a maternal effect gene required for normal early development in mice. *CURBIO*, 13(23):2110–2117.

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561):131–137.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S. W. M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R. M., van Lohuizen, M., Reinders, M., Wessels, L., and van Steensel, B. (2010). Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Molecular Cell*, 38(4):603–613.

Peric-Hupkes, D. and van Steensel, B. (2011). Role of the Nuclear Lamina in Genome Organization and Gene Expression. *Cold Spring Harbor Symposia on Quantitative Biology*, 75(0):517–524.

Pernis, B., Chiappino, G., Kelus, A. S., and Gell, P. G. (1965). Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *The Journal of experimental medicine*, 122(5):853–876.

Phillips, D. M. (1963). The presence of acetyl groups of histones. *The Biochemical journal*, 87:258–263.

Pickersgill, H., Kalverda, B., de Wit, E., Talhout, W., Fornerod, M., and van Steensel, B. (2006). Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nature Genetics*, 38(9):1005–1014.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.-K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development*, 135(18):3081–3091.

Pollex, T. and Heard, E. (2012). Recent advances in X-chromosome inactivation research. *Current Opinion in Cell Biology*, 24(6):825–832.

Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, 463(7284):1101–1105.

Raj, A. and van Oudenaarden, A. (2009). Single-molecule approaches to stochastic gene expression. *Annual review of biophysics*, 38:255–270.

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA

methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10):5237–5242.

Rando, O. J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Current Opinion in Genetics & Development*, 22(2):148–155.

Rastan, S. and Robertson, E. J. (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. *Journal of embryology and ....*

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796):593–599.

Reeves, R. H., Irving, N. G., Moran, T. H., Wohn, A., Kitt, C., Sisodia, S. S., Schmidt, C., Bronson, R. T., and Davisson, M. T. (1995). A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nature Genetics*, 11(2):177–184.

Renfree, M. B., Suzuki, S., and Kaneko-Ishino, T. (2013). The origin and evolution of genomic imprinting and viviparity in mammals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1609):20120151.

Rhind, N. and Gilbert, D. M. (2013). DNA replication timing. *Cold Spring Harbor Perspectives in Biology*, 5(8).

Rhoades, K. L., Singh, N., Simon, I., Glidden, B., Cedar, H., and Chess, A. (2000). Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. *CURBIO*, 10(13):789–792.

Ribich, S., Tasic, B., and Maniatis, T. (2006). Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proceedings of the National Academy of Sciences of the United States of America*, 103(52):19719–19724.

Riggs, A. D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenetics and cell genetics*, 14(1):9–25.

Röber, R. A., Weber, K., and Osborn, M. (1989). Differential timing of nuclear lamin A/C expression in the various organs of the mouse embryo and the young animal: a developmental study. *Development*, 105(2):365–378.

Rothbart, S. B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A. I., Barsyte-Lovejoy, D., Martinez, J. Y., Bedford, M. T., Fuchs, S. M., Arrowsmith, C. H., and Strahl, B. D. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nature Structural & Molecular Biology*, 19(11):1155–1160.

Rougeulle, C. (2003). Promoter-restricted H3 Lys 4 di-methylation is an epigenetic mark for monoallelic expression. *Human Molecular Genetics*, 12(24):3343–3348.

Rouhi, A., Gagnier, L., Takei, F., and Mager, D. L. (2006). Evidence for epigenetic maintenance of Ly49a monoallelic gene expression. *Journal of immunology (Baltimore, Md. : 1950)*, 176(5):2991–2999.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T. C., Robins, A. J., Dalton, S., and Gilbert, D. M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Research*, 20(6):761–770.

Sago, H., Carlson, E. J., Smith, D. J., Kilbridge, J., Rubin, E. M., Mobley, W. C., Epstein, C. J., and Huang, T. T. (1998). Ts1Cje, a partial trisomy 16 mouse model for Down syndrome, exhibits learning and behavioral abnormalities. *Proceedings of the National Academy of Sciences of the United States of America*, 95(11):6256–6261.

Santoro, R. and Grummt, I. (2001). Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Molecular Cell*, 8(3):719–725.

Santoro, R., Li, J., and Grummt, I. (2002). The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nature Genetics*, 32(3):393–396.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. *Developmental biology*, 241(1):172–182.

Schatten, G., Maul, G. G., Schatten, H., Chaly, N., Simerly, C., Balczon, R., and Brown, D. L. (1985). Nuclear lamins and peripheral nuclear antigens during fertilization and embryogenesis in mice and sea urchins. *Proceedings of the National Academy of Sciences of the United States of America*, 82(14):4727–4731.

Schatz, D. G., Oettinger, M. A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. *Cell*, 59(6):1035–1048.

Scheffer, M. P., Eltsov, M., and Frangakis, A. S. (2011). Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. *Proceedings of the National Academy of Sciences*, 108(41):16992–16997.

Schlesinger, S., Selig, S., Bergman, Y., and Cedar, H. (2009). Allelic inactivation of rDNA loci. *Genes & Development*, 23(20):2437–2447.

Schulz, E. G. and Heard, E. (2013). Role and control of X chromosome dosage in mammalian development. *Current Opinion in Genetics & Development*, 23(2):109–115.

Schwaiger, M., Stadler, M. B., Bell, O., Kohler, H., Oakeley, E. J., and Schübeler, D. (2009). Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome. *Genes & Development*, 23(5):589–601.

Searle, A. G. (1968). Comparative genetics of coat colour in mammals. *Comparative genetics of coat colour in mammals*.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Molecular Cell*.

Seisenberger, S., Peat, J. R., Hore, T. A., Santos, F., Dean, W., and Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1609):20110330.

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., and Sakano, H. (2003). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science*, 302(5653):2088–2094.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3):458–472.

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., Trombetta, J. T., Gennert, D., Gnirke, A., Goren, A., Hacohen, N., Levin, J. Z., Park, H., and Regev, A. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*.

Sharif, J., Muto, M., Takebayashi, S.-I., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., and Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, 450(7171):908–912.

Sharman, G. B. (1971). Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature*, 230(5291):231–232.

Shykind, B. M. (2005). Regulation of odorant receptors: one allele at a time. *Human Molecular Genetics*, 14 Spec No 1:R33–9.

Silvers, W. K. (1958). Origin and identity of clear cells found in hair bulbs of albino mice. *The Anatomical Record*, 130(2):135–144.

Sinsheimer, R. L. (1955). The action of pancreatic deoxyribonuclease. II. Isomeric dinucleotides. *The Journal of biological chemistry*, 215(2):579–583.

Skene, P. J. and Henikoff, S. (2013). Histone variants in pluripotency and disease. *Development*, 140(12):2513–2524.

Skok, J. A., Brown, K. E., Azuara, V., Caparros, M. L., Baxter, J., Takacs, K., Dillon, N., Gray, D., Perry, R. P., Merkenschlager, M., and Fisher, A. G. (2001). Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nature Immunology*, 2(9):848–854.

Smallwood, S. A., Tomizawa, S.-I., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S. R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genetics*, 43(8):811–814.

Smith, E. R., Zhang, X.-Y., Capo-Chichi, C. D., Chen, X., and Xu, X.-X. (2011). Increased expression of Syne1/nesprin-1 facilitates nuclear envelope structure changes in embryonic stem cell differentiation. *Developmental Dynamics*, 240(10):2245–2255.

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*.

Smith, Z. D. and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, 14(3):204–220.

Solter, D. (1988). Differential imprinting and expression of maternal and paternal genomes. *Annual review of genetics*, 22:127–146.

Song, J., Rechkoblit, O., Bestor, T. H., and Patel, D. J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science*, 331(6020):1036–1040.

Spada, F., Haemmer, A., Kuch, D., Rothbauer, U., Schermelleh, L., Kremmer, E., Carell, T., Längst, G., and Leonhardt, H. (2007). DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *The Journal of Cell Biology*, 176(5):565–571.

Spector, D. L. (2001). Nuclear domains. *Journal of Cell Science*, 114(Pt 16):2891–2893.

Spector, D. L. (2006). SnapShot: Cellular Bodies. *Cell*, 127(5):1071.e1–1071.e2.

Stack, S. M., Brown, D. B., and Dewey, W. C. (1977). Visualization of interphase chromosomes. *Journal of Cell Science*, 26:281–299.

Stewart, C. and Burke, B. (1987). Teratocarcinoma stem cells and early mouse embryos contain only a single major lamin polypeptide closely resembling lamin B. *Cell*, 51(3):383–392.

Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765):41–45.

Subrahmanyam, R. and Sen, R. (2012). Epigenetic features that regulate IgH locus recombination and expression. *Current topics in microbiology and immunology*, 356:39–63.

Sullivan, G. J., Bridger, J. M., Cuthbert, A. P., Newbold, R. F., Bickmore, W. A., and McStay, B. (2001). Human acrocentric chromosomes with transcriptionally silent nucleolar organizer regions associate with nucleoli. *THE EMBO JOURNA L*, 20(11):2867–2874.

Sullivan, T., Escalante-Alcalde, D., Bhatt, H., Anver, M., Bhat, N., Nagashima, K., Stewart, C. L., and Burke, B. (1999). Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *The Journal of Cell Biology*, 147(5):913–920.

Surani, M. A., Barton, S. C., and Norris, M. L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, 308(5959):548–550.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929):930–935.

Takagi, N. and Abe, K. (1990). Detrimental effects of two active X chromosomes on early mouse development. *Development*, 109(1):189–201.

Takagi, N., Sugawara, O., and Sasaki, M. (1982). Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. *Chromosoma*, 85(2):275–286.

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–676.

Takei, F., McQueen, K. L., Maeda, M., Wilhelm, B. T., Lohwasser, S., Lian, R. H., and Mager, D. L. (2001). Ly49 and CD94/NKG2: developmentally regulated expression and evolution. *Immunological reviews*, 181:90–103.

Takizawa, T., Gudla, P. R., Guo, L., Lockett, S., and Misteli, T. (2008a). Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes & Development*, 22(4):489–498.

Takizawa, T., Meaburn, K. J., and Misteli, T. (2008b). The meaning of gene positioning. *Cell*, 135(1):9–13.

Tanamachi, D. M., Hanke, T., Takizawa, H., Jamieson, A. M., and Raulet, D. R. (2001). Expression of natural killer receptor alleles at different Ly49 loci occurs independently and is regulated by major histocompatibility complex class I molecules. *The Journal of experimental medicine*, 193(3):307–315.

Tang, F., Barbacioru, C., Nordman, E., Bao, S., Lee, C., Wang, X., Tuch, B. B., Heard, E., Lao, K., and Surani, M. A. (2011). Deterministic and Stochastic Allele Specific Gene Expression in Single Mouse Blastomeres. *PLoS ONE*, 6(6):e21208.

Tasic, B., Nabholz, C. E., Baldwin, K. K., Kim, Y., Rueckert, E. H., Ribich, S. A., Cramer, P., Wu, Q., Axel, R., and Maniatis, T. (2002). Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Molecular Cell*, 10(1):21–33.

Thomas, B. J., Rubio, E. D., Krumm, N., Broin, P. Ó., Bomsztyk, K., Welcsh, P., Greally, J. M., Golden, A. A., and Krumm, A. (2011). Allele-specific transcriptional elongation regulates monoallelic expression of the IGF2BP1 gene. *Epigenetics & Chromatin*, 4(1):14.

Thompson, N. E., Steinberg, T. H., Aronson, D. B., and Burgess, R. R. (1989). Inhibition of in vivo and in vitro transcription by monoclonal antibodies prepared against wheat germ RNA polymerase II that react with the heptapeptide repeat of eukaryotic RNA polymerase II. *The Journal of biological chemistry*, 264(19):11511–11520.

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science*, 282(5391):1145–1147.

Thorvaldsen, J. L., Duran, K. L., and Bartolomei, M. S. (1998). Deletion of the H19 differen-

tially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes & Development*, 12(23):3693–3702.

Trieu, M., Ma, A., Eng, S. R., Fedtsova, N., and Turner, E. E. (2003). Direct autoregulation and gene dosage compensation by POU-domain transcription factor Brn3a. *Development*, 130(1):111–121.

Ueda, T., Abe, K., Miura, A., Yuzuriha, M., Zubair, M., Noguchi, M., Niwa, K., Kawase, Y., Kono, T., Matsuda, Y., Fujimoto, H., Shibata, H., Hayashizaki, Y., and Sasaki, H. (2000). The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes to Cells*, 5(8):649–659.

Valiante, N. M., Uhrberg, M., Shilling, H. G., Lienert-Weidenbach, K., Arnett, K. L., D'Andrea, A., Phillips, J. H., Lanier, L. L., and Parham, P. (1997). Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors. *Immunity*, 7(6):739–751.

van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G. J., Ariyurek, Y., den Dunnen, J. T., and Lamond, A. I. (2010). High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Molecular biology of the cell*, 21(21):3735–3748.

Vielle, A., Lang, J., Dong, Y., Ercan, S., Kotwaliwale, C., Rechtsteiner, A., Appert, A., Chen, Q. B., Dose, A., Egelhofer, T., Kimura, H., Stempor, P., Dernburg, A., Lieb, J. D., Strome, S., and Ahringer, J. (2012). H4K20me1 contributes to downregulation of X-linked genes for C. elegans dosage compensation. *PLoS Genetics*, 8(9):e1002933.

Wang, J., Valo, Z., Bowers, C. W., Smith, D. D., Liu, Z., and Singer-Sam, J. (2010). Dual DNA Methylation Patterns in the CNS Reveal Developmentally Poised Chromatin and Monoallelic Expression of Critical Genes. *PLoS ONE*, 5(11):e13843.

Wang, J., Valo, Z., Smith, D., and Singer-Sam, J. (2007). Monoallelic expression of multiple genes in the CNS. *PLoS ONE*, 2(12):e1293.

Wang, X., Su, H., and Bradley, A. (2002). Molecular mechanisms governing Pcdh-gamma gene expression: evidence for a multiple promoter and cis-alternative splicing model. *Genes & Development*, 16(15):1890–1905.

Wernig, M., Lengner, C. J., Hanna, J., Lodato, M. A., Steine, E., Foreman, R., Staerk, J., Markoulaki, S., and Jaenisch, R. (2008). A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotechnology*, 26(8):916–924.

Wheway, G., Abdelhamed, Z., Natarajan, S., Toomes, C., Inglehearn, C., and Johnson, C. A. (2013). Aberrant Wnt signalling and cellular over-proliferation in a novel mouse model of Meckel-Gruber syndrome. *Developmental biology*, 377(1):55–66.

Williams, R. R. E. (2006). Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *Journal of Cell Science*, 119(1):132–140.

Woodcock, C. L. (1994). Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. *The Journal of Cell Biology*, 125(1):11–19.

Wossidlo, M., Arand, J., Sebastiano, V., Lepikhov, K., Boiani, M., Reinhardt, R., Schöler, H., and Walter, J. (2010). Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *The EMBO Journal*, 29(11):1877–1888.

Wutz, A., Rasmussen, T. P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature Genetics*, 30(2):167–174.

Xie, W., Barr, C. L., Kim, A., Yue, F., Lee, A. Y., Eubanks, J., Dempster, E. L., and Ren, B. (2012). Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell*, 148(4):816–831.

Xie, X., Mikkelsen, T. S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E. S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17):7145–7150.

Yagi, T. (2008). Clustered protocadherin family. *Development, growth & differentiation*, 50 Suppl 1:S131–40.

Yamanaka, S. and Blau, H. M. (2010). Nuclear reprogramming to a pluripotent state by three approaches. *Nature*, 465(7299):704–712.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature*, 453(7194):519–523.

Yokota, S., Hirayama, T., Hirano, K., Kaneko, R., Toyoda, S., Kawamura, Y., Hirabayashi, M., Hirabayashi, T., and Yagi, T. (2011). Identification of the cluster control region for the protocadherin-beta genes located beyond the protocadherin-gamma cluster. *Journal of Biological Chemistry*, 286(36):31885–31895.

Young, J. M., Friedman, C., Williams, E. M., Ross, J. A., Tonnes-Priddy, L., and Trask, B. J. (2002). Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Human Molecular Genetics*, 11(5):535–546.

Yunger, S., Rosenfeld, L., Garini, Y., and Shav-Tal, Y. (2010). Single-allele analysis of transcription kinetics in living mammalian cells. *Nature Publishing Group*, 7(8):631–633.

Zentner, G. E. and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. *Nature Structural & Molecular Biology*, 20(3):259–266.

Zernicka-Goetz, M. and Huang, S. (2010). Stochasticity versus determinism in development: a false dichotomy? *Nature Reviews Genetics*, 11(11):743–744.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, 322(5902):750–756.

Zhao, R., Bodnar, M. S., and Spector, D. L. (2009). Nuclear neighborhoods and gene expression. *Current Opinion in Genetics & Development*, 19(2):172–179.

Zhou, Y., Santoro, R., and Grummt, I. (2002). The chromatin remodeling complex NoRC targets

HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. *THE EMBO JOURNA L*, 21(17):4632–4640.

Zhu, J.-K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annual review of genetics*, 43:143–166.

Zwemer, L. M., Zak, A., Thompson, B., Kirby, A., Daly, M. J., Chess, A., and Gimelbrant, A. (2012). Autosomal monoallelic expression in the mouse. *Genome Biology*, 13(2):R10.

Appendix A

List of random autosomal monoallelically expressed genes

A.1 List of monoallelically expressed genes in ESCs

Gene ID	Gene Name	Class
ENSMUSG00000078184	B020018G12Rik	A
ENSMUSG00000032666	1700025G04Rik	B
ENSMUSG00000074635	3110070M22Rik	B
ENSMUSG00000021185	9030617O03Rik	B
ENSMUSG00000029695	Aass	B
ENSMUSG00000028127	Abcd3	B
ENSMUSG00000087439	AC118260.5	B
ENSMUSG00000086294	AC142101.1	B
ENSMUSG00000086627	AC164431.1	B
ENSMUSG00000078445	AC168276.1	B
ENSMUSG00000022477	Aco2	B
ENSMUSG00000022229	Atp12a	B
ENSMUSG00000034112	Atp2c2	B
ENSMUSG00000036905	C1qb	B
ENSMUSG00000053819	Camk2d	B
ENSMUSG00000040189	Ccdc114	B
ENSMUSG00000068547	Clca6	B
ENSMUSG00000001119	Col6a1	B
ENSMUSG00000054196	Cthrc1	B
ENSMUSG00000004267	Eno2	B
ENSMUSG00000038776	Ephx1	B
ENSMUSG00000033373	Fntb	B
ENSMUSG00000031714	Gab1	B
ENSMUSG00000037280	Galnt6	B
ENSMUSG00000069170	Gpr98	B
ENSMUSG00000020176	Grb10	B
ENSMUSG00000054128	H2-T3	B
ENSMUSG00000059447	Hadhb	B
ENSMUSG00000024423	Impact	B
ENSMUSG00000025825	Iscu	B
ENSMUSG00000018362	Kpna2	B
ENSMUSG00000036853	Mcoln3	B
ENSMUSG00000042814	Mcts2	B
ENSMUSG00000051855	Mest	B
ENSMUSG00000031647	Mfap3l	B
ENSMUSG00000003948	Mmd	B
ENSMUSG00000023939	Mrpl14	B
ENSMUSG00000034422	Parp14	B
ENSMUSG00000029231	Pdgfra	B
ENSMUSG00000006494	Pdk1	B
ENSMUSG00000019817	Plagl1	B

ENSMUSG00000027695	Pld1	B
ENSMUSG00000027750	Postn	B
ENSMUSG00000010609	Psen2	B
ENSMUSG00000036858	Ptcra	B
ENSMUSG00000070953	Rabepk	B
ENSMUSG00000004952	Rasa4	B
ENSMUSG00000005774	Rfx5	B
ENSMUSG00000086391	RP23-119N24.3	B
ENSMUSG00000080989	RP23-136C22.3	B
ENSMUSG00000086730	RP23-13C17.5	B
ENSMUSG00000084291	RP23-158O11.5	B
ENSMUSG00000045464	RP23-16O15.2	B
ENSMUSG00000044751	RP23-173F16.2	B
ENSMUSG00000075276	RP23-263O6.4	B
ENSMUSG00000078640	RP23-57F11.1	B
ENSMUSG00000022032	Scara5	B
ENSMUSG00000028780	Sema3c	B
ENSMUSG00000025020	Slit1	B
ENSMUSG00000020409	Slu7	B
ENSMUSG00000051111	Sv2c	B
ENSMUSG00000020059	Sycp3	B
ENSMUSG00000056130	Ticam2	B
ENSMUSG00000029810	Tmem176b	B
ENSMUSG00000027656	Wisp2	B
ENSMUSG00000021287	Xrcc3	B
ENSMUSG00000027514	Zbp1	B

A.2 List of monoallelically expressed genes in NPCs

Gene ID	Gene Name	Class
ENSMUSG00000006931	1110036O03Rik	A
ENSMUSG00000045107	1810063B07Rik	A
ENSMUSG00000038059	2010002N04Rik	A
ENSMUSG00000031851	2310079N02Rik	A
ENSMUSG00000021747	4930452B06Rik	A
ENSMUSG00000073755	5730409E04Rik	A
ENSMUSG00000021792	5730469M10Rik	A
ENSMUSG00000073858	AC122912.1	A
ENSMUSG00000072949	Acot1	A
ENSMUSG00000060923	Acyp2	A
ENSMUSG00000018340	Anxa6	A
ENSMUSG00000042082	Arsb	A
ENSMUSG00000030847	Bag3	A
ENSMUSG00000008999	Bmp7	A
ENSMUSG00000022440	C1qtnf6	A
ENSMUSG00000073418	C4b	A
ENSMUSG00000040373	Cacng5	A
ENSMUSG00000021373	Cap2	A
ENSMUSG00000022947	Cbr3	A
ENSMUSG00000027435	Cd93	A
ENSMUSG00000024769	Cdc42bpg	A
ENSMUSG00000044303	Cdkn2a	A
ENSMUSG00000026042	Col5a2	A
ENSMUSG00000031825	Crispld2	A
ENSMUSG00000016256	Ctsz	A
ENSMUSG00000040327	Cul9	A
ENSMUSG00000039519	Cyp7b1	A
ENSMUSG00000073609	D2hgdh	A
ENSMUSG00000029821	Dfna5	A
ENSMUSG00000039095	En2	A
ENSMUSG00000029122	Evc	A
ENSMUSG00000024691	Fam111a	A
ENSMUSG00000001555	Fkbp10	A
ENSMUSG00000002732	Fkbp7	A
ENSMUSG00000031451	Gas6	A
ENSMUSG00000027316	Gfra4	A
ENSMUSG00000028005	Gucy1b3	A
ENSMUSG00000025534	Gusb	A
ENSMUSG00000059325	Hopx	A
ENSMUSG00000074811	Hps6	A
ENSMUSG00000060591	Ifitm2	A

ENSMUSG00000055980	Irs1	A
ENSMUSG00000041921	Metapl1	A
ENSMUSG00000039533	Mmd2	A
ENSMUSG00000055430	Nap1l5	A
ENSMUSG00000042684	Npl	A
ENSMUSG00000003849	Nqo1	A
ENSMUSG00000029468	P2rx7	A
ENSMUSG00000048347	Pcdhb18	A
ENSMUSG00000045498	Pcdhb3	A
ENSMUSG00000051242	Pcdhb9	A
ENSMUSG00000019577	Pdk4	A
ENSMUSG00000036218	Pdzrn4	A
ENSMUSG00000026664	Phyh	A
ENSMUSG00000060675	Pla2g16	A
ENSMUSG00000023913	Pla2g7	A
ENSMUSG00000028494	Plin2	A
ENSMUSG00000049409	Prokr1	A
ENSMUSG00000020415	Pttg1	A
ENSMUSG00000030559	Rab38	A
ENSMUSG00000026475	Rgs16	A
ENSMUSG00000021464	Ror2	A
ENSMUSG00000042312	S100a13	A
ENSMUSG00000042216	Sgsm1	A
ENSMUSG00000041540	Sox5	A
ENSMUSG00000024776	Stambpl1	A
ENSMUSG00000023885	Thbs2	A
ENSMUSG00000036975	Tmem177	A
ENSMUSG00000073968	Trim68	A
ENSMUSG00000058672	Tubb2a	A
ENSMUSG00000062591	Tubb4	A
ENSMUSG00000035799	Twist1	A
ENSMUSG00000034777	Vax2	A
ENSMUSG00000024076	Vit	A
ENSMUSG00000085906	AC020971.2	A+B
ENSMUSG00000072591	AC122789.1	A+B
ENSMUSG00000021814	Anxa7	A+B
ENSMUSG00000057789	Bak1	A+B
ENSMUSG00000030077	Chl1	A+B
ENSMUSG00000065979	Cpped1	A+B
ENSMUSG00000050248	Evc2	A+B
ENSMUSG00000028268	Gbp3	A+B
ENSMUSG00000034245	Hdac11	A+B

ENSMUSG00000008540	Mgst1	A+B
ENSMUSG00000046949	Nqo2	A+B
ENSMUSG00000046768	Rhoj	A+B
ENSMUSG00000037455	1110021L09Rik	B
ENSMUSG00000025466	1810014F10Rik	B
ENSMUSG00000071632	2510002D24Rik	B
ENSMUSG00000033111	3830406C13Rik	B
ENSMUSG00000062822	4833420G17Rik	B
ENSMUSG00000052688	5430435G22Rik	B
ENSMUSG00000025971	9430016H08Rik	B
ENSMUSG00000020620	Abca8b	B
ENSMUSG00000028970	Abcb1b	B
ENSMUSG00000046858	AC107766.1	B
ENSMUSG00000089955	AC107766.1	B
ENSMUSG00000085644	AC108401.1	B
ENSMUSG00000075536	AC113595.2	B
ENSMUSG00000078697	AC119848.1	B
ENSMUSG00000060647	AC121569.2	B
ENSMUSG00000089856	AC122789.1	B
ENSMUSG00000075444	AC123606.1	B
ENSMUSG00000073233	AC124134.1	B
ENSMUSG00000087651	AC125276.1	B
ENSMUSG00000079599	AC129184.1	B
ENSMUSG00000061461	AC134463.1	B
ENSMUSG00000079149	AC134581.1	B
ENSMUSG00000062862	AC138791.1	B
ENSMUSG00000086101	AC158956.1	B
ENSMUSG00000056938	Acbd4	B
ENSMUSG00000040272	Accs	B
ENSMUSG00000017697	Ada	B
ENSMUSG00000053441	Adamts19	B
ENSMUSG00000030000	Add2	B
ENSMUSG00000020473	Aebp1	B
ENSMUSG00000038729	Akap2	B
ENSMUSG00000090182	AL929021.1	B
ENSMUSG00000035561	Aldh1b1	B
ENSMUSG00000030088	Aldh1l1	B
ENSMUSG00000004849	Ap1s1	B
ENSMUSG00000009585	Apobec3	B
ENSMUSG00000032812	Arap1	B
ENSMUSG00000037509	Arhgef4	B
ENSMUSG00000062031	Athl1	B

ENSMUSG00000031441	Atp11a	B
ENSMUSG00000007097	Atp1a2	B
ENSMUSG00000078349	AW011738	B
ENSMUSG00000080717	B230307C23Rik	B
ENSMUSG00000042215	Bag2	B
ENSMUSG00000046598	Bdh1	B
ENSMUSG00000038286	Bphl	B
ENSMUSG00000071317	Bves	B
ENSMUSG00000024592	C330018D20Rik	B
ENSMUSG00000021991	Cacna2d3	B
ENSMUSG00000035547	Capn5	B
ENSMUSG00000051483	Cbr1	B
ENSMUSG00000022665	Ccdc80	B
ENSMUSG00000025510	Cd151	B
ENSMUSG00000005087	Cd44	B
ENSMUSG00000026031	Cflar	B
ENSMUSG00000029161	Cgref1	B
ENSMUSG00000056486	Chn1	B
ENSMUSG00000090258	Churc1	B
ENSMUSG00000006782	Cnp	B
ENSMUSG00000044681	Cnpy1	B
ENSMUSG00000040690	Col16a1	B
ENSMUSG00000031502	Col4a1	B
ENSMUSG00000027570	Col9a3	B
ENSMUSG00000056941	Commd7	B
ENSMUSG00000051811	Cox6b2	B
ENSMUSG00000034796	Cpne7	B
ENSMUSG00000027408	Cpxm1	B
ENSMUSG00000042109	Csdc2	B
ENSMUSG00000028804	Csmd2	B
ENSMUSG00000005054	Cstb	B
ENSMUSG00000028179	Cth	B
ENSMUSG00000083282	Ctsf	B
ENSMUSG00000031924	Cyb5b	B
ENSMUSG00000034445	Cybasc3	B
ENSMUSG00000015224	Cyp2j9	B
ENSMUSG00000022150	Dab2	B
ENSMUSG00000045608	Dbx2	B
ENSMUSG00000039450	Dcxr	B
ENSMUSG00000040296	Ddx58	B
ENSMUSG00000025815	Dhtkd1	B
ENSMUSG00000027560	Dok5	B

ENSMUSG00000047205	Dusp18	B
ENSMUSG00000034488	Edil3	B
ENSMUSG00000078135	Eid1	B
ENSMUSG00000048988	Elfn1	B
ENSMUSG00000029675	Eln	B
ENSMUSG00000021728	Emb	B
ENSMUSG00000022505	Emp2	B
ENSMUSG00000004267	Eno2	B
ENSMUSG00000015766	Eps8	B
ENSMUSG00000075703	Ept1	B
ENSMUSG00000027714	Exosc9	B
ENSMUSG00000052397	Ezr	B
ENSMUSG00000029851	Fam115c	B
ENSMUSG00000070044	Fam149a	B
ENSMUSG00000078670	Fam174b	B
ENSMUSG00000054942	Fam73a	B
ENSMUSG00000024778	Fas	B
ENSMUSG00000064080	Fbln2	B
ENSMUSG00000040913	Fbxw4	B
ENSMUSG00000003355	Fkbp11	B
ENSMUSG00000025175	Fn3k	B
ENSMUSG00000081683	Fzd10	B
ENSMUSG00000034793	G6pc3	B
ENSMUSG00000025579	Gaa	B
ENSMUSG00000018567	Gabarap	B
ENSMUSG00000035473	Galm	B
ENSMUSG00000015027	Galns	B
ENSMUSG00000047261	Gap43	B
ENSMUSG00000027199	Gatm	B
ENSMUSG00000028214	Gem	B
ENSMUSG00000056966	Gjc3	B
ENSMUSG00000071543	Gm10336	B
ENSMUSG00000043004	Gng2	B
ENSMUSG00000029464	Gpn3	B
ENSMUSG00000069170	Gpr98	B
ENSMUSG00000049583	Grm5	B
ENSMUSG00000060803	Gstp1	B
ENSMUSG00000001665	Gstt3	B
ENSMUSG00000023079	Gtf2ird1	B
ENSMUSG00000019188	H13	B
ENSMUSG00000061232	H2-K1	B
ENSMUSG00000053835	H2-T24	B

ENSMUSG00000058189	Hist1h2bm	B
ENSMUSG00000025188	Hps1	B
ENSMUSG00000041548	Hspb8	B
ENSMUSG00000058258	Idi1	B
ENSMUSG00000035551	Igfbpl1	B
ENSMUSG00000022969	Il10rb	B
ENSMUSG00000002897	Il17ra	B
ENSMUSG00000032763	Ilvbl	B
ENSMUSG00000001504	Irx2	B
ENSMUSG00000015533	Itga2	B
ENSMUSG00000027009	Itga4	B
ENSMUSG00000025348	Itga7	B
ENSMUSG00000073859	Itpripl2	B
ENSMUSG00000025736	Jmjd8	B
ENSMUSG00000027895	Kcnc4	B
ENSMUSG00000044708	Kcnj10	B
ENSMUSG00000037624	Kcnk2	B
ENSMUSG00000063229	Ldha	B
ENSMUSG00000033880	Lgals3bp	B
ENSMUSG00000019906	Lin7a	B
ENSMUSG00000000693	Loxl3	B
ENSMUSG00000025507	Lrdd	B
ENSMUSG00000021579	Lrrc14b	B
ENSMUSG00000039246	Lyplal1	B
ENSMUSG00000045854	Lyrm2	B
ENSMUSG00000037523	Mavs	B
ENSMUSG00000042814	Mcts2	B
ENSMUSG00000030621	Me3	B
ENSMUSG00000030291	Med21	B
ENSMUSG00000028655	Mfsd2a	B
ENSMUSG00000054612	Mgmt	B
ENSMUSG00000019823	Mical1	B
ENSMUSG00000034121	Mks1	B
ENSMUSG00000020287	Mpg	B
ENSMUSG00000031760	Mt3	B
ENSMUSG00000015222	Mtap2	B
ENSMUSG00000045636	Mtus1	B
ENSMUSG00000020814	Mxra7	B
ENSMUSG00000067818	Myl9	B
ENSMUSG00000031478	Nek3	B
ENSMUSG00000022454	Nell2	B
ENSMUSG00000037499	Nenf	B

ENSMUSG00000039873	Neurl2	B
ENSMUSG00000020248	Nfyb	B
ENSMUSG00000031773	Nlrc5	B
ENSMUSG00000026946	Nmi	B
ENSMUSG00000022421	Nptxr	B
ENSMUSG00000017176	Nt5c3l	B
ENSMUSG00000024228	Nudt12	B
ENSMUSG00000029153	Ociad2	B
ENSMUSG00000029012	Orc5l	B
ENSMUSG00000030774	Pak1	B
ENSMUSG00000090053	Palm2	B
ENSMUSG00000089945	Palm2-Akap2	B
ENSMUSG00000033377	Palmd	B
ENSMUSG00000038507	Parp12	B
ENSMUSG00000022438	Parvb	B
ENSMUSG00000001497	Pax9	B
ENSMUSG00000049100	Pcdh10	B
ENSMUSG00000007440	Pcdha11	B
ENSMUSG00000051486	Pcdhb11	B
ENSMUSG00000047307	Pcdhb13	B
ENSMUSG00000047910	Pcdhb16	B
ENSMUSG00000051599	Pcdhb2	B
ENSMUSG00000045689	Pcdhb4	B
ENSMUSG00000045876	Pcdhb8	B
ENSMUSG00000030513	Pcsk6	B
ENSMUSG00000078931	Pdf	B
ENSMUSG00000031595	Pdgfrl	B
ENSMUSG00000070526	Peg12	B
ENSMUSG00000042632	Pla2g6	B
ENSMUSG00000021822	Plau	B
ENSMUSG00000024197	Plin3	B
ENSMUSG00000031775	Pllp	B
ENSMUSG00000031570	Ppapdc1b	B
ENSMUSG00000042133	Ppig	B
ENSMUSG00000037826	Ppm1k	B
ENSMUSG00000026778	Prkcq	B
ENSMUSG00000039405	Prss23	B
ENSMUSG00000024337	Psmb9	B
ENSMUSG00000028378	Ptgr1	B
ENSMUSG00000072946	Ptgr2	B
ENSMUSG00000063235	Ptpmt1	B
ENSMUSG00000026204	Ptprn	B

ENSMUSG00000004044	Ptrf	B
ENSMUSG00000040511	Pvr	B
ENSMUSG00000072566	Pvt1	B
ENSMUSG00000033684	Qsox1	B
ENSMUSG00000027953	Rag1ap1	B
ENSMUSG00000025921	Rdh10	B
ENSMUSG00000025350	Rdh5	B
ENSMUSG00000020037	Rfx4	B
ENSMUSG00000021719	Rgs7bp	B
ENSMUSG00000042671	Rgs8	B
ENSMUSG00000048330	Ric3	B
ENSMUSG00000035226	Rims4	B
ENSMUSG00000086290	RP23-115D21.2	B
ENSMUSG00000044751	RP23-173F16.2	B
ENSMUSG00000075276	RP23-263O6.4	B
ENSMUSG00000047905	RP23-265P8.1	B
ENSMUSG00000086998	RP23-29H5.7	B
ENSMUSG00000086061	RP23-308N2.7	B
ENSMUSG00000087596	RP23-317F9.4	B
ENSMUSG00000086859	RP23-41E14.4	B
ENSMUSG00000041959	S100a10	B
ENSMUSG00000034173	Scand3	B
ENSMUSG00000025203	Scd2	B
ENSMUSG00000026504	Sdccag8	B
ENSMUSG00000000753	Serpinf1	B
ENSMUSG00000070436	Serpinh1	B
ENSMUSG00000024548	Setbp1	B
ENSMUSG00000022436	Sh3bp1	B
ENSMUSG00000030638	Sh3gl3	B
ENSMUSG00000024600	Slc27a6	B
ENSMUSG00000027661	Slc2a10	B
ENSMUSG00000049922	Slc35c1	B
ENSMUSG00000027894	Slc6a17	B
ENSMUSG00000025889	Snca	B
ENSMUSG00000034891	Sncb	B
ENSMUSG00000044349	Snhg11	B
ENSMUSG00000006050	Sra1	B
ENSMUSG00000070003	Ssbp4	B
ENSMUSG00000024172	St6gal2	B
ENSMUSG00000004043	Stat5a	B
ENSMUSG00000033855	Ston1	B
ENSMUSG00000020903	Stx8	B

ENSMUSG00000028369	Svep1	B
ENSMUSG00000019769	Syne1	B
ENSMUSG00000000384	Tbrg4	B
ENSMUSG00000027868	Tbx15	B
ENSMUSG00000020034	Tcp11l2	B
ENSMUSG00000035686	Thrsp	B
ENSMUSG00000031639	Tlr3	B
ENSMUSG00000025572	Tmc6	B
ENSMUSG00000026109	Tmeff2	B
ENSMUSG00000069763	Tmem100	B
ENSMUSG00000024666	Tmem138	B
ENSMUSG00000026939	Tmem141	B
ENSMUSG00000025933	Tmem14a	B
ENSMUSG00000043140	Tmem186	B
ENSMUSG00000050777	Tmem37	B
ENSMUSG00000028822	Tmem50a	B
ENSMUSG00000025505	Tmem80	B
ENSMUSG00000035413	Tmem98	B
ENSMUSG00000053475	Tnfaip6	B
ENSMUSG00000022074	Tnfrsf10b	B
ENSMUSG00000090170	Tnfsf12	B
ENSMUSG00000000934	Top1mt	B
ENSMUSG00000074607	Tox2	B
ENSMUSG00000019842	Traf3ip2	B
ENSMUSG00000044528	Tram1l1	B
ENSMUSG00000025413	Ttc4	B
ENSMUSG00000029201	Ugdh	B
ENSMUSG00000005501	Usp40	B
ENSMUSG00000079490	Vmn2r-ps112	B
ENSMUSG00000034040	Wbscr17	B
ENSMUSG00000030170	Wnt5b	B
ENSMUSG00000028329	Xpa	B
ENSMUSG00000021287	Xrcc3	B
ENSMUSG00000087598	Zfp111	B
ENSMUSG00000042097	Zfp239	B
ENSMUSG00000028389	Zfp37	B
ENSMUSG00000059878	Zfp422	B
ENSMUSG00000045598	Zfp553	B
ENSMUSG00000023284	Zfp605	B
ENSMUSG00000067928	Zfp760	B
ENSMUSG00000054716	Zfp771	B

Appendix B

Epigenetic drug screen results

Table showing results from first two trials of the epigenetic drug screen. Compound name, target category and specific target are shown. Drugs highlighted in blue showed weak activation in 1 or 2 trials, yellow indicate drugs which led to more convincing activation in 2 or more trails. Trial 1 was conducted for 4 days using 10μM drug. Growth and cell morphology characteristics are noted. The expressed allele for each gene are listed. Highlighted in yellow are partial activation of the previously silent allele, blue indicates very weak activation. x indicates sequencing traces which were not of sufficient quality to determine allele information.

	Contents	Target Category	Specific Target	Trial #1: 4 days 10μM			Trial #2: variable dosage, 4 days		
				Notes	Gas6	Serpinh1	dosage (μM)	Gas6	Cap2
A1	DMSOneg control	negative control			x	A	10	C	A
A2	3-amino Benzamide	PARP	PARP		x	x	25	C	x
A3	Scriptaid	HDAC	HDAC	slowed growth	x	x	5		
A4	F-Amidine (trifluoroacetate salt)	Peptidyl arginine deiminase	PAD4	neurospheres	X	x	25		A
A5	5-Azacytidine	DNA methylation	DNA methyltransferases	slowed growth	C	A	10	C	A
A6	IOX1	2OG oxygenases	2OG oxygenases, JMJD2a, JMJD2E, PHF8, PHD2, FIH		x	A	25	C	A
A7	Tenovin-6	HDAC	SIRT1, SIRT2, SIRT3	cell death			5		
A8	M 344	HDAC	HDAC1, HDAC6		C/T	A	10	C (T)	A/G
A9	RG-108	DNA methylation	DNA methyltransferases	neurospheres	x	x	25	C	A
A10	Sinefungin				x	A	25	C	A
A11	Piceatannol	protein kinases	wide range of tyrosine/threonine protein kinases		x	A	25		
B2	SB 939	HDAC	all HDAC except HDAC6, HDAC7	cell death			5	x	x
B4	JGB174	HDAC	SIRT1		x	x	10	C	A
B5	Decitabine	DNA methylation	DNA methyltransferases	elongated	C/T	x	10	x	x
B6	MI-2 (hydrochloride)	histone methyltransferase	menin	slowed growth	C	x	10	x	x
B7	Sodium Butyrate	HDAC	HDAC	neurospheres	x	A	10	C	A
B8	Oxamflatin	HDAC	HDAC	cell death			5		
B9	2',3',5'-triacetyl-5-Azacytidine	DNA methylation	DNA methyltransferases		x	A	25	C	A
B10	trans-Resveratrol	antioxidant	COX-1		x	A	25	C	A (G)
B11	EX-527	HDAC	SIRT1		C (T)	A	10	x	A
C2	PCI 34051	HDAC	HDAC8	neurospheres	x	A	10		
C3	Apicidin	HDAC	HDAC, HDAC3	slowed growth			5		
C4	UNC0638	histone methyltransferase	G9a	cell death			5		
C5	(+)-JQ1	BRD4	BRD4	neurospheres	x	x	10	C	A
C6	MI-nc (hydrochloride)	histone methyltransferase	Menin		x	A	25	C	A
C7	Anacardic Acid	HAT	p300, pCAF	cell death			5		
C8	Salermide	HDAC	SIRT1, SIRT2		x	A	25	C	A
C9	S-Adenosylhomocysteine	DNA methylation	activated methyl cycle, cystein biosynthesis	neurospheres	x	x	25	C	A
C10	2,4-DPD	HIF	HIF-PH	neurospheres	x	x	25	C	A (G)
C11	SAHA	HDAC	Class I and II HDAC	slowed growth	C	x	5		
D2	4-iodo-SAHA	HDAC	HDAC1, HDAC6	cell death			5	x	x
D3	HC Toxin	HDAC	HDACs	cell death			5		
D4	Phthalazinone pyrazole	Aurora A Kinase	Aurora A kinase	neurospheres	x	x	5	x	x
D5	(-)-JQ1	none	no affinity for Brd4	neurospheres	x	A	10		
D6	Gemcitabine	DNA replication/repair	Gadd45a	elongated	x	x	10	x	x
D7	AGK2	HDAC	SIRT2		x	x	25	C	A (G)
D8	Mirin	DNA repair	Mre11-Rad50-Nbs1	neurospheres	x	A	25		

				Trial #1: 4 days 10μM			Trial #2: variable dosage, 4 days		
	Contents	Target Category	Specific Target	Notes	Gas6	Serpinh1	dosage (μM)	Gas6	Cap2
D9	UNC0224	histone methyltransferase	G9a	neurospheres	x	A	25	C	A (G)
D10	DMOG	HIF	HIF-PH		x	A	25	C	A (G)
D11	2-PCPA (hydrochloride)	LSD1	LDS1		x	A	25	C (T)	A
E2	Sirtinol	HDAC	SIRT1, SIRT2		C	A	25		
E3	UNC0321 (trifluoroacetate salt)	histone methyltransferase	G9a		X	A	25	C	A
E4	Isoliquiritigenin	antioxidant	quinone reductase I	slowed growth	C	A	5	C	A
E5	BSI-201	DNA repair	PARP1		X	A	25	C	A
E6	Lomeguatrib	DNA repair	O6-methylguanine-DNA methyltransferase (MGMT)		x	A	25	C	A
E7	CAY10603	HDAC	HDAC6	elongated	C	A	10	C (T)	A (G)
E8	Pimelic Diphenylamide 106	HDAC	Class I HDAC		C (T)	A	10	C (T)	A
E9	Chidamide	HDAC	HDAC	elongated	C (T)	A	10	C	C/A
E10	Trichostatin A	HDAC	HDAC1	cell death			5		
E11	Nicotinamide				x	A	25	C	A (G)
F2	Tubastatin A (trifluoroacetate salt)	HDAC	HDAC6	elongated	C	x	25		
F4	CCG-100602	Rho	RhoA/C		x	x	25		
F6	Daminozide	histone demethylase	KDM2A, PHF8, KDM7A		x	x	25	x	x
F8	(S)-HDAC-42	HDAC	HDAC	cell death			5		
F10	CAY10398	HDAC	HDAC1	elongated	C	A	25		
G2	Garcinol	HAT	p300, pCAF	cell death			5		
G3	(-)-Neplanocin A	methyltransferase	SAH hydrolase		C	A	5	C	A
G4	Zebularine	DNA methylation	DNA methyltransferases		x	x	25	C	A (G)
G6	Valproic Acid (sodium salt)	HDAC	Class I HDAC		x	x	25	x	A (G)
G8	MS-275	HDAC	HDAC1	slowed growth	C/T		10	C	A/G
G9	3-Deazaneplanocin A	EZH2	EZH2		C		25	C	
G10	2,4-Pyridinedicarboxylic Acid				x	A	25	C	A
G11	N-Oxalylglycine				C	x	25	C	A
H2	Ellagic Acid	histone methyltransferase	cytochrome P450, CARM1	neurospheres	x	A	25	C	A (G)
H5	AG-014699	PARP	PARP1	slowed growth	C	A	5	C	A (G)
H6	Tenovin-1	HDAC	SIRT1, 2	cell death			5	C	A (G)
H7	CBHA	HDAC	HDAC1, HDAC3	elongated	C	x	25	C (A)	A (G)
H8	HNHA	HDAC	HDAC		x	A	25	C	A (G)
H10	CAY10433	HDAC	HDAC	elongated	C	A	25	C	A (G)
H11	Suramin (sodium salt)				x	A	25	C (A)	A

Appendix C

Detailed experimental protocols

C.1 ESC to NPC differentiation protocol

Differentiation of mouse ESCs to NPCs

10% N2B27 media:

250ml	DMEM/F12 1x, (Gibco, common stock)
250ml	Neurobasal medium (1x, Gibco 21103049)
0.5ml	N2 supplement (100x, Gibco 17502-048, stored at -20)
4ml (20mg)	Insulin (Sigma I-1882 made to 5mg/ml in PBS, stored at -20)
0.667ml	BSA Fraction V Solution (7.5% Gibco 15260-001)
1.0ml	B27 supplement (50x, Gibco 0080085SA, stored at -20)
5ml	P/S

sterile filter before use. Keep 4 degrees in dark. Use within 2 weeks.

Alternatively of N2 supplement plus insulin (N2+) you can use home-made modified N2 supplement. The only difference between Mod N2 and N2+ is the concentration of insulin:

	1x	100x Mod N2	10ml Mod N2	100x N2 +
Insulin (Sigma I-1882)	25mg/L	2.5mg/ml	25mg	0.5mg/ml
Apotransferin (Sigma T1147)	100ug/ml	10mg/ml	100mg	10mg/ml
Progesterone (Sigma P8783)	6ng/ml	600ng/ml	6ug	630ng/ml
Putrescine (Sigma P5780)	16ug/ml	1.6mg/ml	16mg	1.611mg/ml
Sodium selenite (Sigma P5261)	30nM	3uM	5.1882ug	3uM

N2 expansion media:

500ml	DMEM/F2	
333ul	7.5% BSA Fraction V	final 50ug/ml
20ul	250ug/ml mEGF	final 10ng/ml
20ul	250ug/ml mFGF	final 10ng/ml
5ml	P/S	

filter, make 50ml aliquots. Add to 50ml prior to use:

50ul	1mg/ml laminin	final 1ug/ml
500ul	100x N2 supplement	final 1x

Protocol:

1. Grow ES cells on MEF feeders until confluent
2. Trypsinise, soak on gelatin to remove MEFs 1hr at 37 degrees
3. Collect ESC, count cells
4. Plate onto gelatin coated 10cm plate at different densities ranging from $0.5x10^6$ cells to $2.0x10^6$ cells per 10cm plate in N2B27 media. Typically use 3 or 4 different densities as differentiation efficiency varies.
5. Differentiate cells for 6 days. Change media every 2 days.
6. Trypsinise plate and count cells. Suspend between $1.5x10^6$ and $3.0x10^6$ cells total in uncoated T75 flask in N2 expansion media.

7. Aggregates will start to grow out in suspension over the next 4 days.
8. Collect day 3 or 4 aggregates by mild centrifugation.
9. Plate aggregates onto 1x 10cm gelatin coated plate in N2 expansion media.
10. NPC should adhere within a few days. Once cells have settled, change N2 expansion media every 2-3 days. Split as required.

C.2 RNA-sequencing library preparation

Long RNA-seq protocol (paired-end stranded library)

Start with 10ug of total RNA.

PolyA+ isolation (Qiagen Oligotex kit)

* use DEPC-treated water
* preheat Oligotex suspension to 37 degrees, mix by vortexing then keep at room temp
* heat water bath or heating block to 70 degrees, heat 400ul of buffer OEB per sample
* ensure that buffer OBB does not have precipitates by prewarming at 37 degrees for 10 minutes then place at room temperature.
* perform all steps at room temperature unless otherwise indicated.
* all centrifugation steps should be performed in a microcentrifuge tube at max speed (14,000g to 18,000g)

1. Pipet 10ug total RNA into an RNase-free 1.5ml microcentrifuge tube and adjust the volume of water to 250ul.
2. Add 250ul buffer OBB, 15ul oligotex suspension. Mix thoroughly by vortexing or flicking the tube
3. Incubate 3 minutes at 70 degrees to disrupt secondary structure
4. Remove sample from waterbath/heating block and place at room temperature for 12 minutes to allow hybridisation between oligo dT30 and polyA tails
5. Centrifuge 2 minutes at 14,000-18,000g, room temperature. Collect and save the supernatant (polyA minus fraction). It doesn't matter if not all of the supernatant is collected.
6. Resuspend the pellet in 1ml buffer OW2 by pipetting. Make sure pellet is completely resuspended. Centrifuge 12,000g 2 minutes. Carefully remove supernatant.
7. Wash again in 1ml buffer OW2. Be careful when removing supernatant, often it is necessary to remove all but ~100ul, spin down again and then remove the rest.
8. Add 100ul preheated buffer OEB (70 degrees). Resuspend by pipetting, place back at 70 degrees for 10 seconds before centrifuging 2 minutes 12,000g room temp.
9. Transfer supernatant containing polyA+ RNA to new microcentrifuge.
10. Resuspend again with 100ul preheated buffer OEB. Add the supernatant to the polyA+ fraction.
11. For second round of polyA+ purification repeat steps 1-10. Otherwise continue to ethanol precipitation.
12. Spin polyA+ RNA in spin filter column for 1 minutes at 18,000g to remove any remaining oligotex suspension from the polyA+ RNA. Transfer flowthrough to a new tube as the Ambion tubes don't close very well.
13. Add 1ul glycoblue, 1/10V 3M sodium acetate pH5.5, 3V 100% EtOH. Incubate -70 for at least 30 minutes
14. Centrifuge 30 minutes 4 degrees 15,000g
15. Wash 1x with 70% EtOH (-20degrees), remove EtOH.
16. Either airdry or put in speedvac for 4 minutes to remove residual EtOH which can interfere with subsequent reactions.
17. Resuspend pellect in 10ul H2O on ice for 5 minutes.

Ribominus treatment

*use 10ul of polyA+ RNA from previous step or <10ug of total RNA.
*set a waterbath or heat block to 70 degrees

1. Add to 1-10ug of RNA, 10ul of ribominus probe, 100ul hybridsation buffer
2. Incubate at 70 degrees for 5 minutes to denature the RNA.
3. Cool sample slowly over 30 minutes by placing tube in 37 degree heat block to allow sequence specific hybridisation.
4. Prepare beads during the incubation:
    a. Vortex ribominus beads thoroughtly, pipet 750ul into a sterile 1.5ml tube
    b. Place on magnet for 1 minute, remove supernatant.
    c. Add 750ul sterile DEPC water, vortex, place on magnet, discard supernatant
    d. Repeat wash with 750ul water
    e. Resuspend in 750ul hybridisation buffer and transfer 250ul to a new tube.
    f. Place the tube with 500ul on magnet for 1 minute, remove supernatant and resuspend in 200ul hybridisation buffer.
    g. Keep both tubes at 37 degrees until needed.
5. After 37 degree incubation, transfer ~120ul RNA-probe sample to the prepared ribominus beads (200ul beads). Mix well
6. Incubate 37 degrees for 15 minutes, gently mix occasionally.
7. Briefly centrifuge, place on magnet for 1 minute. DO NOT DISCARD SUPERNATANT AS THIS CONTAINS THE RNA!
8. Place the tube with 250ul beads on magent 1 minute, remove supernatant
9. Transfer ribominus RNA from the first tube to the second tube of beads. Mix well by pipetting.
10. Incubate 37 degrees for 15 minutes, gently mix occasionally.
11. Place tube on magnetic separator for 1 minute, transfer the supernatant containing ribominus RNA to a small filter column and spin at max speed for 2 minutes to remove any residual magnetic particles.
12. Transfer flow through to a new tube
13. Add 1/10V 3M sodium acetate pH5.5, 3V 100% EtOH (glycoblue from polyA purification will still be present). Incubate -70 for at least 30 minutes
14. Centrifuge 30 minutes 4 degrees 15,000g
15. Wash 1x with 70% EtOH (-20degrees), remove EtOH.
16. Either airdry or put in speedvac for 4 minutes to remove residual EtOH which can interfere with subsequent reactions.
17. Resuspend pellect in 4ul H2O on ice for 5 minutes.

cDNA-1st strand synthesis

*Add all of the polyA+ ribominus RNA from 10ug of total RNA.
*if have 2 or more samples make up mastermixes for all steps

1. To 4ul of RNA add:
   - 1.6ul random primers (50ng/ul, invitrogen)
   - 2ul polydT20 (50uM, invitrogen)
   - 1ul NIST spike-ins
2. Start PCR program:
   - 98 degrees 2 min
   - 70 degrees 5 min
   - 0.1deg/sec to 15 degrees
   - PAUSE
3. As soon as 15 degrees is reached (after ~15 minutes), add:
   - 4ul Superscript III 1st strand buffer (5x, invitrogen)
   - 1ul 0.1M MgCl2 (diluted from 1M MgCl2 Ambion stock)
   - 1ul 10mM dNTPs (invitrogen)
   - 2ul 0.1M DTT
   - 1ul RNase Inhibitor (Ambion 20U/ul)
   - 0.5ul H2O
4. The reaction total should be 17.9ul
5. Resume PCR program:
   - 15 degrees 30 min
   - PAUSE
6. After 30 min at 15 degrees, pause program and add:
   - 1.0ul actinomycin-D (120ng/ul in 10mM Tris pH7.6, dilute from 1mg/ml stock before use)
   - 1.1ul superscript III enzyme (invitrogen)
7. The reaction total should be 20ul
8. Resume PCR program (approx 1 hour 40 minutes)
   - 0.1deg/seec to 25 degrees
   - 25 degrees 10 min
   - 0.1deg/sec to 42 degrees
   - 42 degrees 45 min
   - 0.1 deg/sec to 50 degrees
   - 50 degrees 15 min
   - 75 degrees 15 min
   - 4 degrees hold
9. Bring total reaction volume to 100ul with H2O. Add 5 volumes of buffer PB
10. Add to minelute Qiagen spin column.
11. Centrifuge 1 minute 10,000g
12. Wash column 1x with buffer PE
13. Centrifuge 1 minute 10,000g
14. Remove flow through, centrifuge 1 minute 12,000g
15. Add 16ul of EB to column, sit 1 minute at room temp, spin 12,000g.
16. Elute again with 15ul EB. Pool sample (~30ul).

2nd strand synthesis

* add enzymes last in order listed in protocol to prevent RNase H activity before DNA pol is present.
* prepare reaction on ice

1. Prepare 2nd strand mix:

2ul	5x first strand buffer (invitrogen)
15ul	5x second strand buffer (invitrogen)
0.5ul	0.1M MgCl2
1ul	0.1 M DTT
2ul	dUNTP mix (10mM each of dATP, dCTP, dGTP, dUTP)
0.5ul	E. coli DNA ligase (10U/ul)
2ul	E. coli DNA polymerase I (10U/ul)
0.5ul	RNase H (2U/ul)
21.5ul	RNase free H2O

2. Add 45ul second strand mix to 30ul of purified 1st strand reaction, bringing total reaction volume to 75ul
3. Incubate 2 hours at 16 degrees, hold at 4 degrees in PCR machine
4. Bring total reaction volume to 100ul with H2O. Add 5 volumes of buffer PB
5. Add to minelute Qiagen spin column.
6. Centrifuge 1 minute 10,000g
7. Wash column 1x with buffer PE
8. Centrifuge 1 minute 10,000g
9. Remove flow through, centrifuge 1 minute 12,000g
10. Add 26ul of EB to column, sit 1 minute at room temp, spin 12,000g.
11. Elute again with 25ul EB. Pool sample (~50ul).
12. Save 1.5ul to run on bioanalyser DNA high-sensitivity chip (pre-fragmentation)

Fragmentation of ds cDNA using Covaris

* If machine is off: switch machine on, ensure chambers are filled with autoclaved DI water. Run degas program prior to fragmenting samples (~30 minutes)

1. Transfer 50ul sample to covaris microtube using a pipette
2. Place in machine by snapping into place
3. Run program ‘degas100ulsnapcap60sec’
4. sonication takes 60 seconds
5. run 1ul on DNA high-sensitivity chip (post-fragmentation). Fragmentation size should have a peak at 200-300.

End-Repair cDNA

48ul	sample
27ul	H2O
10ul	T4 DNA ligase buffer with 10mM ATP
4ul	10mM dNTP mix
5ul	T4 DNA polymerase 3U/ul(NEB M0203)
1ul	Klenow DNA polymerase 5U/ul(NEB M0210)
5ul	T4 PNK 10U/ul(NEB M0201)
100ul	

Incubate room temperature 30 minutes

Add 500ul PB, clean-up using Qiagen minelute columns. Elute 2 x 16ul

Addition of single A base

32ul	eluted cDNA
5ul	NEB buffer 2
10ul	1mM dATP
3ul	Klenow fragment 3' to 5' exo –5U/ul (NEB M0212)
50ul	

Incubate 37 degrees, 30 minutes

Bring volume to 100ul with 50ul H2O, Add 500ul PB, minelute columns, Elute 1 x 19ul

Adapter Ligation

19ul	eluted cDNA
25ul	2x Rapid DNA ligase buffer (Enzymatics B101)
1ul	Illumina Paired-End adapter oligo mix
5ul	DNA T4 ligase (Enzymatics 600U/ul)
50ul	

Incubate room temperature 30 minutes

Bring volume to 100ul with 50ul H2O, Add 500ul PB, minelute columns. Elute 1 x 15ul

UNG treatment

15ul	eluted cDNA
1.7ul	500mM KCl
1ul	UNG (Roche N808-0096)

Indubate 37 degrees 15 minutes, 95 degrees 10 minutes. Hold on ice.

Gel purification

Add 10ul of loading dye to 17.7ul UNG treated sample

Run on 2% ultra-pure agarose gel for 2 hours at 90V. Use 100bp ladder and have a spare lane between samples.
Cut out 200bp band and another band at about 250bp. It is normal not to see anything on the gel, cut out gel anyway. Freeze larger slice.

Weigh out gel slice (~120g). Add 3V buffer QG, dissolve 15-20 minutes at 55 degrees. Add 1V isopropanol. Load onto minelute column, spin through. Wash 1x 0.5ml buffer QG, 1x 0.75ml buffer PE. Dry spin x1. Eulte 2 x 15ul buffer EB

PCR amplification

Use 15ul of eluted cDNA from gel purification. Save other 15ul incase PCR does not work.

15ul	eluted cDNA
1ul	PE primer 1.0 (100uM HPLC purified)
1ul	PE primer 2.0 (100uM HPLC purified)
50ul	2x HF Phusion Mix (Finzymes)
33ul	H2O (incase need to add more or less cDNA, can adjust this amount)
100ul	

Cycle conditions:

98 degrees	1 min	
98 degrees	10s	
60 degrees	30s	18 cycles
72 degrees	30s	
72 degrees	5 min	
hold at 4 degrees		

Add 500ul PB, minelute clean-up, elute 1x15ul

Gel Purification

Add 10ul of loading dye to 15ul eluted PCR product

Run on 1% ultra-pure agarose gel for 2 hours at 90V. Use 100bp ladder and have a spare lane between samples.
Cut out band about 100pb larger than cDNA band.
Gel purify as above, elute 2x 25ul EB.

Dilute to 100ul with 50ul H2O. Add 10ul Sodium acetate, 330ul EtOH. Precipitate 30 min at -80 degrees or overnight at -20. Centrifuge 30 min at max speed. Wash 1x in 70% EtOH. Airdry or speedvac for 4 min. Resuspend in 25ul H20.

Library Quantitation

Run Agilent DNA high sensitivity chip. Run 2 dilutions of sample at 1:20 and 1:30. At least one of the dilutions should be between max and half max of loading peak height. Calculate the peak size (should be consistent between dilutions) and take the average of the concentration.

Dilute library to 10nM.

Send 25ul or half of library, which ever is less, to sequencers. Keep remaining library as backup.

Reagents required – separate stocks of everything to prevent contamination!!!

Oligotex mRNA midi kit (12 reactions)	Qiagen Cat # 70042	$673
Glycoblue (300 reactions)	Ambion Cat # AM9515	$65.25
Ribominus kit (8 reactions)	Invitrogen Cat # A10837-08	$482
Superscript III RT (2,000U)	Invitrogen Cat # 18080-093	$80
Random primers	Invitrogen Cat # 48190-011	$138
Oligo-dT20 primers	Invitrogen Cat # 18418-020	$130
NIST spike-ins	from Gingeras lab	-
RNAse Inhibitor	Ambion Cat # AM2690	$82
1M MgCl2	Ambion Cat # AM9530G	$27.90
10mM Tris-HCl pH7.6	Sigma Cat # T2444-100mL	$21.50
Actinomycin-D (5mg)	Invitrogen 11805-017	$78.25
5x second strand buffer	Invitrogen 10812-014	$110
dUTP	Roche #11934554001	?
dNTPs	Roche # 11969064001	?
E. Coli DNA ligase	Invitrogen Cat # 18052-019	$39
E. coli DNA polymerase I	Invitrogen Cat # 18010-017	$99.25
RNase H	Invitrogen Cat # 18021-014	$128
Bioanalyser high-sensitivity DNA chips	Agilent Cat # 5067-4626	$453
Bioanalyser RNA nano chip reagents	Agilent Cat # 5067-1512	$362
Covaris microtube snap-cap (25 tubes)	Covaris Cat #520045	$125
T4 DNA ligase buffer with 10mM ATP		
T4 DNA polymerase (3U/ul)	NEB Cat # M2030	
Klenow DNA polumerase (5U/ul)	NEB Cat # M0210	
T4 PNK (10U/ul)	NEB Cat # M0201	
NEB buffer 2	NEB	
Klenow fragment 3’ to 5’ exo – (5U/ul)	NEB Cat # M0212	
2x Rapid Ligation Buffer	Enzymatics B101	
T4 DNA ligase (600U/ul)	Enzymatics 12 2012	
Illumina Paired-end adapter Oligo Mix	Illumina – got aliquot from Gingeras Lab	
Uracil N-Glycoslyase (UNG) AmpErase	ABI N8080096	$143
PE primer 1.0	Order HPLC purified from IDT	
PE primer 2.0	Order HPLC purified from IDT	
2x HF phusion mix	Finzymes	

C.3 Histone Chromatin immunoprecipitation protocol

Histone ChIP protocol

Sample fixation:

1. Pre-warm glycine to ensure it is correctly resuspended prior to starting.
2. Harvest cells and resuspend to 1 x 10^6 cells/ml in D-PBS. 1 x 10cm plate of NPCs will yield approximately 3 x 10^6 cells.
3. Add 27µl of 37% formaldehyde per 1ml of cell suspension (final 1%). Mix by gentle inversion of the falcon tube.
4. Incubate exactly 5 minutes at room temperature
5. Add 200µl of 2.5M glycine per 1ml of cell suspension. Mix by gentle inversion of the falcon tube.
6. Incubate 5 minutes at room temperature
7. Spin down 3000g for 10 minutes at 4°C
8. Resuspend in ice-cold TBS.
9. Aliquot 5x10^6 cells per 1.5ml eppendorf tube. Each sample will be sufficient for 2 ChIP plus input control.
10. Spin down, 3000g for 3 minutes at 4°C
11. Reuspend and incubate samples in 1ml Lysis buffer + protease inhibitors added, 10 minutes on ice
12. Spin down 1000g for 3 minutes at 4°C
13. Wash nuclei in 0.5ml Lysis buffer + protease inhibitors
14. Snap freeze pellet in liquid nitrogen. Store samples at -80°C for up to 6 months.

Antibody+Bead preparation:

1. Resuspend M-280 sheep anti-mouse or anti-rabbit beads (Invitrogen) and aliquot 40µl per IP (approximately 2x10^7 beads).
2. Add 400µl 0.5% BSA, 1x PBS per sample
3. Place on magnetic stand for 1 minute. Remove supernatant
4. Resuspend in 0.5ml 0.5% BSA, 1x PBS per sample. Incubate 30 minutes on rotator at 4°C.

5. Add antibodies or 5µl normal mouse/rabbit IgG (3mg/ml) in 500µl RIPA-150, 0.5% BSA.
6. Incubate 4-6 hours at 4°C with rotation.

Chromatin immunoprecipitation:

7. Fill sonicator bath with ice slurry to pre-cool machine at least 30 minutes prior to starting.
8. Resuspend nuclei in 500µl Dilution Buffer I + protease inhibitors
9. Incubate 5 minutes on ice
10. Sonicate 20 cycles on maxium setting 30 seconds ON, 30 seconds OFF. Optimal setting depends on the cell type and fixation conditions and needs to be optimized. Use polystyrene 15ml falcon tubes (BD 352095) for sonication. Probes extend into the tube. Add a small amount of ice every 4 cycles to ensure samples stay cold.
11. Spin down samples in 1.5ml low binding RNAse, DNAse free 1.5ml tubes maximum speed, 10 minutes at 4°C.
12. During chromatin centrifugation step, wash antibody bound beads twice in 1ml RIPA-150, 0.5% BSA.
13. Collect chromatin supernatant. Add 500µl Dilution Buffer II and 500µl RIPA-150 + protease inhibitors per sample.
14. Add 500µl chromatin supernatant per IP. Make sure you keep 200µl chromatin aside as input control. Keep input at 4°C until required.
15. Incubate chromatin with antibody bound beads overnight (at least 16 hours) at 4°C with rotation
16. Wash beads 2x in 1ml RIPA-150 + 0.5x protease inhibitors
17. Wash beads 2x in 1ml RIPA-500 + 0.1mM PMSF
18. Wash beads 1x in 1ml TE
19. Elute in 100µl elution buffer (TE, 1% SDS) for 15 minutes at 65°C with shaking
20. Collect supernatant
21. Repeat elution with another 100µl elution buffer for 15 minutes at 65°C with shaking

22. Pool eluted samples (total 200µl).
23. Add 200µl post-elution buffer to each sample (total 400µl)
24. To the input, add 160µl H_2O, 18µl 5M NaCl, 3.6µl 0.5M EDTA, 18µl 10% SDS
25. Incubate samples and input overnight at 65°C in waterbath. Ensure tubes are sealed by wrapping with parafilm
26. Add 2µl of 10mg/ml RNAse per sample (final 50µg/ml). Incubate 37°C for 30 minutes
27. Add 5µl of 20ml/ml proteinase K (finale 250µg/ml). Incubate 55°C for 1 hour.
28. Clean-up either by phenol-chloroform extraction or using Qiagen PCR clean-up columns.
29. Final volume 50µl. Use 2µl for Q-PCR
30. Check chromatin smear by running 2-5µl of input on 2% agarose gel. Smear should be between 200-400bp.

Lysis Buffer:

stock	final concentration	10ml
1M Tris-HCl pH8.0	10mM	100µl
5M NaCl	10mM	20µl
10% NP40	0.5%	0.5ml
H_2O		9.4ml

Dilution Buffer I

stock	final concentration	5ml
1M Tris-HCl pH8.0	50mM	250µl
0.5M EDTA	2mM	20µl
20% SDS	0.2%	50µl
5M NaCl	134mM	134µl
20% Triton X-100	0.88%	220µl
10% Na-Deoxycholate	0.088%	44µl
H_2O		4.3ml

Dilution Buffer II

stock	final concentration	5ml
1M Tris-HCl pH8.0	50mM	250µl
5M NaCl	167mM	167µl
20% Triton X-100	1.1%	275µl
10% Na-Deoxycholate	0.11%	55µl
H_2O		4.25ml

Elution Buffer

stock	final concentration	5ml
1M Tris-HCl pH8.0	10mM	50µl
0.5M EDTA	1mM	10µl
10% SDS	1%	500µl
H_2O		4.44ml

Post-Elution Buffer

stock	final concentration	5ml
1M Tris-HCl pH8.0	10mM	50µl
0.5M EDTA	9mM	90µl
5M NaCl	600mM	600µl
H_2O		4.26ml

RIPA Buffer

stock	100ml RIPA-150	100ml RIPA-500
1M Tris-HCl pH8.0	5ml	5ml
5M NaCl	3ml	10ml
0.5M EDTA	200µl	200µl
10% SDS	1ml	1ml
20% triton X-100	5ml	5ml
10% Na-deoxycholate	1ml	1ml
H_2O	up to 100ml	up to 100ml

C.4 RNA polymerase Chromatin immunoprecipitation protocol

RNA polymerase II ChIP protocol

Sample fixation:

1. Pre-warm glycine to ensure it is correctly resuspended prior to starting.
2. Harvest cells and resuspend to 1 x 10^6 cells/ml in D-PBS. 1 x 10cm plate of NPCs will yield approximately 3 x 10^6 cells.
3. Add 27µl of 37% formaldehyde per 1ml of cell suspension (final 1%). Mix by gentle inversion of the falcon tube.
4. Incubate exactly 5 minutes at room temperature
5. Add 200µl of 2.5M glycine per 1ml of cell suspension. Mix by gentle inversion of the falcon tube.
6. Incubate 5 minutes at room temperature
7. Spin down 3000g for 10 minutes at 4°C
8. Resuspend in ice-cold TBS.
9. Aliquot 5x10^6 cells per 1.5ml eppendorf tube. Each sample will be sufficient for 2 ChIP plus input control.
10. Spin down, 3000g for 3 minutes at 4°C
11. Reuspend and incubate samples in 1ml Lysis buffer + protease inhibitors added, 10 minutes on ice
12. Spin down 1000g for 3 minutes at 4°C
13. Wash nuclei in 0.5ml Lysis buffer + protease inhibitors
14. Snap freeze pellet in liquid nitrogen. Store samples at -80°C for up to 6 months.

Antibody+Bead preparation:

1. Resuspend M-280 sheep anti-mouse beads (Invitrogen) and aliquot 40µl per IP (approximately 2x10^7 beads).
2. Add 400µl 0.5% BSA, 1x PBS per sample
3. Place on magnetic stand for 1 minute. Remove supernatant
4. Resuspend in 0.5ml 0.5% BSA, 1x PBS per sample. Incubate 30 minutes on rotator at 4°C.

5. Add 5µl mouse α RNApolII antibody (abcam 817) or 5µl normal mouse IgG (3mg/ml) in 500µl RIPA-150, 0.5% BSA.
6. Incubate 4-6 hours at 4°C with rotation.

Chromatin immunoprecipitation:

7. Fill sonicator bath with ice slurry to pre-cool machine at least 30 minutes prior to starting.
8. Resuspend nuclei in 500µl Sonication buffer + protease inhibitors
9. Incubate 5 minutes on ice
10. Sonicate 16 cycles on maximum setting 30 seconds ON, 30 seconds OFF. Optimal setting depends on the cell type and fixation conditions and needs to be optimized. Use polystyrene 15ml falcon tubes (BD 352095) for sonication. Probes extend into the tube. Add a small amount of ice every 4 cycles to ensure samples stay cold.
11. Spin down samples in 1.5ml low binding RNAse, DNAse free 1.5ml tubes maximum speed, 10 minutes at 4°C.
12. During chromatin centrifugation step, wash antibody bound beads twice in 1ml RIPA-150, 0.5% BSA.
13. Collect chromatin supernatant. Add 1ml Sonication buffer + protease inhibitors per sample.
14. Add 500µl chromatin supernatant per IP. Make sure you keep 200µl chromatin aside as input control. Keep input at 4°C until required.
15. Incubate chromatin with antibody bound beads overnight (at least 16 hours) at 4°C with rotation
16. Wash beads 3x in 1ml Sonication Buffer + protease inhibitors
17. Wash beads 1x in 1ml High Salt Sonication Buffer + 0.1mM PMSF
18. Wash beads 1x in 1ml LiCl buffer
19. Wash beads 1x in 1ml TE
20. Elute in 100µl elution buffer (TE, 1% SDS) for 15 minutes at 65°C with shaking
21. Collect supernatant
22. Repeat elution with 100µl elution buffer for 15 minutes at 65°C with shaking

23. Pool eluted samples (total 200µl).
24. Add 200µl post-elution buffer to each sample (total 400µl)
25. To the input, add 160µl H_2O, 18µl 5M NaCl, 3.6µl 0.5M EDTA, 18µl 10% SDS
26. Incubate samples and input overnight at 65°C in water bath. Ensure tubes are sealed by wrapping with parafilm
27. Add 2µl of 10mg/ml RNAse per sample (final 50µg/ml). Incubate 37°C, 30 min
28. Add 5µl of 20ml/ml proteinase K (finale 250µg/ml). Incubate 55°C for 1 hour.
29. Clean-up either by phenol-chloroform extraction or using Qiagen PCR clean-up columns.
30. Final volume 50µl. Use 2µl for Q-PCR
31. Check chromatin smear by running 2-5µl of input on 2% agarose gel. Smear should be between 200-400bp.

Lysis Buffer:

stock	final concentration	10ml
1M Tris-HCl pH8.0	10mM	100µl
5M NaCl	10mM	20µl
10% NP40	0.5%	0.5ml
H_2O		9.4ml

Sonication Buffer

stock	final concentration	5ml
1M Tris-HCl pH7.5	50mM	250µl
5M NaCl	140mM	140µl
0.5M EDTA	1mM	10µl
0.2M EGTA	1mM	25µl
20% Triton X-100	1%	250µl
10% Na-deoxycholate	0.1%	50µl
20% SDS	0.1%	250µl
H_2O		4.25ml

High Salt Sonication Buffer

stock	final concentration	5ml
1M Tris-HCl pH7.5	50mM	250µl
5M NaCl	500mM	500µl
0.5M EDTA	1mM	10µl
0.2M EGTA	1mM	25µl
20% Triton X-100	1%	250µl
10% Na-deoxycholate	0.1%	50µl
20% SDS	0.1%	250µl
H_2O		3.66ml

LiCl buffer

stock	final concentration	5ml
1M Tris-HCl pH8.0	20mM	100µl
0.5M EDTA	1mM	10µl
5M LiCl	250mM	250µl
10% NP-40	0.5%	500µl
10% Na-deoxycholate	0.5%	500µl
H_2O		3.64ml

Elution Buffer

stock	final concentration	5ml
1M Tris-HCl pH8.0	50mM	250µl
0.5M EDTA pH8.0	10mM	100µl
20% SDS	1%	250µl
H_2O		4.4ml

C.5 ChIP-sequencing protocol

ChIP-Seq protocol

Sample fixation:

1. Pre-warm glycine to ensure it is correctly resuspended prior to starting.
2. Harvest cells and resuspend to 1 x 10^6 cells/ml in D-PBS. 1 x 10cm plate of NPCs will yield approximately 3 x 10^6 cells.
3. Add 27µl of 37% formaldehyde per 1ml of cell suspension (final 1%). Mix by gentle inversion of the falcon tube.
4. Incubate exactly 5 minutes at room temperature
5. Add 200µl of 2.5M glycine per 1ml of cell suspension. Mix by gentle inversion of the falcon tube.
6. Incubate 5 minutes at room temperature
7. Spin down 3000g for 10 minutes at 4°C
8. Resuspend in ice-cold TBS.
9. Aliquot 5x10^6 cells per 1.5ml eppendorf tube. Each sample will be sufficient for 2 ChIP plus input control.
10. Spin down, 3000g for 3 minutes at 4°C
11. Reuspend and incubate samples in 1ml Lysis buffer + protease inhibitors added, 10 minutes on ice
12. Spin down 1000g for 3 minutes at 4°C
13. Wash nuclei in 0.5ml Lysis buffer + protease inhibitors
14. Snap freeze pellet in liquid nitrogen. Store samples at -80°C for up to 6 months.

Antibody+Bead preparation:

1. Resuspend M-280 sheep anti-mouse or anti-rabbit beads (Invitrogen) and aliquot 80µl per IP (approximately 4x10^7 beads).
2. Add 400µl 0.5% BSA, 1x PBS per sample
3. Place on magnetic stand for 1 minute. Remove supernatant
4. Resuspend in 0.5ml 0.5% BSA, 1x PBS per sample. Incubate 30 minutes on rotator at 4°C.
5. Add 4mg antibodies or in 500µl RIPA-150, 0.5% BSA.

6. Incubate 4-6 hours at 4°C with rotation.

Chromatin immunoprecipitation:

7. Fill sonicator bath with ice slurry to pre-cool machine at least 30 minutes prior to starting.
8. Resuspend nuclei in 500µl Dilution Buffer I + protease inhibitors. Will need 2 tubes each owith $5x10^6$ cells per IP.
9. Incubate 5 minutes on ice
10. Sonicate 24 cycles on maxium setting 30 seconds ON, 30 seconds OFF. Optimal setting depends on the cell type and fixation conditions and needs to be optimized. Use polystyrene 15ml falcon tubes (BD 352095) for sonication. Probes extend into the tube. Add a small amount of ice every 4 cycles to ensure samples stay cold.
11. Spin down samples in 1.5ml low binding RNAse, DNAse free 1.5ml tubes maximum speed, 10 minutes at 4°C.
12. During chromatin centrifugation step, wash antibody bound beads twice in 1ml RIPA-150, 0.5% BSA.
13. Collect chromatin supernatant. Add 500µl Dilution Buffer II and 500µl RIPA-150 + protease inhibitors per sample.
14. Add 2x1.5ml chromatin supernatant per IP. You will have two tubes per IP.
15. Incubate chromatin with antibody bound beads overnight (at least 16 hours) at 4°C with rotation
16. Wash beads 2x in 1ml RIPA-150 + 0.5x protease inhibitors
17. Wash beads 2x in 1ml RIPA-500 + 0.1mM PMSF
18. Wash beads 1x in 1ml TE
19. Elute in 100µl elution buffer (TE, 1% SDS) for 15 minutes at 65°C with shaking. Pool the two tubes together with the first 100µl wash.
20. Collect supernatant
21. Repeat elution with another 100µl elution buffer for 15 minutes at 65°C with shaking
22. Pool eluted samples (total 200µl).

23. Add 200µl post-elution buffer to each sample (total 400µl)
24. Incubate samples and input overnight at 65°C in waterbath. Ensure tubes are sealed by wrapping with parafilm
25. Add 2µl of 10mg/ml RNAse per sample (final 50µg/ml). Incubate 37°C for 30 minutes
26. Add 5µl of 20ml/ml proteinase K (finale 250µg/ml). Incubate 55°C for 1 hour.
27. Clean-up either by phenol-chloroform extraction or using Qiagen PCR clean-up columns.
28. Final volume 30µl.
29. Check chromatin smear by running High Sensitivity DNA chip (Bioanalyser). Will need 5-10ng DNA within 200-600bp range.
30. Add 31µl resuspension buffer from Illumina TruSeq mRNA preparation kit. All reagents from here on come from this kit.
31. Perform end repair. Add 40µl End Repair Mix. Incubate 30 minutes at 30°C
32. Clean-up using 140µl AmpureX beads. Resuspend in 17.5µl Resuspension buffer
33. Perform A-tail reaction. Add 12.5µl End Repair Mix. Incubate 37°C for 30 min.
34. Ligate adapters. Add 2.5µl resuspension buffer, 2.5µl ligation mix and 2.5µl adapter. Incubate 10 minutes at 30°C
35. Add 5µl stop ligation buffer. Bring the 42.5µl to a total volume of 100µl by adding 57.5µl H_2O.
36. Clean-up with Qiagen PCR purification columns. Elute 30µl elution buffer
37. Run on 2% agarose gel for 2 hours at 90V. Ensure there is a spare lane between samples.
38. Gel purify band at 200bp range. Collect 300bp gel piece as back-up. Elute 30µl elution buffer.
39. PCR amplify with 7.5µl primer cocktail and 37.5µl PCR mix. 15 cycles.
40. Clean-up with Ampure beads adding 75µl to 75µl PCR reaction
41. Resuspend in a total 22.5µl resuspension buffer
42. Transfer 20µl to a new tube.
43. Run 1:10 and 1:20 dilutions on High Sensitivity DNA chip (Agilent bioanalyser). Peak of library should be around 300bp. Dilute to 10nM and sequence (SR36).

Lysis Buffer:

stock	final concentration	10ml
1M Tris-HCl pH8.0	10mM	100µl
5M NaCl	10mM	20µl
10% NP40	0.5%	0.5ml
H_2O		9.4ml

Dilution Buffer I

stock	final concentration	5ml
1M Tris-HCl pH8.0	50mM	250µl
0.5M EDTA	2mM	20µl
20% SDS	0.2%	50µl
5M NaCl	134mM	134µl
20% Triton X-100	0.88%	220µl
10% Na-Deoxycholate	0.088%	44µl
H_2O		4.3ml

Dilution Buffer II

stock	final concentration	5ml
1M Tris-HCl pH8.0	50mM	250µl
5M NaCl	167mM	167µl
20% Triton X-100	1.1%	275µl
10% Na-Deoxycholate	0.11%	55µl
H_2O		4.25ml

Elution Buffer

stock	final concentration	5ml
1M Tris-HCl pH8.0	10mM	50µl
0.5M EDTA	1mM	10µl
10% SDS	1%	500µl

H_2O	4.44ml

Post-Elution Buffer

stock	final concentration	5ml
1M Tris-HCl pH8.0	10mM	50µl
0.5M EDTA	9mM	90µl
5M NaCl	600mM	600µl
H_2O		4.26ml

RIPA Buffer

stock	100ml RIPA-150	100ml RIPA-500
1M Tris-HCl pH8.0	5ml	5ml
5M NaCl	3ml	10ml
0.5M EDTA	200µl	200µl
10% SDS	1ml	1ml
20% triton X-100	5ml	5ml
10% Na-deoxycholate	1ml	1ml
H_2O	up to 100ml	up to 100ml

C.6 Nick Translation protocol

Nick Translation Protocol

Reagents from Nick Translation Kit (Abbott Molecular Cat. 32-801300 or homemade reagents)
Labeled dUTP from Enzo life sciences

1. Make reaction mixture:
    - 22-xµl water
    - xµl DNA (2µg total)
    - 2.5µl 0.2mM labeled dUTP
    - 5µl 0.1mM dTTP
    - 10µl dNTP mix
    - 5µl 10x nick translation buffer
2. mix well, add 5µl nick translation enzyme

3. PCR reaction:
    - 15°C 10 hours
    - 70°C 10 min
    - hold at 4°C

4. transfer to 1.5ml eppendorf and add:
    - 1µl 0.5M EDTA
    - 1µl linear acrylamide
    - 5µl 3M NaOAc
    - 125µl 100% EtOH (ice cold)

5. precipitate at -20°C over night or -80°C 2 hours

6. centrifuge max speed 1 hr 4°C

7. remove supernatant

8. wash 2x with 1ml 75% EtOH

9. air dry pellet 15 min 37°C incubator

10. resuspend 50ul H2O. Place on shaking 37°C heat block to completely resuspended.

11. Run 5ul on 2% agarose gel. Smear pattern should be between 50-400nt.

12. Store at -20°C in dark. Use 3-5ul per FISH reaction

C.7 RNA-DNA FISH protocol

DAY 1

RNA FISH Hybridization:

1) Fix cells in 2% PFA for 20 min at room temperature

2) Wash in 1x PBS, 3 x 10min.

3) Permeablize cells in 1xPBS/0.5% Triton X-100/5mM VRC, 5 minutes on ice.

4) Wash in 1x PBS, 3 x 10min

5) Equilibrate in 2x SSC, 10 min @ room temperature

6) Probe preparation:
    a. Place 2ul probe (~40ng), 5ul each Cot1 DNA, yeast tRNA, and ssDNA in a tube, and dehydrate in Speed-Vac. (~20min)
    b. Meanwhile, prepare 2x Hybridization Buffer: 4xSSC, 20% Dextran Sulfate in H2O. Mix and keep on 37°C heating block.
    c. Remove probe from Speed-Vac and add 10ul formamide to resuspend. Place on the 37°C heating block, with shaking, for at least 30 min.

7) Denature the probe for 10 min @ 90-92°C. Then, incubate on ice for at least 2 minutes.

8) Add 10ul 2X Hybridization Buffer to the denatured probe and mix well.

9) Spot 20ul probe mixture onto a clean slide, and place the coversip, cell side down, onto the probe.

10) Seal with rubber cement, and hybridize overnight at 37°C in a humid chamber.

DAY 2

Before you start:

- Turn the water bath up to 80°C. Put an empty coplin jar into the 80°C water bath to pre-heat.
- Set one heating block to 92-95°C.
- Place an empty coplin jar on ice to chill.
- Warm 50% Formamide/2X SSC solution to 37°C in a coplin jar.
- Warm some 2X SSC and 1X SSC to 37°C. (10 ml per coplin jar)

Post-hybridization washes:

1) Remove rubber cement from coverslips, and add some 2X SSC to the sides of the coverslip to help loosen it from the slide.

2) Transfer the coverslips to the coplin jar containing 50% formamide/2X SSC (pH 7.0), incubate at 37°C for 30 minutes. If a shaking water bath is available, gently agitate the coplin jars during the incubation.

2) Wash in 2X SSC, 15 minutes at 37°C.

3) Wash in 1X SSC, 15 minutes at 37°C.

4) Equilibrate the coverslips in 4X SSC, 5 minutes at room temp.

5) Incubate with DAPI stain (1:10,000 in 4X SSC) for 3 minutes at room temp.

6) Wash in 4X SSC, 2 x 5 minutes at room temp.

7) Mount coverslips on clean slides, using 15ml mounting medium. Blot excess mounting medium with a paper towel, and then seal coverslips with nail polish.

8) Image cells, be sure to save coordinates of images on microscope to return to the same cells.

DAY 3:

Coverslip Preparation:

1) Mark position of coverslip on slide with black marker. Remove nailpolish and carefully remove coverslip from slide. Use of 1x PBS can assist coverslip removal. Make a small break in one corner and mark the slide appropriately such that you can correctly orient the slide following DNA-FISH.

2) Wash slide 3 times in 1xPBS, 5 minutes each, to remove mounting medium.

3) Fix cells in 4% PFA, 10min @ room temp.

2) Wash in 1X PBS, 3 x 10 min.

3) Add 0.1mg/ml RNAse A in 1X PBS, incubate at 37°C x 1.5 hour.

4) During RNase treatment, start probe preparation, as described below.

Probe Preparation (Keep probe out of direct light):

1) For each slide to be hybridized, add the following to a 1.5 ml tube:

   5ml mouse Cot1 DNA (~10mg)
   5ml salmon sperm DNA
   5ml yeast tRNA
   2-5ml nick translated probe (40-50ng)
   Total volume: ~20ml

2) Dehydrate probe in Speed-Vac. (~30 min)

3) Add 10ml formamide to each tube to resuspend probe.

4) Place tubes in the shaking incubator @ 37°C for at least 30 minutes to resuspend.

5) Make 2X Hybridization Buffer, 10ml for each slide:
   20% Dextran Sulfate
   4X SSC
   dH2O

6) Keep 2X Hybridization Buffer in shaking incubator at 37°C until coverslips are ready for hybridization.

Denaturation of Cellular DNA Before Hybridization:

1) After RNase treatment is complete, transfer coverslips to coplin jars containing 2X SSC.

2) Wash in 2X SSC, 2 x 5 min at RT, then 5 min at 42°C.

3) Make formamide solution: 70% formamide, 2X SSC.

4) Microwave the formamide solution until just boiling. (~20 sec)

5) Transfer the hot solution to the coplin jar that is already in the 80°C water bath, and monitor the temperature of the solution in the coplin jar with a thermometer.

6) When the temperature of the solution goes down to 80°C, transfer the coverslips in to the coplin jar in the 80°C water bath.

7) Denature the coverslips in the 80°C bath for 2-10 minutes (depending on cell type).

8) Just before the denaturation step is complete, add ice-cold (-20°C) 70% ethanol to the coplin jar that was pre-chilled on ice.

9) QUICKLY transfer coverslips to the coplin jar containing 70% ethanol. Incubate on ice for 5 minutes.

10) Dehydrate the coverslips through an ethanol series: 95% ethanol x 5 minutes, then 100% ethanol for at least 5 minutes. The coverslips can stay in 100% ethanol until the probes are ready for hybridization.

Final Probe Preparation and Hybridization:

1) Just before the cells are ready for hybridization, denature the probe (resuspended in formamide) in the 92-95°C heating block for 10 minutes.

2) Transfer probes directly to ice, and incubate for at least 2 minutes.

3) Add 10ml pre-warmed 2X Hybridization Buffer to the denatured probe, mix well, and spot the probe on a clean slide. (~20ml total, final concentrations in probe: 10% dextran sulfate, 50% formamide, 2X SSC)

4) Remove the denatured coverslip from the 100% ethanol, blot with Whatman paper, and briefly air dry.

5) Turn the coverslip, cell side down, onto the probe on the slide.

6) Seal the coverslip with rubber cement, place the slides in a humid chamber, and incubate at 37°C overnight.

<u>DAY 4</u>

Before you begin:

- Warm 50% Formamide/2X SSC solution to 37°C in a coplin jar.
- Warm some 2X SSC and 1X SSC to 37°C. (10 ml per coplin jar)

Post-hybridization washes:

1) Remove rubber cement from coverslips, and add some 2X SSC to the sides of the coverslip to help loosen it from the slide.

2) Transfer the coverslips to the coplin jar containing 50% formamide/2X SSC (pH 7.0), incubate at 37°C for 30 minutes. If a shaking water bath is available, gently agitate the coplin jars during the incubation.

3) Wash in 2X SSC, 15 minutes at 37°C.

4) Wash in 1X SSC, 15 minutes at 37°C.

5) Equilibrate the coverslips in 4X SSC, 5 minutes at room temp.

6) Incubate with DAPI stain (1:10,000 in 4X SSC) for 3 minutes at room temp.

7) Wash in 4X SSC, 2 x 5 minutes at room temp.

8) Mount coverslips on original slides using marks as a guide, using 15ml mounting medium. Blot excess mounting medium with a paper towel, and then seal coverslips with nail polish.

Appendix D

Additional publications arising during this thesis

D.1 Chromatin organisation and transcriptional regulation

This review covers the advances made in the past 2-3 years regarding the structure and organisation of chromatin in the nucleus and how this relates to gene expression.

Specific contributions: MAE-M wrote sections regarding gene position in relation to chromosome territories, nuclear periphery and transcription dependent gene movement. MRH wrote sections regarding higher order chromatin structure and genome-wide chromatin interactions, and compiled the manuscript. MRH, MAE-M and DLS contributed to reviewing the literature, and editing the final manuscript.

Citation: Michael R Hübner, Melanie A Eckersley-Maslin and David L Spector. *Current Opinion in Genetics & Development* 2013 23:89-95

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Genetics & Development

Chromatin organization and transcriptional regulation

Michael R Hübner[1], Mélanie A Eckersley-Maslin[1,2] and David L Spector[1,2]

Cell type specific transcriptional regulation must be adhered to in order to maintain cell identity throughout the lifetime of an organism, yet it must be flexible enough to allow for responses to endogenous and exogenous stimuli. This regulation is mediated not only by molecular factors (e.g. cell type specific transcription factors, histone and DNA modifications), but also on the level of chromatin and genome organization. In this review we focus on recent findings that have contributed to our understanding of higher order chromatin structure and genome organization within the nucleus. We highlight new findings on the dynamic positioning of genes relative to each other, as well as to their chromosome territory and the nuclear lamina, and how the position of genes correlates with their transcriptional activity.

Addresses
[1] Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, NY 11724, United States
[2] Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, United States

Corresponding author: Spector, David L (spector@cshl.edu)

Current Opinion in Genetics & Development 2013, **23**:89–95

This review comes from a themed issue on **Genome architecture and expression**

Edited by **Genevieve Almouzni** and **Frederick Alt**

For a complete overview see the Issue and the Editorial

Available online 24th December 2012

0959-437X/$ – see front matter,

http://dx.doi.org/10.1016/j.gde.2012.11.006

Introduction

Although chromatin was first described 130 years ago [1], the organization and dynamics of chromatin in the interphase nucleus *in vivo*, and how this organization relates to transcriptional regulation, is still not fully understood. Here we review recent advances in electron microscopy and light microscopy, as well as biochemical and molecular biology approaches that have shed new light on this fundamental question in biology.

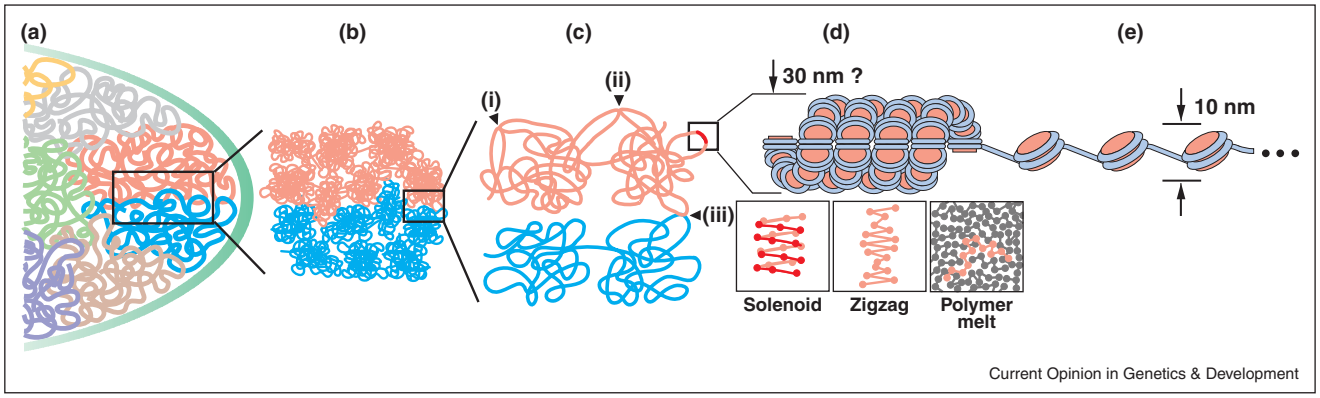
Higher order chromatin structure

DNA in the eukaryotic cell nucleus exists as a complex with histone proteins. 147 bp of DNA are wrapped in 1.7 negatively supercoiled turns around the nucleosome core particle comprised of two H3-H4 and two H2A–H2B histone dimers. Nucleosomes are separated from each other by 10–80 bp linker DNA associated with linker histone H1 (reviewed in [2]). This DNA–nucleosome complex forms a 10 nm diameter fiber resembling 'beads on a string' [3,4] (Figure 1e). The 10 nm chromatin fiber has been shown *in vitro* to form a higher order helical fiber 30 nm in diameter (Figure 1d) containing 6–11 nucleosomes per turn [5,6] which has been proposed to form even higher order chromatin fibers in interphase [7], and a 200–300 nm chromonema structure in mitotic chromosomes [8,9]. Two models have been proposed to describe the 30 nm fiber (Figure 1d). First, an interdigitated one-start solenoid structure where each nucleosome interacts with its fifth or sixth neighbor [10]. Secondly, a two-start zigzag ribbon where every second nucleosome interacts [11,12]. In a molecular tweezer experiment using 25-nucleosome repeat arrays *in vitro*, it has been determined that the extension characteristics and force of 4 pN required to fully extend the array from a 30 nm to a 10 nm fiber is consistent with a solenoid structure [13].

While it has been extensively studied *in vitro*, evidence for the existence of the 30 nm fiber *in vivo* is limited. It has been proposed that the 30 nm fiber is the preferred structure in chromatin preparations with low chromatin concentrations and low ionic strength where intra-molecular nucleosome interactions are favored over inter-molecular interactions (reviewed in [14,15]). Moreover, alcohol dehydration and embedding procedures used in electron microscopy sample preparations, as well as the 'Widom 601 nucleosome positioning' sequence used for some of these studies probably favor the formation of the 30 nm fiber *in vitro* (reviewed in [14]), all factors which call into question its existence *in vivo*.

In interphase cells, the 30 nm fiber has so far only been observed in two specialized systems: starfish spermatozoids [16], and chicken erythrocyte nuclei [16,17]. In contrast to the majority of cells, these two model systems are largely transcriptionally inactive, they contain a more highly charged histone H1 isoform, low abundance of non-histone chromatin proteins, and a longer nucleosome repeat length [18], suggesting that the 30 nm fiber might be involved in heterochromatic transcriptional repression and compaction [17]. However, this compaction may not be sufficient for transcriptional silencing, as the structure of the 30 nm fiber in avian erythrocyte nuclei is loose enough to permit the access of even large proteins to the chromatin fiber [17,19]. Interestingly, in mouse rod photoreceptor cells which have concentric areas of varying chromatin compaction, the central and most compact area shows an amorphous phase with no chromatin fibers, whereas the more peripheral layer with intermediate levels of chromatin compaction shows a 30 nm fiber, and the least condensed region shows only the 10 nm

Figure 1



Chromatin organization in the mammalian nucleus. **(a)** Chromosomes are organized in chromosome territories. **(b)** Chromosome territories are comprised of fractal globules, and fractal globules from adjacent chromosome territories can interdigitate. **(c)** Chromatin fibers interact (i) within a fractal globule (frequent), (ii) between fractal globules of the same chromosome territory (rare), or (iii) between adjacent chromosome territories (very rare). **(d)** Chromatin may form a 30 nm fiber with a solenoid zigzag, or polymer melt organization (see text). **(e)** Chromatin is resolved as a 10 nm 'beads on a string' fiber consisting of nucleosomes.

fiber [20]. This suggests that chromatin within these cells can exist in multiple distinct structures.

In order to study the decompaction and transcriptional activation of condensed chromatin from human cells that mimics *in vivo* characteristics, Reinberg and colleagues reconstituted 5 kb of DNA surrounding the RAR/RXR responsive PEPCK promoter with native histones isolated from HeLa cells, as well as histone H1, the core histone chaperone RSF, and the histone H1 chaperone NAP-1 [21]. This resulted in a highly compacted 30 nm chromatin fiber which became decondensed upon transcriptional activation. By contrast, mitotic HeLa S3 chromosomes observed in a close-to-native state by small-angle X-ray scattering and cryo-electron microscopy (cryo-EM) of vitreous sections, fail to show a higher order chromatin structure beyond the 10 nm fiber [22••,23]. Similarly, cryo-EM of rodent and plant interphase chromatin has been shown to be homogeneous and disorganized [24]. Furthermore, chromatin organization was studied by a combination of electron spectroscopic imaging and electron tomography, which does not involve contrast agents and creates a three dimensional image of chromatin *in situ* [25••]. Using this technique, open chromatin or condensed chromatin within chromocenters in mouse embryonic fibroblasts, as well as in mouse spleen lymphocytes and liver tissue cells showed the 10 nm fiber, but did not exhibit any evidence for a 30 nm or higher order chromatin organization [25••]. Therefore, rather than being ordered into a 30 nm fiber, chromatin has been described as a dynamic disordered and interdigitated state comparable with a 'polymer melt', where nucleosomes that are not linear neighbors on the DNA strand interact within a chromatin region [14,22••,23] (Figure 1d). It has been proposed that these regions represent drops of viscous fluid in which the radial position of genes within these drops may influence their transcriptional activity [14]. This fluid and irregular chromatin arrangement might permit a more dynamic and flexible organization of the genome than the rigid 30 nm fiber would provide [14,22••], and would consequently facilitate dynamic processes such as transcription, DNA replication, DNA repair and enhancer-promoter interactions [22••]. Furthermore, the irregular spacing and concentration of nucleosomes seen *in vivo* has been shown to be incompatible with the 30 nm fiber [26], further supporting the polymer melt model.

In recent years, considerable effort has been made to study chromatin in conditions that are close to the living state and an increasing amount of data suggests that chromatin organization above the 10 nm fiber probably does not exist in most mammalian cells. New super-resolution imaging techniques are promising tools to further evaluate the organization and dynamics of chromatin in living cells in the near future.

Genome-wide chromatin interactions

The development of the Chromosome Conformation Capture (3C) and 3C-related genome-wide techniques (circularized chromosome conformation capture (4C), carbon copy chromosome conformation capture (5C), Hi-C) has given us an insight into the structure and long-range interactions of chromatin at the molecular level *in vivo* (reviewed in [27,28]). In yeast, 3C analysis of transcriptionally active chromatin shows local variations in chromatin compaction, and does not support the presence of a 30 nm fiber [29]. A seminal study by Dekker and colleagues provided a model of the local chromatin environment of normal human lymphoblasts

on the megabase scale as a fractal globule, where chromatin partitions into adjacent regions with minimal interdigitation [30••] (Figure 1b), consistent with the diffusion and binding properties caused by molecular crowding of chromatin binding proteins [31,32]. The fractal globules ultimately associate on the chromosome level to form chromosome territories [30••] (Figure 1a, b), which can be observed in interphase nuclei using light microscopy techniques. In addition, the fractal globule model suggests a mechanism for the interaction of genomic sites that are distant within a chromosome or on different chromosomes, which might lead to chromosomal translocations in cancer. Interestingly, the three dimensional chromatin structure revealed by Hi-C experiments directly correlates with chromosomal rearrangements and somatic copy number variations reported in human cancer cells [33]. Similarly, the translocation frequency of the Igh and Myc loci which are located on different chromosomes in mouse B lymphocytes directly correlates to their contact frequency in a 4C-seq experiment [34]. Furthermore, the actual observed intra-and inter-chromosomal translocation frequency has been shown to correlate with the contact probability in a Hi-C experiment in G1 arrested mouse pro-B cells [35••].

Within the fractal globule, chromatin is organized into discrete domains. A Hi-C analysis in mouse ES cells identified 2200 topological domains in which chromatin with a median size of 880 kb occupying about 91% of the genome interacts locally [36••]. These topological domains are enriched in housekeeping genes and SINE elements, and are separated by topological boundary regions with characteristics of insulator elements, such as CTCF-binding and a segregation of the heterochromatic H3K9me3 mark [36••]. This organization of the topological domains is conserved between different human cell types, as well as between human and mouse [36••]. A follow up study by the same group using the ChIP-seq technique found a significant overlap of topological domains with cis-regulatory enhancer-promoter units in 19 embryonic and adult mouse tissues and cell types [37]. Similarly, a 4.5 Mb region encompassing Xist on the X chromosome in mouse ES cells was shown to partition into discrete topologically associating domains (TADs) that are 200 kb to 1 Mb in size, and are present on both the active and inactive X chromosome in male and female ES cells [38••]. While they are enriched in, they do not require H3K27me3, H3K9me2 nor lamina-associated domains (LADs) for their maintenance [38••]. Within a TAD, genes are transcriptionally co-regulated, and while the TADs as a whole do not change, the internal TAD contacts rearrange upon ES cell differentiation supporting the link between chromatin structure and transcription [38••]. Similarly, a study of the active and inactive X-chromosome in human SATO3 lymphoblast cells revealed that transcription disrupts intrachromosomal interactions, leading to local chromatin decompaction at promoters [39]. A 5C study as part of the ENCODE project analyzed the interactions of transcriptional start sites (TSS) in 44 regions representing 1% of the genome in three human cell lines [40]. More than 1000 mostly asymmetric long-range interactions with distal elements resembling promoters and enhancers were identified within these regions [40]. However, in contrast to another study [37], ~60% of the interactions were found in only one of the three cell lines analyzed indicating a cell-type specific chromatin folding [40]. Therefore, it remains to be determined how conserved these long-range interactions are between cell types or species. In addition to intra-chromosomal contacts, tethered chromosome conformation capture (TCC) experiments in human lymphoblastoid cells revealed that inter-chromosomal contacts are indiscriminate between chromosome territories and their contact probability is a function of their transcriptional activity and position within the territory [41]. Consequently, inter-chromosomal contacts are about 70 times less frequent than intra-chromosomal contacts and may be present only in a fraction of cells where both interacting regions are accessible [41] (Figure 1).

The fractal globule model has provided exciting initial insights into genome-wide short-range and long-range gene interactions involved in transcriptional regulation and chromosomal translocations in cancer. However, current 3C methodology surveys chromatin topology within dynamic populations of cells. At the single cell level, chromatin interactions are likely to be dynamic, some being stochastic, and their frequency may depend on the cell cycle and additional factors. Therefore, an examination of chromatin topology of single cells is needed to assess cell-to-cell differences as well as changes during the cell cycle and stages of differentiation in order to fully understand the relationship of gene interactions to cellular function.

Gene position in relation to chromosome territories

From the higher order fractal globule structure, chromatin is further organized into chromosome territories, where each chromosome, rather than being intertwined, occupies its own distinct region of the nucleus (reviewed in [42,43]). In order to study the contacts and interdigitation of chromosome territories, Bickmore and colleagues used fluorescently-labeled pooled sequence-capture probes to show that the exons of mouse chromosome 2 predominantly localize at the surface of the chromosome territory [44]. This is consistent with genes looping out of their chromosome territory and allows for interactions with regions of other chromosomes. Pulse-labeling experiments have revealed that only 1% of chromatin from different chromosomes co-localize in interphase cells [45]. Thus it is likely that these inter-chromosomal interactions occur transiently and/or that these are rare events, as has also been proposed by genome-wide mapping of

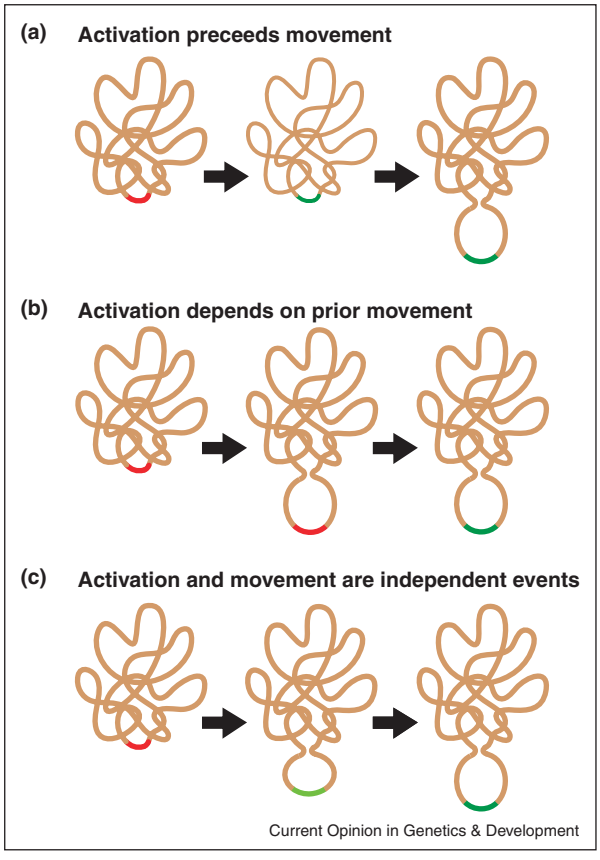
chromosome interactions [41] (Figure 1). The importance of inter-chromosomal interactions for gene regulation still remains to be elucidated, but it has been proposed that some co-regulated genes can colocalize in interchromatin granules or transcription factories [46–48]. However, it remains to be demonstrated if looping out from a chromosome territory is an active process preceding transcription, or if it is a consequence of gene activation (Figure 2). Treatment with the histone deacetylase inhibitor TSA results in increased chromatin mobility [49] and an increase in inter-chromosomal co-localization [45], suggesting that gene activation may not be a consequence of gene movement and co-localization, and that the two processes might indeed be independent from each other (Figure 2c).

Gene position in relation to the nuclear periphery

Beyond the organization of chromatin in chromosome territories, the radial position of genes within the nucleus has been implicated in gene regulation. In particular, it has been suggested that the low transcriptional activity of perinuclear heterochromatin is a consequence of nuclear lamina-mediated gene silencing [50]. The nuclear lamina which is comprised of a meshwork of type V intermediate filament proteins (lamins) and other associated proteins (reviewed in [51]) provides the interface between the inner nuclear membrane, nuclear pore complex and the nearby chromatin. Associations of large regions of chromatin, termed lamin associated domains (LADs) with the nuclear lamina is generally associated with transcriptional repression [52], however relocation to the periphery is not always sufficient for gene silencing [53], nor is it necessary as many inactive loci are located within the nucleoplasm away from the nuclear periphery. Nonetheless the association with, and disassociation of gene loci from the nuclear lamina and corresponding changes in transcriptional status, for example during embryonic stem cell differentiation [52], implicates this nuclear compartment in the regulation of gene expression.

Recent studies have advanced our understanding of how genes relocate to and from the nuclear periphery. In *S. cerevisiae* the *INO1* gene relocates to the nuclear pore complex (NPC) upon transcriptional activation [54]. This relocation is controlled by two upstream 8 bp and 20 bp DNA elements termed 'DNA zip codes' which are sufficient for relocation and clustering at the NPC [55••], suggesting that the genome itself encodes for its spatial organization. DNA elements can also mediate gene repositioning in mammalian cells. The *IgH* and *Cyp3a* loci are located within LADs that dissociate from the nuclear lamina in cell types in which these genes are actively transcribed [56]. Integration of BACs containing these genomic regions into a control locus relocates the locus to the nuclear periphery [57••]. Through a series of truncation experiments, Singh and colleagues identified a 4–6 kb minimal sequence element at these loci that is sufficient to target the surrounding DNA region to the nuclear periphery and consequently attenuate transcription of a reporter gene [57••]. This sequence element is enriched with the *GAGA* motif, which when inserted as 10 copies in a 400 bp array, is sufficient to target a DNA locus to the lamina. The sequestration at the lamina could be partially inhibited through knockdown of either the zinc finger protein cKrox, which binds the GAGA motif, or the histone deacetylase HDAC3 [57••]. Therefore, chromatin modifications, in addition to the DNA sequence elements, may also be involved in positioning genes at the nuclear periphery. This is further supported by findings implicating histone deactylases in targeting the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene to the nuclear periphery in non-expressing cells [58]. In addition to histone acetylation levels, histone H3 lysine 9 methylation has also been suggested to influence DNA positioning at the nuclear lamina in *C. elegans* embryos [59••]. An RNAi screen identified 29 factors that, when knocked-down, led to activation of a

Figure 2



Transcriptional activity influences chromatin topology. **(a)** Transcriptional activation of a gene may precede its movement within the nucleus. **(b)** An inactive gene may get activated subsequent to its movement to a site that is favorable to transcriptional activation. **(c)** Transcriptional activation and gene movement may be independent of each other. Red, inactive gene; green, active gene.

peripheral repressed reporter array. Interestingly, only 2 of these factors resulted in additional movement of the array into the nucleoplasm, demonstrating that movement is not required for gene activation. Conversely, movement of a reporter gene under an inactive promoter from the periphery was not accompanied by transcriptional activation [59••]. Therefore, while there is a correlation between gene expression levels and nuclear periphery positioning, the two processes are not necessarily dependent on each other. Additional characterization of the factors that are required for gene positioning relative to the nuclear periphery and other nuclear structures represents an interesting area for future research.

Transcription dependent gene movement

When considering gene positioning, either relative to topological associated domains, chromatin territories, or a nuclear structure such as the lamina, an important consideration is whether changes in gene position occur before or following changes in gene expression (Figure 2). For example, the HOX gene cluster in mammals change during differentiation from a single domain marked by H3K27me3 to a bimodal domain in which the active Hox genes occupy a separate region rich in H3K4me3 distinct from the inactive regions [60••]. However, it is unknown whether the structural changes that accompany gene activation are necessary for transcription to occur, or whether they are a secondary event stabilizing the gene expression program in the cells. The two alleles of the imprinted *Kcnq1* locus have recently been shown to associate in early embryogenesis at sites of high RNA polymerase II occupancy [61]. This suggests a role for gene transcription in mediating the pairing, however cause and effect again remain unknown. Similarly, the long noncoding RNAs TUG1 and MALAT1/NEAT2 have been implicated in the relocation of growth control genes between Polycomb bodies and interchromatin granule clusters [62]. Long range chromosomal interactions between the *ifrγ* cytokine gene and its receptor genes *ifrγR1* and *ifrγR2* are also associated with gene expression [63]. This interaction persists following inhibition of transcription with the RNA polymerase II inhibitor α-amanitin, implying that gene transcription is not required to maintain the intergenic interactions. However, it remains to be determined whether transcription is required for their establishment. Along these lines, chromatin looping may directly affect transcription, rather than being the result of transcriptional co-regulation. This was shown by zinc-finger mediated tethering of the GATA1 associated protein Ldb1, or merely its self-association domain, to the β-globin promoter in erythroid cells [64••]. This led to the formation of a chromatin loop between the promoter and the locus control region (LCR), and to expression of the β-globin gene in the absence of GATA1 [64••]. In order to untangle the interconnection of gene transcription and gene movement, live cell systems, in which one can follow the activation or silencing of individual endogenous genes with respect to their chromosome territory or a nuclear compartment, will be required. These types of experiments will be critical to extending our understanding of the role of nuclear organization in the regulation of gene expression.

Outlook

In recent years, light microscopy and electron microscopy approaches, as well as the emergence of genome-wide 3C-related studies have broadened our understanding of the three-dimensional organization of chromatin within the nuclear space, and how it relates to transcriptional regulation. However, many fundamental questions remain unanswered. Although increasing evidence from experiments that are close to the native chromatin state do not support the 40 year old concept of higher order chromatin structure, there is still a lack of understanding with regard to the structure of chromatin in the living cell, and whether or not a 30 nm fiber or even higher order chromatin organization exists in live interphase mammalian cells. Chromatin may have very different structures within a cell depending on multiple factors, such as the radial position within the nucleus, the cell cycle stage, the differentiation state of the cell, transcriptional activity, nucleosome occupancy, DNA and histone modifications, histone variants, long-range chromatin interactions, or any combination of these factors.

Although 3C-related techniques have provided significant insight into genome-wide chromatin association frequencies within a population of cells, these techniques currently do not tell us how dynamic such interactions are in and among single cells. It remains to be determined what the frequency and duration of these interactions are, how they relate to the cell cycle and differentiation, and if they are the cause or consequence of transcriptional regulation.

While recent advances in imaging and molecular approaches have provided significant insights into chromatin organization and gene interactions, ongoing studies examining individual living and fixed cells will provide the basis for further advances.

Acknowledgements

We thank the members of the Spector lab for helpful discussions, Megan Bodnar and Cinthya Zepeda-Mendoza for critically reading the manuscript and James Duffy for help with preparing the figures. Research in the Spector lab is supported by grants from NIGMS 42694, NCI 5P01CA013106-40, and NCI 2P30CA45508-24.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Flemming W: *Zellsubstanz, Kern und Zelltheilung*. Leipzig: F. C. W. Vogel; 1882.

2. Felsenfeld G, Groudine M: **Controlling the double helix**. *Nature* 2003, **421**:448-453.

3. Olins AL, Olins DE: **Spheroid chromatin units (v bodies)**. *Science* 1974, **183**:330-332.

4. Woodcock CF: **Ultrastructure of inactive chromatin**. *J Cell Biol* 1973, **59**:368a.

5. Finch JT, Klug A: **Solenoidal model for superstructure in chromatin**. *Proc Natl Acad Sci USA* 1976, **73**:1897-1901.

6. Gerchman SE, Ramakrishnan V: **Chromatin higher-order structure studied by neutron scattering and scanning transmission electron microscopy**. *Proc Natl Acad Sci USA* 1987, **84**:7802-7806.

7. Belmont AS, Bruce K: **Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure**. *J Cell Biol* 1994, **127**:287-302.

8. Rattner JB, Lin CC: **Radial loops and helical coils coexist in metaphase chromosomes**. *Cell* 1985, **42**:291-296.

9. Widom J, Klug A: **Structure of the 300A chromatin filament: X-ray diffraction from oriented samples**. *Cell* 1985, **43**:207-213.

10. Robinson PJ, Fairall L, Huynh VA, Rhodes D: **EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure**. *Proc Natl Acad Sci USA* 2006, **103**:6506-6511.

11. Schalch T, Duda S, Sargent DF, Richmond TJ: **X-ray structure of a tetranucleosome and its implications for the chromatin fibre**. *Nature* 2005, **436**:138-141.

12. Dorigo B, Schalch T, Kulangara A, Duda S, Schroeder RR, Richmond TJ: **Nucleosome arrays reveal the two-start organization of the chromatin fiber**. *Science* 2004, **306**: 1571-1573.

13. Kruithof M, Chien FT, Routh A, Logie C, Rhodes D, van Noort J: **Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber**. *Nat Struct Mol Biol* 2009, **16**:534-540.

14. Maeshima K, Hihara S, Eltsov M: **Chromatin structure: does the 30-nm fibre exist in vivo?** *Curr Opin Cell Biol* 2010, **22**:291-297.

15. Bian Q, Belmont AS: **Revisiting higher-order and large-scale chromatin organization**. *Curr Opin Cell Biol* 2012, **24**:359-366.

16. Woodcock CL: **Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length**. *J Cell Biol* 1994, **125**:11-19.

17. Scheffer MP, Eltsov M, Frangakis AS: **Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei**. *Proc Natl Acad Sci USA* 2011, **108**:16992-16997.

18. Grigoryev SA, Woodcock CL: **Chromatin organization – the 30 nm fiber**. *Exp Cell Res* 2012, **318**:1448-1455.

19. Makde RD, England JR, Yennawar HP, Tan S: **Structure of RCC1 chromatin factor bound to the nucleosome core particle**. *Nature* 2010, **467**:562-566.

20. Kizilyaprak C, Spehner D, Devys D, Schultz P: **In vivo chromatin organization of mouse rod photoreceptors correlates with histone modifications**. *PLoS ONE* 2010, **5**:e11039.

21. Li G, Margueron R, Hu G, Stokes D, Wang YH, Reinberg D: **Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo**. *Mol Cell* 2010, **38**:41-53.

22. •• Nishino Y, Eltsov M, Joti Y, Ito K, Takata H, Takahashi Y, Hihara S, Frangakis AS, Imamoto N, Ishikawa T *et al.*: **Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure**. *EMBO J* 2012, **31**:1644-1653.
This study shows, by cryo-electron microscopy and synchrotron X-ray scattering, that HeLa mitotic chromosomes are not organized above the 10 nm fiber.

23. Eltsov M, Maclellan KM, Maeshima K, Frangakis AS, Dubochet J: **Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ**. *Proc Natl Acad Sci USA* 2008, **105**: 19732-19737.

24. Bouchet-Marquis C, Dubochet J, Fakan S: **Cryoelectron microscopy of vitrified sections: a new challenge for the analysis of functional nuclear architecture**. *Histochem Cell Biol* 2006, **125**:43-51.

25. •• Fussner E, Strauss M, Djuric U, Li R, Ahmed K, Hart M, Ellis J, Bazett-Jones DP: **Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres**. *EMBO Rep* 2012, **13**:992-996.
Open and condensed chromatin in mouse somatic cells does not exhibit a 30 nm fiber when observed via electron spectroscopic imaging and tomography approaches.

26. Diesinger PM, Heermann DW: **Depletion effects massively change chromatin properties and influence genome folding**. *Biophys J* 2009, **97**:2146-2153.

27. Sanyal A, Bau D, Marti-Renom MA, Dekker J: **Chromatin globules: a common motif of higher order chromosome structure?** *Curr Opin Cell Biol* 2011, **23**:325-331.

28. de Wit E, de Laat W: **A decade of 3C technologies: insights into nuclear organization**. *Genes Dev* 2012, **26**:11-24.

29. Dekker J: **Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction**. *J Biol Chem* 2008, **283**:34532-34540.

30. •• Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO *et al.*: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome**. *Science* 2009, **326**:289-293.
This seminal paper identifies, by Hi-C experiments the fractal globule organization of chromosome territories in human cells.

31. Bancaud A, Huet S, Daigle N, Mozziconacci J, Beaudouin J, Ellenberg J: **Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin**. *EMBO J* 2009, **28**:3785-3798.

32. Bancaud A, Lavelle C, Huet S, Ellenberg J: **A fractal model for nuclear organization: current evidence and biological implications**. *Nucleic Acids Res* 2012, **40**:8783-8792.

33. Fudenberg G, Getz G, Meyerson M, Mirny LA: **High order chromatin architecture shapes the landscape of chromosomal alterations in cancer**. *Nat Biotechnol* 2011, **29**:1109-1113.

34. Hakim O, Resch W, Yamane A, Klein I, Kieffer-Kwon KR, Jankovic M, Oliveira T, Bothmer A, Voss TC, Ansarah-Sobrinho C *et al.*: **DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes**. *Nature* 2012, **484**:69-74.

35. •• Zhang Y, McCord RP, Ho YJ, Lajoie BR, Hildebrand DG, Simon AC, Becker MS, Alt FW, Dekker J: **Spatial organization of the mouse genome and its role in recurrent chromosomal translocations**. *Cell* 2012, **148**:908-921.
A Hi-C analysis of mouse pro-B cells reveals that intrachromosomal and interchromosomal translocations are linked to the spatial organization and contact probabilities of chromosomes.

36. •• Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: **Topological domains in mammalian genomes identified by analysis of chromatin interactions**. *Nature* 2012, **485**:376-380.
This study identifies local chromatin interaction domains termed 'topological domains' which are conserved across cell types and species, and their boundaries have characteristics of insulator elements.

37. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B: **A map of the cis-regulatory sequences in the mouse genome**. *Nature* 2012, **488**:116-120.

38. •• Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J *et al.*: **Spatial partitioning of the regulatory landscape of the X-inactivation centre**. *Nature* 2012, **485**:381-385.
5C analysis identified 'topologically associated domains' (TADs) on the mouse X-chromosome, and shows that TADs dynamically associate with 'lamin associated domains' (LADs) during differentiation.

39. Naughton C, Sproul D, Hamilton C, Gilbert N: **Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures**. *Mol Cell* 2010, **40**:397-409.

40. Sanyal A, Lajoie BR, Jain G, Dekker J: **The long-range interaction landscape of gene promoters**. *Nature* 2012, **489**:109-113.

41. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L: **Genome architectures revealed by tethered chromosome conformation capture and population-based modeling**. *Nat Biotechnol* 2012, **30**:90-98.

42. Cremer T, Cremer C: **Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present**. *Eur J Histochem* 2006, **50**:223-272.

43. Cremer T, Cremer C: **Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories**. *Eur J Histochem* 2006, **50**:161-176.

44. Boyle S, Rodesch MJ, Halvensleben HA, Jeddeloh JA, Bickmore WA: **Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis**. *Chromosome Res* 2011, **19**:901-909.

45. Olivares-Chauvet P, Fennessy D, Jackson DA, Maya-Mendoza A: **Innate structure of DNA foci restricts the mixing of DNA from different chromosome territories**. *PLoS ONE* 2011, **6**:e27527.

46. Hu Q, Kwon YS, Nunez E, Cardamone MD, Hutt KR, Ohgi KA, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG *et al.*: **Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules**. *Proc Natl Acad Sci USA* 2008, **105**:19199-19204.

47. Brown JM, Green J, das Neves RP, Wallace HA, Smith AJ, Hughes J, Gray N, Taylor S, Wood WG, Higgs DR *et al.*: **Association between active genes occurs at nuclear speckles and is modulated by chromatin environment**. *J Cell Biol* 2008, **182**:1083-1097.

48. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS *et al.*: **Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells**. *Nat Genet* 2010, **42**:53-61.

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

50. Reddy KL, Zullo JM, Bertolino E, Singh H: **Transcriptional repression mediated by repositioning of genes to the nuclear lamina**. *Nature* 2008, **452**:243-247.

51. Dechat T, Adam SA, Taimen P, Shimi T, Goldman RD: **Nuclear lamins**. *Cold Spring Harb Perspect Biol* 2010, **2**:a000547.

52. Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Graf S, Flicek P, Kerkhoven RM, van Lohuizen M *et al.*: **Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation**. *Mol Cell* 2010, **38**:603-613.

53. Kumaran RI, Spector DL: **A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence**. *J Cell Biol* 2008, **180**:51-65.

54. Brickner JH, Walter P: **Gene recruitment of the activated INO1 locus to the nuclear membrane**. *PLoS Biol* 2004, **2**:e342.

55. •• Brickner DG, Ahmed S, Meldi L, Thompson A, Light W, Young M, Hickman TL, Chu F, Fabre E, Brickner JH: **Transcription factor binding to a DNA zip code controls interchromosomal clustering at the nuclear periphery**. *Dev Cell* 2012, **22**:1234-1246.
DNA sequences were identified in *S. cerevisiae* and *S. pombe* that are sufficient for targeting genes to the nuclear pore complex (NPC) during transcriptional induction.

56. Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, Singh H: **Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development**. *Science* 2002, **296**:158-162.

57. •• Zullo JM, Demarco IA, Pique-Regi R, Gaffney DJ, Epstein CB, Spooner CJ, Luperchio TR, Bernstein BE, Pritchard JK, Reddy KL *et al.*: **DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina**. *Cell* 2012, **149**:1474-1487.
In mouse cells, lamin associated domains mediate the post-mitotic association of genes with the nuclear lamina which leads to their transcriptional silencing.

58. Muck JS, Kandasamy K, Englmann A, Gunther M, Zink D: **Perinuclear positioning of the inactive human cystic fibrosis gene depends on CTCF. A-type lamins and an active histone deacetylase**. *J Cell Biochem* 2012, **113**:2607-2621.

59. •• Towbin BD, Gonzalez-Aguilera C, Sack R, Gaidatzis D, Kalck V, Meister P, Askjaer P, Gasser SM: **Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery**. *Cell* 2012, **150**:934-947.
Methylation of H3K9 is involved in the autonomous and self-reinforcing perinuclear anchoring of heterochromatin in *C. elegans*.

60. •• Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D: **The dynamic architecture of Hox gene clusters**. *Science* 2011, **334**:222-225.
The collinear activation of Hox genes during mouse development coincides with a relocalization of these genes from a transcriptionally inactive domain rich in H3K27me3 to an active domain rich in H3K4me3.

61. Krueger C, King MR, Krueger F, Branco MR, Osborne CS, Niakan KK, Higgins MJ, Reik W: **Pairing of homologous regions in the mouse genome is associated with transcription but not imprinting status**. *PLoS ONE* 2012, **7**:e38983.

62. Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Grinstein JD, Dorrestein PC, Rosenfeld MG: **ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs**. *Cell* 2011, **147**:773-788.

63. Deligianni C, Spilianakis CG: **Long-range genomic interactions epigenetically regulate the expression of a cytokine receptor**. *EMBO Rep* 2012, **13**:819-826.

64. •• Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA: **Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor**. *Cell* 2012, **149**:1233-1244.
Tethering of the GATA1 associated Ldb1 looping factor to the β-globin promoter in mouse erythroid cells leads to the formation of a chromatin loop with the locus control region (LCR) and is sufficient for transcriptional activation of the β-globin locus.