Defining the chromatin structure of the human genome using size-selected nucleosome mapping.

Janet C. Harwood

Cardiff School of Biosciences,

Cardiff University.

PhD

2015

Acknowledgements.

I would like to thank my primary supervisor Dr. Nick Kent for preparing biological samples for my project, for his general advice, patience and sense of humour throughout this project. I would like to thank the rest of my thesis committee; Prof. Nick Allen for preparing biological samples and general advice and Prof Pete Kille for computing support and general advice.

I would like to thank Heather Stevens for her enthusiasm for this project and The Waterloo Foundation for funding it.

I owe a debt of gratitude to all those who helped me with computing support and advice; to Steffan Adams and Kevin Munn from the School of Biosciences and to Christine Kitchen, Thomas Green and Wayne Lawrence from the Advanced Research Computing team at Cardiff University (ARCCA).

I would like to thank the Daphne Jackson Trust for all their encouragement to undertake a PhD and I would like to thank Sarah Black and Helen Woodfield from Cardiff School of Biosciences for advice, encouragement and for proofreading.

Finally I would like to thank Prof. Adrian Harwood from the Neuroscience and Mental Health Research Institute and Cardiff School of Biosciences for securing funding for this project, for sharing ideas, for his advice and for his unending enthusiasm and support.

Dedication

To my family and friends whose support, encouragement and sense of humour I could not be without.

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed …… (candidate) Date ……

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of ……(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed …… (candidate) Date ……

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references. The views expressed are my own.

Signed …… (candidate) Date ……

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …… (candidate) Date ……

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed …… (candidate) Date ……

Summary

The work in this thesis examines genome-wide and local changes in the patterns of nucleosome positioning throughout the human genome. Nucleosomes are the fundamental repeating unit of chromatin. Their properties and positioning in the genome dictate whether and how proteins involved in gene regulation can access DNA. Nucleosomes are dynamic; their positions can vary considerably at some loci from one cell type to another. Chromatin remodelling complexes can change the structure and the positions of nucleosomes. Their mis-regulation leads to congenital defects affecting pre-natal and early childhood development and is associated with neuro-psychiatric disorders. As mutations in genes that encode chromatin remodelling proteins are associated with human mental health disorders, the work in this thesis focusses on changes that occur in chromatin structure during early neural development.

I have used MNase-seq data to construct genome-wide, high-resolution chromatin particle positioning maps from undifferentiated human induced pluripotent stem cells (hiPSC) and following differentiation to the neuro-progenitor cell (NPC) stage. These maps reveal that a small proportion of the pluripotent genome possesses well-positioned nucleosomes, the number of which increases approximately 8-fold during neural cell development. This is accompanied by changes in the distribution and localisation of nucleosomes between iPS and NPC cells.

Differences in nucleosome positioning during neural cell differentiation were investigated at regulatory regions. Loss and gain of positioned nucleosomes at TSS of pluripotent and neural-specific genes was detected and correlated with gene expression. In addition I investigated the chromatin structure at the binding motifs of two important genome regulators REST and CTCF in detail. Nucleosome positioning is maintained at REST binding motifs during neural cell development. In contrast, at CTCF sites nucleosome repositioning occurs during neural cell development. This work provides insight into the role of chromatin structure in the regulation of human neural cell differentiation.

Contents

Figures

Tables.

Chapter 1 : Introduction

1.1 Overview

Eukaryotic DNA is packaged into the relatively small nucleus of the cell by coiling the DNA around core histone proteins to form nucleosomes. These coils are packaged into higher order 3D structures that ultimately form condensed chromosomes. This complex of nuclear nucleic acids and proteins is known as chromatin. We know that the genome is regulated and that timing of events is important, so how can the genes in this packaged DNA be accessible for the cellular machinery to utilise them and does the organisation of the genome affect the regulation of events in a cell?

It is well known that chromatin structure is dynamic at both a local and a global level. The action of ATP-dependent chromatin remodelling complexes can bring about local changes to nucleosome structure by post-translational modifications of histones or by the exchange of histone variants. Chromatin remodellers also can slide nucleosomes, leading to changes in their positions. It is thought that nucleosome properties and positioning in the genome can dictate accessibility to DNA; hence changes in the structure and positions of nucleosomes can affect gene regulation. Higher-order chromatin structure is also important in the regulation of the genome. It has been shown that chromatin is organised into dynamic functional domains and that this organisation is mediated by architectural proteins such as CTCF. Post-translational modifications of nucleosomes have been studied widely, but the regulation of nucleosome positioning and chromatin structure is less understood. Recent studies have shown that chromatin structure is both organism and cell-type specific. In addition, clinical genetic observations implicate mutations in chromatin remodelling genes in childhood neuro-developmental syndromes and mental health disorders.

Deciphering developmentally associated changes in chromatin structure will provide insight into how chromatin regulation interfaces with human neuro-development and disease. Therefore, the work in this thesis focusses on the changes that occur in chromatin structure during early neural development. In order to do this, the genome- wide positions of nucleosomes in the human genome were mapped at high resolution using MNase-seq. I have constructed and characterised high-resolution chromatin particle positioning maps from human induced pluripotent stem cells (hiPSC) in their undifferentiated state and following differentiation to the neuro-progenitor cell (NPC) stage. Initially I investigated the global patterning of nucleosome positioning in both iPS and NPC. The differences in the distribution and localisation of nucleosomes between iPS and NPC cells was determined, to a) to identify the locations of organised chromatin across entire human genomes and b) to determine the regions in which changes in chromatin structure occur during human neural cell

differentiation. Subsequently, changes in nucleosome positioning during early neural cell differentiation were investigated at regulatory regions relevant to transcription regulation, for example at transcriptional start sites and at the binding motifs of a selection of transcription factors relevant to neural cell development. A detailed analysis is presented for REST which is a transcriptional repressor of neural genes in non-neuronal cells. Finally, in order to investigate whether or not architectural proteins, that affect higher-order chromatin structure, have a role in the regulation of neural cell development, changes in nucleosome positioning were investigated at CTCF binding motifs.

1.2 The human genome

The human genome is large, comprising approximately 3x 10^9 bp of DNA (Lander, Linton et al. 2001). The human genome is organised into 24 pairs of chromosomes, 22 autosomes and two sex chromosomes. Current estimates are that the total number of genes is 60,483, of which 19,814 are protein coding genes, 15,900 are long non-coding RNAs and 9894 are small non-coding RNAs (GENCODE Release (version 22) www.gencodegenes.org). Human protein coding genes vary greatly in size, from several kilobases to several megabases, for example the dystrophin gene is 2.4 Mb (Koenig, Monaco et al. 1988) and they are therefore, along with their regulatory units, dispersed over long distances in the genome. Approximately 50-60% percent of the human genome comprises tandem repeats (de Koning, Gu et al. 2011), known as satellite DNA which is associated with functional centromeres (Verdaasdonk and Bloom 2011) and telomeres (Rosenberg, Hui et al. 1997) and with inactive chromatin .

The major recent important discovery in the work on the human genome is that although only 1% of the human genome possesses protein coding genes, most of the genome is transcribed. Hence much of this transcription generates non-coding RNA. These non-coding RNAs are variable in size, may be sense or antisense transcripts and may overlap. Their functions are a current active area of research.

1.2.1 The non-coding human genome

Recent work has shown the importance of non-coding RNAs in genome regulation (Geisler and Coller 2013), therefore the analysis of chromatin structure in both the coding and non-coding genomes is important. Micro RNAs (miRNAs) are 20-24 nucleotide RNAs that bind to target motifs in messenger RNAs, silencing them (Lim, Lau et al. 2005). They are involved in neurogenesis (Stappert, Borghese et al. 2013), specifically, miR-9 is up-regulated in neural differentiation and is thought to down-regulate REST (Conaco, Otto et al. 2006) (Packer, Xing et al. 2008; Laneve, Gioia et al. 2010).

Long non-coding RNAs (lncRNAs) are involved in the regulation of the epigenome by association with chromatin modifying complexes (Khalil, Guttman et al. 2009). For example, the xist gene encodes a lncRNA that is involved in mammalian X chromosome inactivation (Chow, Yen et al. 2005). In addition and relevant here is that lncRNAs are important in the maintenance of pluripotency (Sheik Mohamed, Gaughwin et al. 2010), (Guttman, Donaghey et al. 2011) and that they can affect gene expression through association with chromatin-modifying complexes (Tsai, Manor et al. 2010).

1.3 Primary chromatin structure

Chromatin was discovered by Walter Fleming in 1878 (Flemming 1965; Paweletz 2001) who observed that thread-like structures in the cell nucleus absorbed aniline dyes (Fig 1.1). In 1963 Luzzati *et al* showed that chromatin had a repeating pattern, with intervals of 100 angstroms, that differed from that of the structure of DNA (Luzzati and Nicolauieff 1963) and images of repeating units in chromatin fibres were shown by electron microscopy by Olins *et al* (Olins and Olins 1974). Further evidence for repeating units in chromatin was shown by digestion of chromatin by micrococcal nuclease (Burgoyne, Hewish et al. 1974; Noll 1974). These repeating units in chromatin were named 'nucleosomes' and shown to flow from lysed nuclei as 10nm fibre structures that resemble beads on a string (Olins and Olins 1974; Oudet, Gross-Bellard et al. 1975; Olins, Carlson et al. 1976). The protein component of a nucleosome comprises histone proteins. Kornberg *et al* isolated histone proteins and studied them in combination with DNA by X ray diffraction and demonstrated that histone H3 and H4 exist as a heterotetramer (Thomas and Kornberg 1975). Later it was demonstrated that histones H2A,H2B, H3 and H4 exist in equal amounts (Olins, Carlson et al. 1976) and that these proteins form an octomer (Thomas and Kornberg 1975; Thomas and Butler 1977) Hence nucleosomes were described as a bead-like structure of 12.5nm diameter that encompass 200 bp of DNA in approximately one negative super-helical turn per bead (Oudet, Germond et al. 1978).

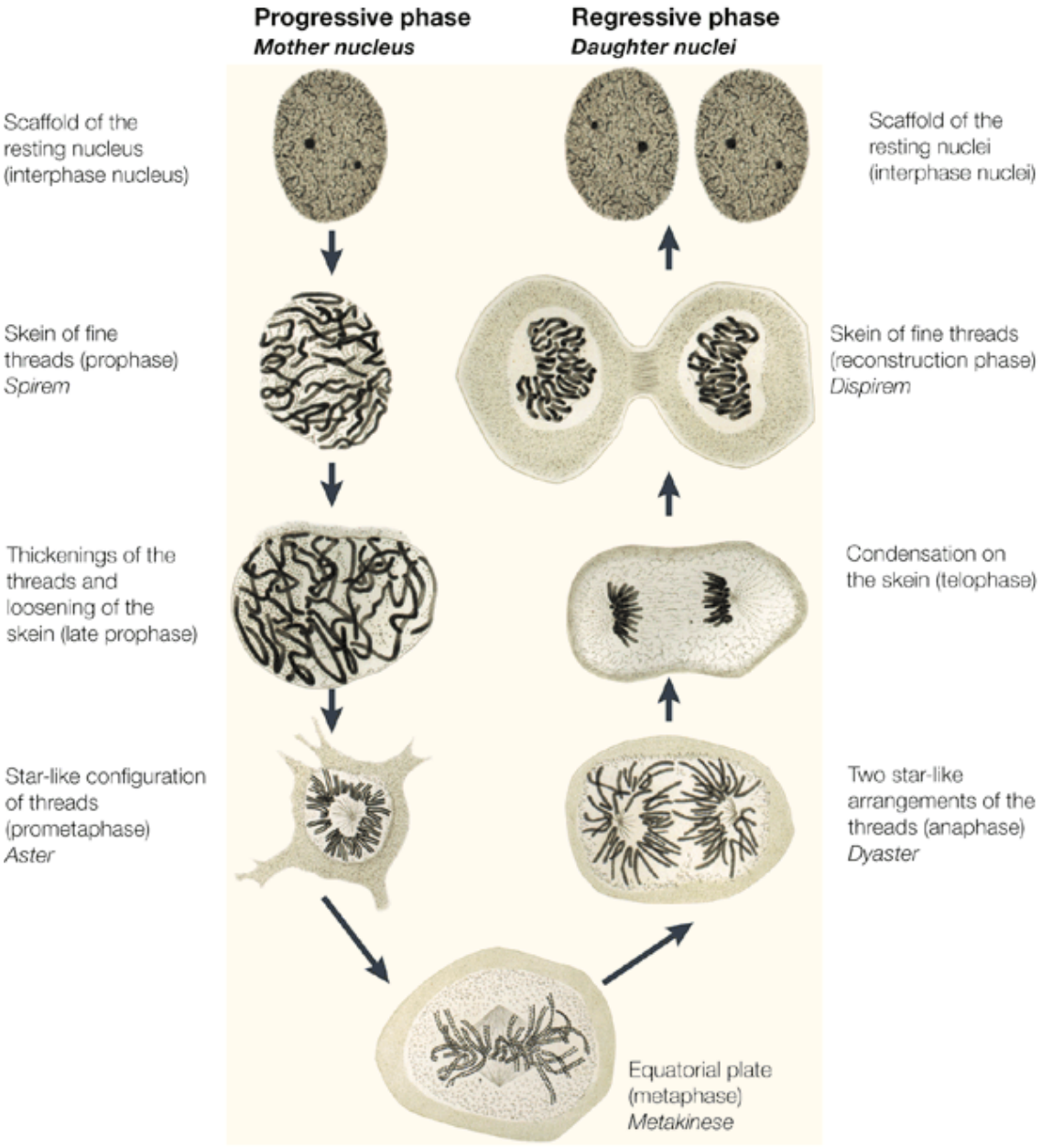


Figure 1.1 Early illustrations of chromatin.

Detection of chromatin by Walter Flemming from 'Zellsubstanz, Kern und Zelltheilung' (Cell substance, nucleus and cell division) (Flemming 1882), reproduced by Paweletz (Paweletz 2001) showing staining of cells with aniline dyes, detecting chromatin structure during different phases of the cell cycle. Reprinted by permission from Macmillan Publishers Ltd: Nature reviews. Molecular cell biology **2**(1): 72-75.Copyright (2001).

More recently the high resolution crystal structure of the nucleosome particle demonstrated that the histone octamer comprises four dimers, two H3-H4 and two H2A-H2B. The H3-H4 dimers form a heterotetramer with which the H2A-H2B dimers interact on opposite sides (fig 1.2). This histone octomer forms the nucleosome core and is wrapped by 146 bp of DNA, comprising 1.65 turns of a left-handed super-helix (Luger, Mader et al. 1997). The centre of the nucleosome where there is an overall pseudo two-fold symmetry is defined as the nucleosome dyad.

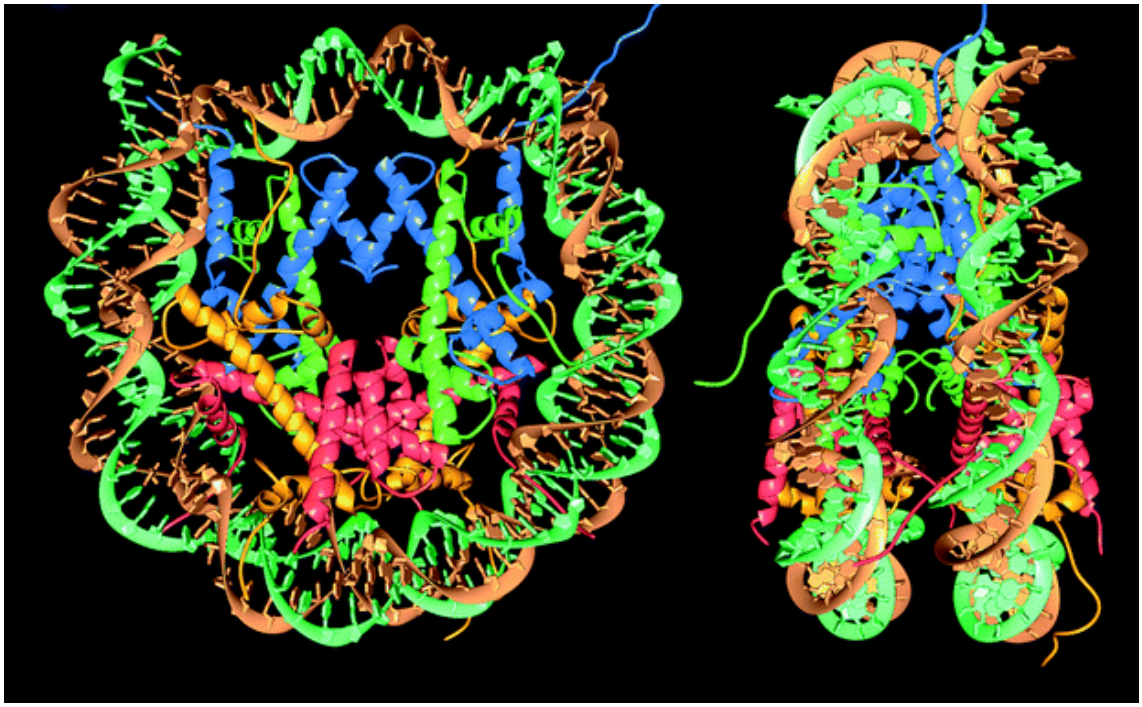


Figure 1.2 Crystal structure of the nucleosome core particle.

The nucleosome core particle comprises an octamer of histone proteins. The histone octamer comprises four dimers, two H3-H4 and two H2A-H2B. The H3-H4 dimers form a heterotetramer with which the H2A-H2B dimers interact on opposite sides. This histone octomer forms the nucleosome core and is wrapped by 146 bp of DNA, comprising 1.65 turns of a left-handed super-helix. Reprinted by permission from Macmillan Publishers Ltd: [Nature] (Luger, Mader et al. 1997) copyright (1997).

Core nucleosomes are joined together by stretches of DNA which are approximately 20-80 bp long, defined as non-nucleosomal DNA and known as linker DNA, to form the 10nm fibre or 'beads on a string' primary chromatin structure. In higher eukaryotes linker DNA may be bound by linker histones, for example histones H1 or H5 (in avian species). It has been shown that linker DNA can vary in length both in different organisms, between cell types from the same organism and at different stages of differentiation.

The nucleosome repeat length (NRL) is given by the distance from the dyad of one nucleosome to the next. Since core canonical nucleosomes in higher eukaryotes are similar in structure, protecting 147 bp of DNA, variation in nucleosome repeat length is attributed to changes in linker length. The NRL can vary from one organism to another; the *S. cerevisiae* genome has a nucleosome repeat length (NRL) of 165 bp (Yuan, Liu et al. 2005), whereas in *C. elegans* it is 175 bp (Valouev, Ichikawa et al. 2008) and early work in human HeLa cells suggested that it is approximately 200 bp (Whitlock and Simpson 1976). There is a wide variation in linker length between different cell types in higher eukaryotes (Compton, Bellard et al. 1976) and recent work has shown that the human nucleosome repeat length is variable from one cell type to another; 193 bp in granulocytes, 203 bp in CD4+ cells (Valouev, Johnson et al. 2011) and 187 bp in lymphoblastoid cells (Gaffney, McVicker et al. 2012). This led to the idea that changes in linker length, resulting in changes in chromatin accessibility may have a role cell differentiation.

In higher eukaryotes, the addition of the linker histone, H1 to the nucleosome core forms 'chromatosomes' (Simpson 1978). H1 is positioned at the dyad axis of the nucleosome core and protects DNA both on entry and exit to the nucleosome core, stabilising core nucleosomes (Allan, Hartman et al. 1980). Linker histones are many and varied. In humans there are eleven H1 variants :the somatic variants H1.1 to H1.5 that are expressed in most cells and six tissue-specific H1 variants, these are encoded by two classes of H1 genes (Happel and Doenecke 2009).

Histone H1 binding to chromatin is dynamic, (Lever, Th'ng et al. 2000) (Misteli, Gunjan et al. 2000). Linker histones are thought to increase chromatin compaction, leading to transcriptional repression, hence H1 is depleted in active chromatin (Fan, Nikitina et al. 2005). Linker histones can act to block the access of chromatin remodelling complexes and affect nucleosome mobility (Hill and Imbalzano 2000) (Ramachandran, Omar et al. 2003) Linker histones can be post-translationally modified, for example by phosphorylation (Garcia, Busby et al. 2004), reversing chromatin compaction and allowing the access of chromatin remodelling complexes (Horn, Carruthers et al. 2002).

In addition, linker histones are involved in the regulation of developmental genes (Nguyen, Gokhan et al. 2014). Recently it has been shown that post-translational modification of H1 by citrullination affects its binding to DNA and that this is important in pluripotency (Christophorou, Castelo-Branco et al. 2014).

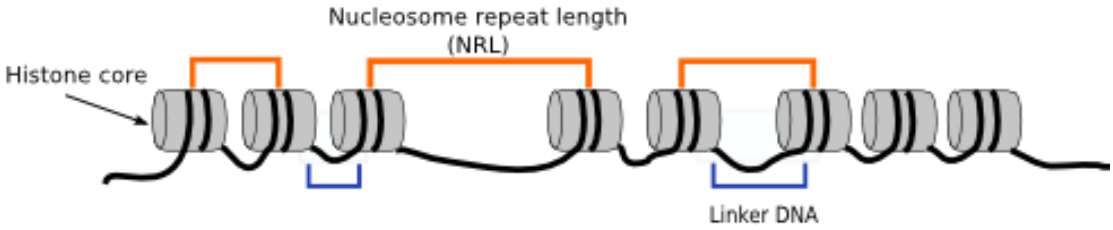


Figure 1.3 Nucleosome Repeat length (NRL).

The nucleosome repeat length is given by the distance from the dyad of one nucleosome to the next. Linker DNA is the length of DNA between nucleosomes. Core canonical nucleosomes in higher eukaryotes protect 147 bp of DNA therefore variation in nucleosome repeat length is attributed to changes in linker length.

1.4 Higher-order chromatin structure

The formation of nucleosomes is thought to be the first step in packaging DNA into the nucleus of a cell in an ordered manner. How the higher order packaging of eukaryotic genomes occurs has been the subject of some controversy and is not completely understood.

1.4.1 The 30nm fibre

The early models of how nucleosomes might be packaged into higher order structures were the solenoid model (Finch and Klug 1976) and the zig-zag model (Woodcock, Frado et al. 1984; Burgoyne 1985) which were derived from *in vitro* studies of chromatin. In the solenoid model, the nucleosomes assemble face-to-face with their nearest neighbour and the linker region is bent, packaging six nucleosomes per turn of the 30nm fibre. Thoma *et al* show evidence for these structures and that H1 is required for this higher order packaging of nucleosomes. (Thoma and Koller 1977; Thoma, Koller et al. 1979).

The zig-zag model proposes that nucleosomes associate with their opposite neighbour and the linker is not bent, hence the characteristics of the structure depends on the linker length and would be more flexible. This model is supported by work from Rydberg et al (Rydberg, Holley et al. 1998).

The early *in vitro* work and models of the 30nm fibre led to the idea that chromatin was generally 'packaged' into this higher order structure when it was inactive. However, more recent work has challenged this idea and suggested that the 30nm fibre can exist, but it is by no means the only way in which primary chromatin is packaged. There is evidence that the 10nm fibre can fold into many forms (Engelhardt 2007), generating many different types of secondary chromatin structure. Cryo-eM, which allows studies on native chromatin in living cells in a hydrated state (Dubochet, Adrian et al. 1988), have shown conflicting results for detection and characterization of the 30nm fibre (Dubochet, Adrian et al. 1988; Maeshima, Hihara et al. 2010; Scheffer, Eltsov et al. 2011) and its existence has been challenged. Eltsov *et al* concluded that there was no evidence for the 30nm fibre in mitotic chromosomes (Eltsov, Maclellan et al. 2008). However, recent Cryo-EM studies have shown that its structure is consistent with a zigzag two-start helix, but the structure differs from that in the original model. (Song, Chen et al. 2014). Taken together, it seems that the 30nm fibre can exist *in vivo*, but it may not exist all the time, or in all of the nucleus. Hence it is not, as previously thought, the dominant form of secondary chromatin structure.

1.4.2 The polymer-melt structure

It has been suggested that that the 10nm fibre exists as a dynamic polymer-melt like structure (Nishino, Eltsov et al. 2012), whereby the 10nm fibres are arranged in a loose, irregular association, rather than coiled into rigid secondary structures. Evidence for this fluid chromatin fibre structure in living cells has been shown recently along with evidence that nucleosomes can move locally, suggesting that chromatin is packaged more loosely than proposed in the original 30nm fibre models (Nozaki, Kaizu et al. 2013).

In this case, *in vivo*, where the relative concentration of nucleosomes is high, localised structures may form through inter-nucleosome fibre interactions. This model would allow the access of large protein complexes necessary for cellular processes to occur, by constant movement and rearrangement of the structure locally (Hihara, Pack et al. 2012).

1.4.3 Fractal globules

Fractal globules are the result of the condensation of a polymer that has topological constraints. The idea that chromatin could form fractal globules (Mirny 2011) is consistent with results from work using chromatin conformation capture techniques (de Wit and de Laat 2012; Sajan and Hawkins 2012; Dekker, Marti-Renom et al. 2013) to study long-range chromatin interactions. These techniques use crosslinking of chromatin, followed by digestion and re-ligation and sequencing to determine which regions of chromatin are in close proximity to one another in the cell. One such technique, Hi-C (Lieberman-Aiden, van Berkum et al. 2009), can be used to look at genome-wide chromatin interactions and the variability in cell to cell chromatin structure is averaged across many cells. More recently, single-cell Hi-C (Nagano, Lubling et al. 2013) has been used to construct a 3D picture of how chromosomes are arranged in single cell.

1.4.4 Chromatin domains

The fractal model of chromatin would mean that the 10nm fibres would not cross one another, could form loops that span large chromosomal domains and have a territorial organisation. This structure could easily unfold, which would be important in allowing access for gene activation. The chromatin loops may be anchored by crosslinking proteins to keep them in place. Territories could extend as far as whole chromosomes, which would occupy a particular space, but could still interact with other chromosomal domains.

Evidence that some of these conditions may be met in this model has come from a large body of recent research in to the 3D structure of chromatin. Interphase chromosomes have been shown occupy distinct territories within the nucleus using spectral karyotyping (SKY)

techniques (Bolzer, Kreth et al. 2005) (Cremer and Cremer 2010) . Within chromosome territories, there are two types of compartments that exist along the chromosome; one of transcriptionally active chromatin (heterochromatin), and one of inactive chromatin (euchromatin). The compartments can interact and they are tissue-specific (Lieberman-Aiden, van Berkum et al. 2009). These compartments are up to 10Mb in size and correlate with the size of a fractal globule, although the relationship between the fractal globule and these chromatin compartments is not yet clear.

Within chromatin compartments, chromatin forms smaller, approximately 500kb loops which may be active or inactive. These loops are known as topologically active domains (TADs) and they were characterised using chromatin conformation capture techniques. Single cell Hi C experiments have shown that these chromatin domains are not randomly placed (Nagano, Lubling et al. 2013). Active chromatin domains tend to be located at the periphery of a chromosome and to associate with other active domains, which may be on other chromosomes. These topological domains are conserved across species, stable across cell types and they can be functionally organised, so that groups of genes are clustered together to perform a particular task and their boundaries are enriched in CTCF (Dixon, Selvaraj et al. 2012). For example, clusters of replication origins that are in the same place tend to fire at the same time, hence these clusters are known as 'replication domains' (Hiratani, Ryba et al. 2008).

Within TADS, there is evidence for the existence of sub-TADs. These comprise a gene and its regulatory elements, brought together by the formation sub-megabase topological domains. These interactions may be anchored by cohesin, separated by boundary elements which are genetically encoded and which tend to be bound by the insulator protein CTCF and may be bound by mediator. Boundary sites are sometimes, but not always co-occupied by the Med12 subunit of the mediator complex, CTCF and Smc, a subunit of the cohesin complex. The combination of proteins found at the boundary elements appear to correlate with the size of the sub-TAD. The overlap of all three of these proteins occurs at boundary sites in ES cells at a high level, suggesting their involvement in cell-type-specific looping.

It has been shown that these sub-TADs bring together tissue-specific enhancers and promoters and may therefore be involved in the regulation of tissue-specific gene expression. (Phillips-Cremins and Corces 2013). Another aspect of global spatial and structural organisation of the genome that may impact on genome regulation is the existence of lamina associated domains (LADs). The nuclear lamina associates with LADs which are transcriptionally repressed. Recent

work has shown that this demarcation of the genome is important in the regulation of replication timing (Pope, Ryba et al. 2014).

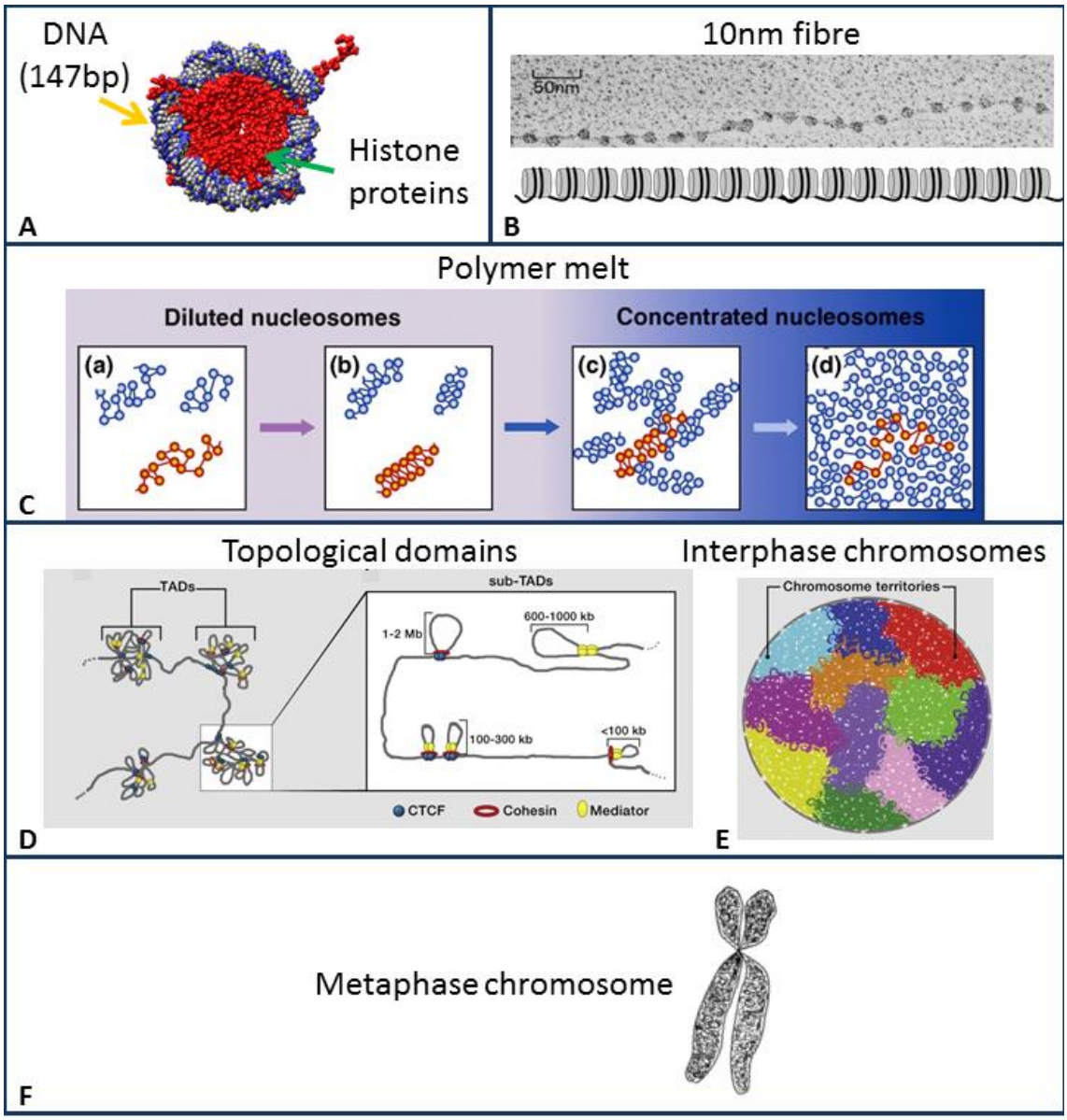


Figure 1.4 Current models of primary and higher-order chromatin structure.

A. The canonical core nucleosome comprises an octamer of histone proteins around which 147 bp of DNA is wrapped (diagram courtesy of Dr. N. Kent). **B.** The top panel shows an electron micrograph of nucleosomes existing as a 'beads on a string' structure which measures approximately 10nm in diameter (courtesy of Victoria Foe). The bottom panel is a cartoon representation of the 10nm fibre. **C**. The polymer melt model for higher order folding of nucleosome fibres. Reprinted from Current opinion in cell biology, **22**(3): 291-297, Maeshima, K., S. Hihara, et al. "Chromatin structure: does the 30-nm fibre exist in vivo?" Copyright (2010), with permission from Elsevier. **D.** Chromatin exists in topologically active domains (TADs) and sub-TADS. **E**. Interphase chromosomes form 'territories' in the nucleus that can interact. D. and E. Reprinted from"Chromatin meets its organizers." Bodnar, M. S. and D. L. Spector Cell **153**(6): 1187-1189 (2013) with permission from Elsevier. **F.** Cartoon of a metaphase chromosome in which chromatin is condensed.

1.5 Chromatin remodelling

The regulated alteration of chromatin structure is known as 'chromatin remodelling', this occurs in two major ways; by alteration of the structure of the core histone proteins or by repositioning nucleosomes. Alteration of the structure of the nucleosome core occurs through post translational modification of the core histone proteins (Strahl and Allis 2000) or by exchange of histone variants (Kamakaka and Biggins 2005). Nucleosome repositioning is brought about by the action of ATP-dependent chromatin remodelling complexes (Clapier and Cairns 2009). Chromatin remodellers are nuclear enzymes that are part of a larger complex and use the energy derived from ATP hydrolysis to slide, evict, eject, or restructure nucleosomes by loosening the DNA around the histone core (Havas, Whitehouse et al. 2001).

1.6 Covalent histone modifications

Histone proteins are basic proteins that have an N-terminal or C-terminal tail. They have a common three dimensional structural motif, known as the histone fold that is found in all eukaryotic histone proteins and is involved in the formation of histone dimers.(Arents and Moudrianakis 1995). The N-terminal tails of histone proteins can protrude from a nucleosome and can contact neighbouring nucleosomes (Luger, Mader et al. 1997). Histone tails are rich in lysine and arginine and the resulting positive charge facilitates their association with negatively charged DNA in forming chromatin. Histone proteins can be post-translationally modified. There are many different types of histone modification, but some of the more characterised ones are methylation (Lachner and Jenuwein 2002), acetylation (Kimura, Matsubara et al. 2005), phosphorylation, ubiquitination sumoylation.

1.6.1 Histone acetylation

Histone acetylases (HATs) modify lysine residues by acetylation and this process is reversible through the action of histone deacetylases (HDACs). Histone acetylation can have a role in transcriptional activation (Allfrey, Faulkner et al. 1964). Histone acetyl transferases (HATs) catalyse the acetylation of conserved lysine amino acids on histone proteins resulting in the activatory H3K9ac and H3K27ac chromatin marks (Berndsen and Denu 2008). Conversely, through the action of the HDACs brings about histone deacetylation, generating repressive histone marks.

More recently it has been shown that that histone acetylation has a role in the regulation of DNA replication (Unnikrishnan, Gafken et al. 2010) and in the repair of double strand breaks (Chen, Carson et al. 2008) in yeast. In addition, histone acetylation may have a direct effect on chromatin structure. Acetylation of histone lysine residues neutralises the charge of the

histone, weakening its interaction with DNA. *In vitro*, this can result in a looser chromatin structure (Hong, Schroth et al. 1993) (Tse, Sera et al. 1998) and may be a mechanism that allows greater access of protein complexes, such as transcription factors, involved in cellular processes, to DNA (Lee, Hayes et al. 1993). It has been shown that acetylated lysine residues are recognised by the bromo-domains in chromatin re-modelling complexes (Kasten, Szerlong et al. 2004), perhaps using these histone marks as a 'docking station'. In addition, it has been shown that co-activator complexes necessary for the activation of transcription can have histone acetylation activity (Kuo, Zhou et al. 1998). Recently it has been shown that modification of lysines that are inside the nucleosome core can destabilise nucleosomes (Di Cerbo, Mohn et al. 2014), providing another level at which chromatin structure alteration can occur.

1.6.2 Histone methylation

There are two types of histone methyl-transferases; those that catalyse the transfer of methyl groups to lysine residues and those that transfer methyl groups to arginine residues on histone H3 and H4 proteins. Mono, di or tri methylation can occur on one lysine side chain, generating a large number of permutations of histone modification. Histone lysines can be mono, di or tri-methylated through the action of histone methyl transferases. These modifications can be reversed through the action of histone demethylases. Certain methylated histone modifications have been correlated with transcriptional activation e.g. H3K4me and H3K36me and others with transcriptional repression e.g. H3K9me and H3K27me. Recently it has been shown that H3K4me3 is associated with enhancers (Pekowska, Benoukraf et al. 2011). In addition, some methylated histone residues are involved recruiting chromatin remodelling complexes, for example in yeast**,** H3K36me3 is involved in recruiting the chromatin re-modelling complex Isw1b. (Maltby, Martin et al. 2012).

However, the current picture of the role of methylation in genome regulation is not simple, methylated histone marks are dynamic and they can change during cell differentiation in human cells (Pan, Tian et al. 2007) It may be that combinations of histone modifications act in concert with the proteins recruited to regulatory regions to regulate the genome.

1.7 Histone Variants

Histone variants differ in their amino acid sequence from the major histone proteins (H1, H2A, H2B, H3 and H4). They can replace the major histone proteins normally found in canonical nucleosomes by incorporation into chromatin in a replication-dependent or replication-independent manner. Replication-independent incorporation of histone variants is brought about by ATP-dependent chromatin remodelling complexes (Mizuguchi, Shen et al. 2004) (Papamichos-Chronakis, Watanabe et al. 2011) or chaperone proteins (Obri, Ouararhni et al. 2014)

Some histone variants are thought to alter the structural properties of nucleosomes and others have been shown to occur in specific places in the genome (Kamakaka and Biggins 2005; Shaytan, Landsman et al. 2015). H2AZ is a histone variant that was first described as a component of pericentric heterochromatin (Rangasamy et al., 2003), but more recently has been shown to be enriched at yeast promoters (Guillemette, Bataille et al. 2005). It is highly conserved, expressed throughout the cell cycle and its incorporation into nucleosomes is replication-independent. The crystal structure of H2AZ-containing nucleosomes is similar to that of canonical nucleosomes (Suto, Clarkson et al. 2000), but recently H2AZ2.2, a splice variant of H2AZ that is found in human cells lines and highly expressed in the brain, has been shown to destabilise nucleosomes (Bonisch, Schneider et al. 2012).Recent genome-wide studies have shown that H2AZ is important in the maintenance of pluripotency. It is enriched at active enhancers and promoters in ES cells, correlates strongly with H3K4 methylation and it is localised at bivalent TSS. Its presence may make chromatin more accessible, allowing the recruitment of protein complexes that mediate gene activation or repression to promoters and enhancers, (Ku, Jaffe et al. 2012; Hu, Cui et al. 2013). In addition, H2AZ is necessary for lineage commitment in ES cells (Creyghton, Markoulaki et al. 2008).

Histone H3.3 is incorporated into nucleosomes in a replication-independent manner (Ahmad and Henikoff 2002). It has been associated with histone replacement in areas of the genome that are undergoing active transcription (Mito, Henikoff et al. 2005; Schwartz and Ahmad 2005) and recently it has been found to be in pericentric chromatin (Goldberg, Banaszynski et al. 2010). The chromo-domain-helicase-DNA-binding protein CHD1 is responsible for in the incorporation of H3.3 into nucleosomes in Drosophila (Konev, Tribus et al. 2007) and recently it has been shown that the incorporation of histone H3.3 is reduced when CHD2 is depleted in human cells (Siggens, Cordeddu et al. 2015). In addition, mutations in H3.3 are associated with paediatric brain tumours (Liu, McEachron et al. 2014).

1.8 Chromatin marks

Chemical modifications of the genome that can be made without any alteration of the DNA sequence are known as 'chromatin marks'. These modifications include a) post translational modification of histone proteins b) the exchange of histone variants and c) DNA methylation. These modifications affect nucleosome structure by affecting the strength of the interactions between the nucleosome core and the DNA wrapped around it. In addition, these modifications can affect the accessibility of other proteins, such as transcription factors, to DNA. Hence, these modifications can regulate gene activity determining how and when genes are switched on and off.

Correlation of the presence of post translational modification of histone proteins with the occurrence of a particular biological event led to the idea of the 'histone code' , whereby certain histone modifications or combinations thereof are associated with the of occurrence particular cellular processes (Strahl and Allis 2000). More recently, chromatin marks have been classified into those that may be involved in gene activation and those that may be involved in repression (Table 1.2). However, recent work has shown that the picture is not as simple as it seemed. Chromatin marks may act individually, in concert, in different regions of the genome and at different times during cell development. It has been shown that post-translational histone modifications may be cell and type specific (Heintzman, Hon et al. 2009), their levels change during the cell cycle (Schulze, Jackson et al. 2009), during cell differentiation (Hawkins, Hon et al. 2010) and in response to environmental factors (Feil and Fraga 2011). In addition they may act in concert with exchange of histone variants, but the exact role and mechanisms of histone modifications in gene regulation is not completely clear. As the number known of chromatin marks has increased, together with data from genome-wide studies which generate detailed information about the locations of these marks, it is becoming clear that the relationship between chromatin marks and genome regulation is complex.

Mark	Location	Effect on transcription
Methylated cytosine	CpG residues	repressive
H3K9me1	5'end genes, introns	
H3K9me3	Heterochromatin and repetitive elements	repressive
H3K9ac	Promoters, regulatory elements	active
H3K20me1	Introns, 5'end of genes	
H3K36me3	Introns,	
H3K79me2	Introns, 5'end of genes	
H3K4me1	Enhancers, introns, downstream of TSS	
H3K4me2	promoters and enhancers	
H3K4me3	Promoters and TSS	repressive
H3K27ac	promoters and enhancers	active
H3K27me3	promoters	repressive
H2AZ	Regulatory elements	

Table 1.1 Summary of chromatin marks, their locations and their effects on transcription.

Chromatin marks have been classified according into those that may be involved in gene activation and those that may be involved in repression. (Dunham 2012; Trynka and Raychaudhuri 2013)

1.8.1 Bivalent promoters

Covalent histone modifications appear to be important development. In embryonic stem cells, the promoters of transcription factor genes that are involved in development possess both an activatory (H3K4me3) and a repressive chromatin mark (H3K27me3) in mouse (Bernstein, Mikkelsen et al. 2006) and in humans (Zhao, Han et al. 2007) , such promoters are known as 'bivalent'. Bivalent promoters have been described as being 'poised ' to express developmental genes, but at the same time differentiation is repressed, thus maintaining pluripotency (Voigt, Tee et al. 2013) and it has been shown that bivalency is important in early neuronal development (Burney, Johnston et al. 2013).

1.8.2 Chromatin states.

Recently, the idea was proposed that rather than particular histone modifications being responsible for gene regulation, that combinations of chromatin marks in particular genomic locations may determine modes of genome regulation. These regions of particular combinations of chromatin marks in the genome are known as 'chromatin states' (Ernst and Kellis 2012). They are dynamic and reversible and can comprise any combination of chromatin marks, covalent histone modifications, methylated DNA, or histone variants which may confer active or repressive chromatin.

Fifteen chromatin states have been defined in human cells (Ernst, Kheradpour et al. 2011) , defining states associated with regulatory regions such as promoters, enhancers, insulators, transcribed regions and large-scale repressed and inactive domains. These types of analyses have led to systematic genome-wide characterisation of regulatory elements and their cell-type specificity (Ernst and Kellis 2013). Computational modelling of these chromatin states using Hidden Markov Modelling (HMM) and tools such as chromHMM, coupled with correlation of these states with gene expression data is leading to linking distinct chromatin signatures and patterns of gene expression during differentiation (Shah, Oldenburg et al. 2014).

1.9 DNA methylation

DNA may be modified by through the action of DNA methylases that methylate cytosine or adenine nucleotides in DNA (Zhang, Huang et al. 2015) (Greer, Blanco et al. 2015). DNA methylation is associated with gene repression, it is associated with X chromosome inactivation (Hellman and Chess 2007). In addition, DNA methylation may be involved in altering the structure of chromatin directly by affecting nucleosome stability and positioning, having a role in the formation of hetero-chromatin and in gene silencing (Razin 1998).

Alternatively, DNA methylation may affect chromatin structure by acting as a docking station for proteins; recruiting for example chromo and bromo-domain containing ATP-dependent chromatin remodelling complexes.

The human genome is 70-80% methylated. Methylation occurs within genes, but promoters tend to be under-methylated (Tazi and Bird 1990). Regions of the genome that are hypo-methylated are known as CpG islands. DNA methylation is mediated by the conserved family of DNA methyl-transferases (DNMTs). In mammals there are three DNMTs; DNMT1, DNMT3a and DNMT3b. *De novo* methylation occurs during embryogenesis, mediated by DNMT3a and DNMT3b (Okano, Bell et al. 1999). DNMT1 is responsible for maintaining the methylation state by binding to hemi-methylated DNA at CpG sites after DNA replication. DNMT1 is essential for viability in human ESC cells and loss of DNMT1 results in a global loss of DNA methylation (Liao, Karnik et al. 2015) suggesting an important role for methylation in human genome regulation.

Much of the work on the role of methylation during cell development been done in mouse models. This has shown that global changes in methylation occur during embryogenesis but methylation patterns change during differentiation and remain stable in differentiated cells. (Smith, Chan et al. 2012) Investigations on the role of methylation in the human genome have shown that only a few of the CpG dinucleotides in the human genome change their methylation state and that these are in regulatory elements, suggesting that this may be a mechanism used to regulate particular sets of genes in differentiation (Ziller, Gu et al. 2013). For example binding of the vertebrate insulator protein CTCF can be cell type-specific and can be regulated by differential DNA methylation (Wang, Maurano et al. 2012) (Ong and Corces 2014).

Recently it has been shown that DNA methylation can affect nucleosome stability and positioning (Collings, Waddell et al. 2013) (Jimenez-Useche, Ke et al. 2013). In addition there is a correlation between DNA methylation and histone methylation patterns. (Smith and Meissner 2013) (Meissner, Mikkelsen et al. 2008). Methylated CpG residues are recognised by methyl-CpG-binding domain proteins (MBD) which are thought to block access to transcription factors by recruiting co-repressor complexes. For example, MeCP2 is an MBD protein that recognises methylated DNA and co-purifies with the SIN3-histone deacetylase (HDAC) complex which deacetylates nucleosomes bringing about a repressed chromatin structure, thus influencing gene expression (Jones, Veenstra et al. 1998).

Hence, the relationship between methylated DNA and chromatin is important since DNA methylation may lead to the recruitment of enzymes that are involved in post-translational

modifications of histone proteins and these changes in chromatin structure ultimately affect gene expression.

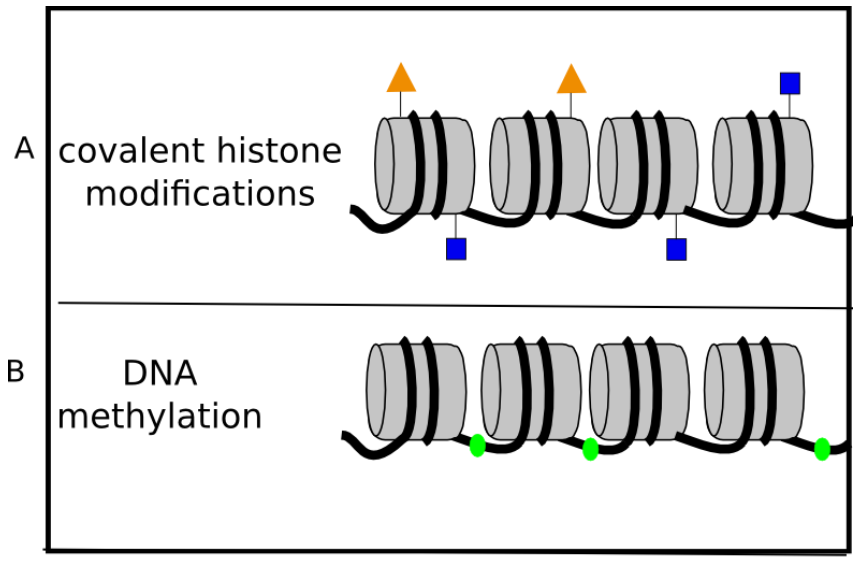


Figure 1.5 Modifications that affect the structure of nucleosomes.

A. Core histone proteins can be post translationally modified, for example by acetylation (orange) and methylation (blue). **B** DNA methylation (green) may be involved in altering the structure of chromatin directly by affecting nucleosome stability and positioning.

1.10 Nucleosome positioning

Nucleosome positioning refers to the location of the nucleosome core relative to the DNA sequence. This is referred to in two ways; translational positioning refers to the linear position of the DNA wrapped around the histone octamer whereas rotational positioning refers to the orientation of the DNA that is wrapped around the nucleosome (Struhl and Segal 2013). A single turn of DNA is approximately 10.5 bp (Wang 1979), therefore a 10 bp change in the translational position of a nucleosome would not affect its rotational position.

The positions of nucleosomes are often considered across a population of cells. They are considered to be 'positioned' when, in a population of cells, the centre of a nucleosome is in the same location in the genome from one cell to the next. Nucleosomes that are not positioned vary in their location from one cell to the next and are termed 'fuzzy' (Fig 1.6) (Mavrich, Ioshikhes et al. 2008).

Recent work has shown that broadly, nucleosome positioning is affected by the action of chromatin remodelling complexes 2) the underlying DNA sequence and 3) the action of transcription factors. However, there is evidence that the positions of adjacent nucleosomes are dictated by the positions others. This is known as 'statistical' positioning and it is thought that a large amount of nucleosome positioning across a genome may be 'statistical'. The next sections will discuss the factors that influence nucleosome positioning.

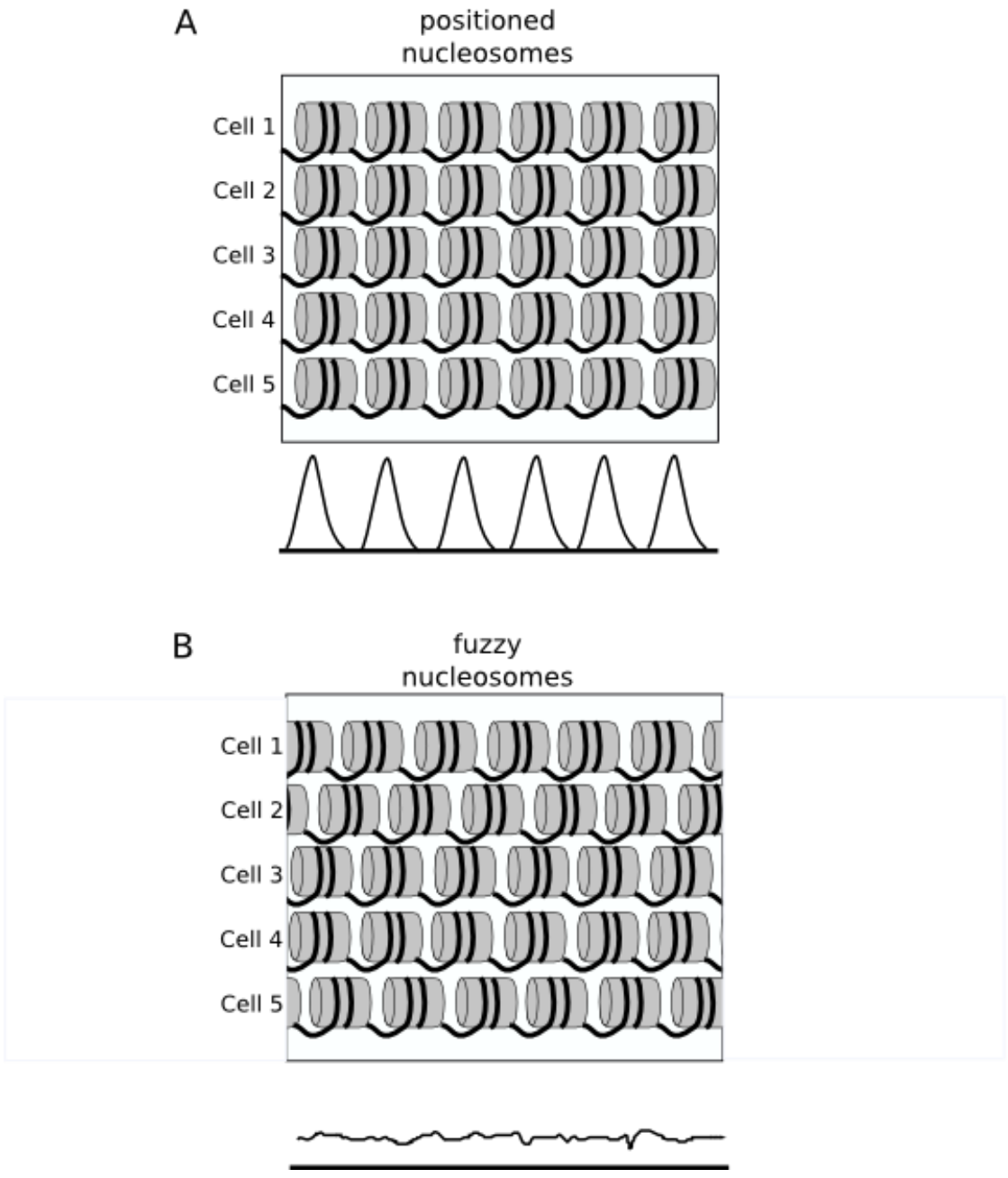


Figure 1.6 Nucleosome positioning.

A. In a population of cells, nucleosomes that are in the same place in the genome from one cell to the next are 'positioned'. **B.** Nucleosomes that are not in the same place in the genome from one cell to the next are not positioned or 'fuzzy'.

1.11 ATP-dependent chromatin remodelling.

Nucleosomes can be placed by the action of ATP-dependent chromatin remodelling enzymes, independently of DNA replication. This was first shown in *S. cerevisiae* in a cell-free system by reconstituting nucleosomes at the Pho5 promoter using yeast extracts and adding histone proteins (Korber and Horz 2004). Recently, further work has shown that ATP-dependent nucleosome sliding and eviction are necessary to achieve *in vivo* patterns of nucleosome positions (Padinhateeri and Marko 2011).

ATP-dependent chromatin re-modelling complexes can change linker length by sliding or disrupting nucleosomes (Havas, Whitehouse et al. 2001), or exchange histone variants (Kamakaka and Biggins 2005) for example H2AZ (Ausio and Abbott 2002), into nucleosomes. All of these complexes contain a Snf2-type ATPase subunit and variable numbers of other subunits. ATP-dependent chromatin re-modelling complexes have been divided into four broad groups according to the structure of the ATPase subunit and their remodelling functions (Clapier and Cairns 2009).

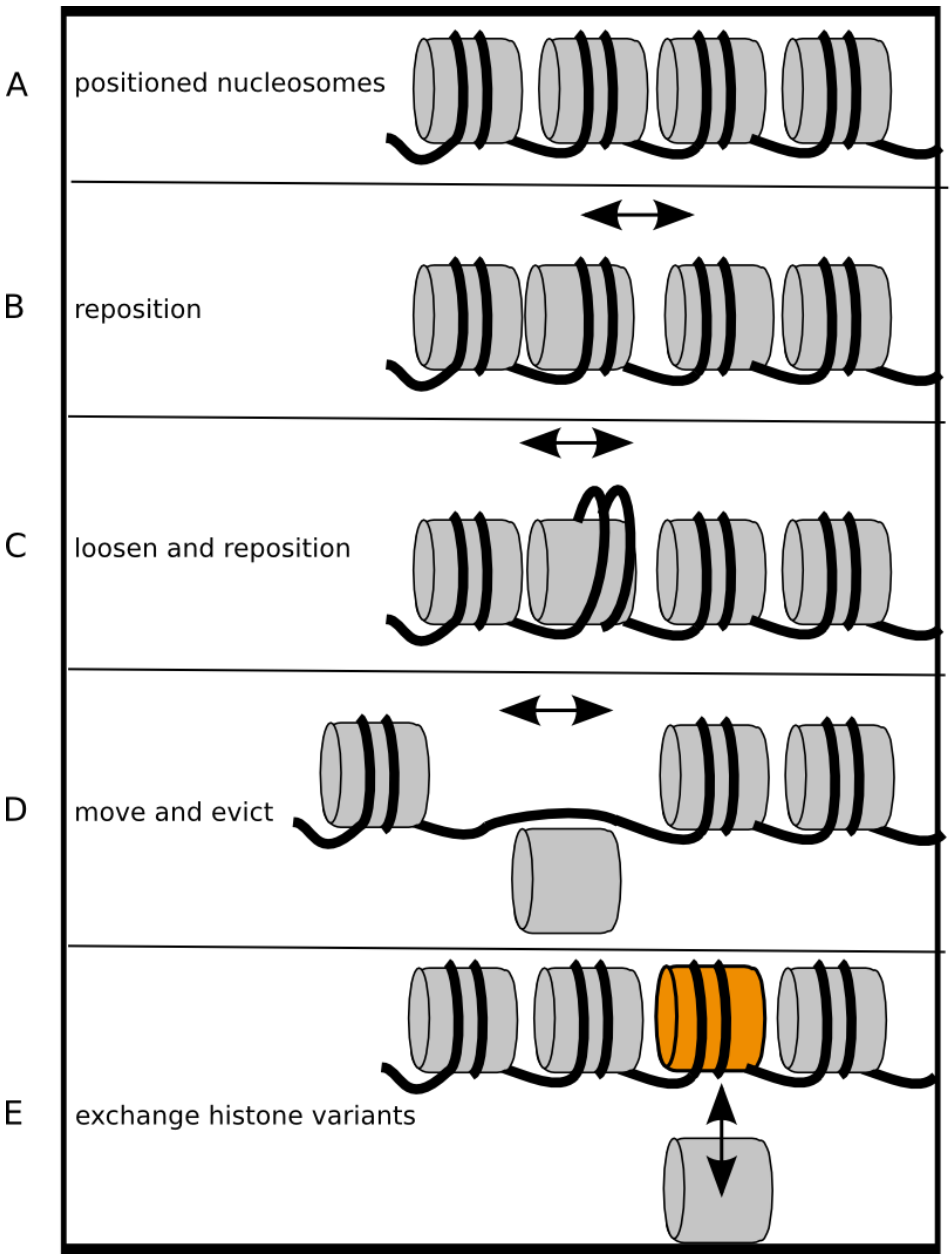


Figure 1.7 ATP dependent re-modelling complexes can alter the position, spacing and structure of nucleosomes.

A. An array of positioned nucleosomes with equal linker lengths. **B**. nucleosomes are repositioned without affecting their structure. **C**. DNA wrapped around the nucleosome core is loosened by altering histone-DNA contacts and nucleosomes are re-positioned. **D**. Nucleosomes are moved and evicted. **E**. Components of the nucleosome core are exchanged for histone variants, destabilising nucleosomes

1.11.1 The SWI/SNF remodellers

Components of the SWI/SNF complex were first identified in *S. cerevisiae* by two groups. The SWI (SWItch) genes were identified as positive regulators of the mating type switching enzyme, HO (Stern, Jensen et al. 1984) and the SNF genes (Sucrose NonFermentable) as a regulator of sucrose metabolism (Neigeborn and Carlson 1984). Later, it was shown that SNF2 had ATPase activity (Laurent, Treich et al. 1993) and that these proteins are members of a multi-subunit complex (Cairns, Kim et al. 1994). In yeast, the RSC remodelling complex (Remodelling the Structure of Chromatin) has similar properties to the SWI/SNF complex, but different biological functions, but is considered to be a member of this family of remodellers (Cairns, Lorch et al. 1996). The SWI/SNF family of re-modellers can both re-position and change the structure of nucleosomes by changing the histone-DNA interactions (Whitehouse, Flaus et al. 1999). These proteins are characterised by having a bromo-domain that recognises acetylated lysines in the N-terminal tails of histones (Hassan, Prochasson et al. 2002). In addition, Kruger *et al* showed that point mutations in histones suppress mutations in SWI/SNF genes (Kruger, Peterson et al. 1995).

The BAF (Brg1 Associated Factors) (Wang, Cote et al. 1996) and PBAF (Polybromo-associated BAF) (Xue, Canman et al. 2000) complexes are human chromatin remodelling complexes from the SWI/SNF family of remodellers. Either of two ATPase subunits, hBRM (human Brahma), the human homologue of SNF2, or BRG1 (Brahma-related Gene 1) can exist in the BAF complex whereas PBAF contains only BRG1. BRG1 is involved in the recruitment REST, the transcriptional repressor of neural genes in non-neuronal cells, to the RE1 binding sites, through recognising H4K8 (Ooi, Belyaev et al. 2006).

1.11.2 The CHD re-modellers

The chromo-domain-helicase-DNA-binding proteins (CHD) are chromatin re-modellers that have been shown to move and evict nucleosomes in *S. cerevisiae* (Stockdale, Flaus et al. 2006) . CHD1 was first characterised in the mouse and shown to have two N-terminal chromo-domains in addition to a DNA binding domain (Delmas, Stokes et al. 1993). CHD remodellers can bind to the methylated H3K4 histone mark in *S. cerevisiae* (Pray-Grant, Daniel et al. 2005) and in humans Flanagan (Flanagan, Mi et al. 2005) through their chromo-domains.

CHD1 is the most characterised of this family of proteins. In the fission yeast *S.pombe* it has been shown to have a role in the regulation of nucleosome spacing (Pointner, Persson et al. 2012). In higher eukaryotes it has been shown to have a role in regulating transcription, associating with HDAC and the transcriptional repressor NcoR in mouse (Tai, Geisterfer et al. 2003) In humans it is recruited to active areas of transcription (Siggens, Cordeddu et al. 2015)

and it can bind to methylated histone H3 (Sims, Chen et al. 2005). CHD1 is important in the maintenance of pluripotency in mouse ES cells (Gaspar-Maia, Alajem et al. 2009) and in the development of *D. discoideum* (Platt, Rogers et al. 2013). These experiments suggest a role for CHD1 in regulating the transcription of genes involved in cell development through chromatin remodelling, whereby the CHD proteins may act as 'readers' of the post translational histone modifications through their chromo-domains.

1.11.3 The INO80 and SWR remodellers

INO80 (Inositol-requiring protein 80) and SWR remodellers are multi-subunit complexes that share protein subunits. Both of them have an HSA (SANT-associated domain) domain and a Snf2-ATPase domain with a long insertion. In humans, the HSA domain recruits IES (Ino Eighty Subunit), actin-related (Arp) proteins (Chen, Conaway et al. 2013) and YY1 (Cai, Jin et al. 2007) and further IES proteins and Arp5 are recruited to the long insertion in the ATPase domain. The INO80 complex was first characterised in *S. cerevisiae* and shown to slide histone octamers along DNA (Shen, Mizuguchi et al. 2000). Later it was shown that this family of remodellers exchange histone H2A with the histone variant H2AZ. In *S. cerevisiae*, SWR-1 incorporates H2AZ into nucleosomes (Mizuguchi, Shen et al. 2004) and INO80 removes H2AZ (Papamichos-Chronakis, Watanabe et al. 2011). When H2AZ is incorporated into +1 nucleosomes, this destabilises the +1 nucleosome leading to gene activation (Yen, Vinayachandran et al. 2013). In addition, INO80 is involved in altering nucleosome spacing (Udugama, Sabri et al. 2011) and it is in DNA repair and replication (Morrison and Shen 2009). Recently the structure of the yeast INO80 complex has been characterised yielding further insight into mechanisms of its function. (Tosi, Haas et al. 2013)

The human IN080 complex has been characterised (Chen, Cai et al. 2011), but there are two complexes which are candidates for being human orthologues of SWR; SRCAP (Ruhl, Jin et al. 2006) and p400/TIP60 (NuA4). As the subunits of the human tip60 (NuR4) complex are homologous to the components of the yeast SWR1 or NuA4 complex, it is thought that this complex corresponds to a hybrid of the two yeast complexes (Auger, Galarneau et al. 2008). Recently a human H2AZ chaperone protein was characterised (Obri, Ouararhni et al. 2014). ANP32E is an H2AZ specific chaperone that is part of the p400/TIP60 complex, able to remove H2AZ from nucleosomes *in vitro* that regulates H2AZ at promoters.

1.11.4 The ISWI remodellers.

The ISWI re-modellers (imitation switch) (Deindl, Hwang et al. 2013) re-position nucleosomes without altering their structure (Tsukiyama, Palmer et al. 1999). Examples of this group of protein are ISW1 and 2 in *S. cerevisiae* and their mammalian orthologues SNF2H and SNF2L.

These proteins have a SANT domain (Boyer, Latek et al. 2004) which can bind histone proteins , a HAND domain and a SLIDE domain which are involved binding in DNA. (Grune, Brzeski et al. 2003).

1.12 The effects of DNA sequence on nucleosome positioning.

Many studies have been carried out in the *S. cerevisiae* genome to determine the factors that influence nucleosome positioning. Studies have shown that although there is no canonical nucleosome positioning sequence in yeast, particular sequences do influence nucleosome positioning. DNA sequence affects the ability of the nucleosome core histones to bind to DNA; certain DNA sequences are more favourable for DNA bending and DNA has to bend around the core histone octomer to form a nucleosome. Drew and Travers *et al* (Drew and Travers 1985) showed that repeating 10 bp patterns of the nucleotides AA/TT creates DNA bending , thus nucleosomes are more likely to form in regions of DNA with these sequences. Conversely, DNA comprising poly dA : dT tracts or poly dC : dG tracts is rigid and does not favour nucleosome formation. In yeast, poly dA : dT tracts of DNA tend to be found at core promoters, where nucleosomes are excluded (Iyer and Struhl 1995; Ioshikhes, Albert et al. 2006; Lee, Tillo et al. 2007). In addition, in yeast, AA and TT dinucleotides are enriched at the 5' and 3' ends of positioned nucleosomes and AT/TA sequences are located at the nucleosome borders near the TSS (Mavrich, Ioshikhes et al. 2008).

In the human genome, the role of DNA sequence in nucleosome positioning has not yet been widely investigated but there is evidence that nucleosomes are depleted at CpG islands (Tazi and Bird 1990) and recently it has been shown that GAA repeats in the human genome lead to nucleosome depletion (Zhao, Xing et al. 2015).

1.13 Statistical positioning of nucleosomes.

It has been proposed that the positions of nucleosomes are affected by the positions of others, i.e. if one nucleosome is in a particular location in the genome, then the positions of adjacent nucleosomes would be dictated by the positions of the first (Kornberg and Stryer 1988) . This idea assumes that there are 'barrier' regions in the genome, that prevent nucleosomes binding and that nucleosomes can move freely outside of barrier regions. In *S. cerevisiae*, it has been suggested that +1 nucleosomes may be positioned according to the underlying DNA sequence. Then +1 nucleosomes may act as barriers against which nucleosomes are positioned in arrays, their packing and organization decaying with distance from the +1 nucleosome (Mavrich, Ioshikhes et al. 2008). Barriers also could be formed by the action of chromatin remodelling complexes, removing or displacing nucleosomes (Cairns 2009) or by DNA sequences that

exclude nucleosomes (Segal and Widom 2009). Further evidence for the fact that some nucleosomes are strongly positioned, forming barriers, and others are not, was shown in experiments by Gossett *et al*. Global depletion of histone proteins tends to lead to the occurrence of fewer well-positioned nucleosomes; it was shown that nucleosomes that were retained after histone H3 depletion were associated with stabilizing histone marks, chromatin remodelling activity or DNA sequences that favour nucleosome positioning (Gossett and Lieb 2012).

In humans, it has been suggested that Pol II may act as a barrier, since positioned nucleosomes are located downstream of stalled Pol II at the promoter of genes (Valouev, Johnson et al. 2011) and nucleosomes tend to be depleted upstream of the TSS in a polII dependent manner (Schones, Cui et al. 2008) . Until recently, the mechanism by which this barrier could be removed was not known, but recent work suggests that CHD1 may be responsible for removing nucleosomes (Skene, Hernandez et al. 2014). In addition, CTCF and tRNA genes can act as barriers or insulators (See section 1.15).

1.14 Patterns of nucleosome positioning in eukaryotic genomes.

Studies of the positions of nucleosomes in eukaryotic genomes have shown that there are areas of the genome that possess arrays of nucleosomes and areas where nucleosomes are depleted. Much of the most detailed analysis of the patterns of nucleosome positions has been undertaken in the *S. cerevisiae* genome (Jiang and Pugh 2009; Park 2009) (Kent, Adams et al. 2011). The yeast genome contains 6,000 small (less than 1kb), tightly packed genes (Goffeau, Barrell et al. 1996) over which nucleosomes are densely packed and tend be uniformly spaced. There are approximately 70,000 positioned nucleosomes which occupy about 81% of the *S. cerevisiae* genome (Lee, Tillo et al. 2007). In *S. cerevisiae*, gene promoters and the 3' ends of genes tend to be nucleosome depleted, whereas gene bodies tend to possess nucleosome arrays (Lee, Tillo et al. 2007). Areas of nucleosome depletion are known as nucleosome free regions (NFR) and strongly positioned nucleosomes have been found to occur near the transcriptional start site (TSS) (Mavrich, Ioshikhes et al. 2008; Jiang and Pugh 2009). Nucleosomes positioned just upstream of the transcriptional start site are known as the -1 nucleosome and the first nucleosome that is positioned downstream of the TSS is known as the +1 nucleosome. These strongly positioned nucleosomes in gene promoters are thought to be important in the regulation of gene expression. Early *in vitro* work showed that enzymes can traverse chromatin displacing nucleosomes during elongation but not during the initiation of transcription suggesting that nucleosomes might limit the access of polymerases to genes. (Lorch, LaPointe et al. 1987; Boeger, Griesenbeck et al. 2003; Henikoff 2008).

Also in *S. cerevisiae*, it has been shown that nucleosomes can restrict the access of activatory transcription factors to their binding sites (Segal, Fondufe-Mittendorf et al. 2006) (Lee, Tillo et al. 2007), thus affecting transcriptional activity and nucleosome loss through histone depletion, resulting in an increase in transcription in many yeast genes (Han and Grunstein 1988). However, there may be more than one type of organisation of chromatin at gene promoters with variation in the mobility and groupings of nucleosomes at and surrounding the transcriptional start site, affecting both expression levels and transcriptional plasticity- the ability to alter gene expression. Tirosh and Barkai *et al* showed that in *S. cerevisiae*, nucleosome occupancy can be high upstream of the TSS in some promoters and that this correlates with lower transcriptional plasticity. In contrast, gene promoters where nucleosome occupancy was high in the distal region of the promoter and not at the TSS, tend to have more mobile nucleosomes and higher transcriptional plasticity, the gene expression change can be up or down, thus higher levels of nucleosome occupancy correlates with more regulation. (Tirosh and Barkai 2008).

In recent years, genome-wide patterns of nucleosome positioning at transcriptional start sites have been well characterised in a wide range of other eukaryotic model organisms; in *S. pombe* (Lantermann, Straub et al. 2010) *D. melanogaster* (Mavrich, Jiang et al. 2008), the cellular slime mould *D. discoideum* (Platt 2013) (Chang, Noegel et al. 2012) and in *C. elegans* (Valouev, Ichikawa et al. 2008). These studies have shown that broadly, there is a pattern of nucleosome positioning at TSS. The TSS itself tends to be devoid of strongly positioned nucleosomes but it is flanked by well positioned nucleosomes on either side; the -1 and +1 nucleosomes. Further by phased arrays of well positioned nucleosomes occur on either side of the -1 and +1 nucleosomes. In these arrays, the strength nucleosome positioning diminishes with the distance from the TSS.

Genome-wide nucleosome maps also have been constructed from higher eukaryotes; from the smaller of the plant genomes; A. *thaliana* and O. *sativa ,* (Chodavarapu, Feng et al. 2010; Li, Liu et al. 2014; Wu, Zhang et al. 2014; Singh, Bag et al. 2015; Zhang, Zhang et al. 2015), from the mouse (Li, Schug et al. 2011) (Teif, Vainshtein et al. 2012) and from human cells (Ozsolak, Song et al. 2007; Schones, Cui et al. 2008; Tillo, Kaplan et al. 2010; Gaffney, McVicker et al. 2012; Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012). These studies have shown that broadly, phased arrays of well positioned nucleosomes flank the NFR at the TSSs, with low numbers of nucleosomes in gene promoters and higher levels in exons (Tilgner, Nikolaou et al. 2009). In the human genome positioned nucleosomes are located at gene promoters (Schones,

Cui et al. 2008) and in regulatory regions such as at CTCF sites (Fu, Sinha et al. 2008) and at enhancers (West, Cook et al. 2014)

Although these organisms appear to have similar patterns of chromatin structure at TSS, they have widely differing genome sizes (Table1.1) and the percentage of the genome that possesses well positioned nucleosomes appears to be inversely proportional to genome size. For example in *S*. *cerevisiae* approximately 85% of the genome possesses well-positioned whereas In the plant genomes, *A*. *thaliana* and *O*. *sativa,* the figure is 12% and 16% of the genome respectively (Zhang, Zhang et al. 2015). This suggests that although there appear to be some common patterns of nucleosome positioning across eukaryotic genomes, there might be variation in the mechanisms by which chromatin regulates the genome.

Organism	Genome size (M bp)	No. of protein coding genes	Reference
S. cerevisiae	12	6,000	(Goffeau, Barrell et al. 1996)
S. pombe	12.5	5,124	http://www.pombase.org
D. discoideum	34	12,257	http://dictybase.org
C. elegans	100	19,735	(Hillier, Coulson et al. 2005) http://www.ensembl.org
A. thaliana	135	27,416	(AGI 2000) https://www.arabidopsis.org
D. melanogaster	143	13,918	http://www.ensembl.org
M. musculus	2500	22,528	(Waterston, Lindblad-Toh et al. 2002) http://www.ensembl.org
H. sapiens	3000	19,814	(Lander, Linton et al. 2001), http://www.gencodegenes.org

Table 1.2 The genome sizes of a selection of model eukaryotic organisms.

1.15 Insulators

1.15.1 Insulator proteins.

Insulator proteins were first characterised in the Drosophila genome (Kellum and Schedl 1991). They seem to have a dual role in genome regulation, acting at a local level as enhancer blockers, preventing the interaction of enhancers and promoters (Cai and Levine 1995). In addition they are present at the boundaries of sub-topologically associated domains (TADs), where genes and their regulatory regions are brought together over long distances, forming loops (Maeda and Karch 2007). Their role seems to be in organising the genome into domains, organising chromatin in a cell type specific manner, resulting in cell type specific gene expression.

CTCF is a DNA binding protein that was first identified as a transcriptional repressor of the chicken c-myc gene (Lobanenkov and Gudvin 1989). Identification and characterisation of the chicken CTCF binding motif showed that CTCF acts by binding to three repeats of the DNA sequence CCCTC, hence the name CCCTC-binding factor (Lobanenkov, Nicolas et al. 1990). CTCF is a highly conserved, well characterised 11- zinc-finger DNA binding protein whose isoforms are differentially expressed in the chicken (Klenova, Nicolas et al. 1993). CTCF is the major insulator so far characterised in human cells (Filippova, Fagerlie et al. 1996) and appears to have many roles (Phillips and Corces 2009). It is both a transcriptional activator and repressor and acts as an insulator protein, blocking the interaction between enhancers and promoters, thus demarcating active and inactive chromatin domains (Handoko, Xu et al. 2011). In addition, CTCF appears to have dominant role in determining chromatin structure on a global level, defining different chromatin compartments (Cuddapah, Jothi et al. 2009) by playing a role in the formation of chromatin loops (Splinter, Heath et al. 2006; Handoko, Xu et al. 2011; Rao, Huntley et al. 2014) facilitating long-range chromosome interactions (MacPherson and Sadowski 2010; Phillips-Cremins and Corces 2013)

A large body of work has shown the association of CTCF at specific DNA binding sites at the stem of chromatin loops, forming topological domains (Dixon, Selvaraj et al. 2012) .Recent work has shown that CTCF binding sites have directionality specified through the differential binding of its zinc fingers. This determines the direction in which chromatin loops are formed, with CTCF sites at the bases of loops having the opposite orientation from each other (Guo, Monahan et al. 2012; Guo, Xu et al. 2015; Vietri Rudan, Barrington et al. 2015). However, the relationship between the transcriptional regulation of genes that determine cell lineage and TAD structures is still not well understood. Recent studies in the mouse genome during differentiation from ESC to NPC, using 5C technology have been carried out. This has shown

that particular combinations of architectural proteins are utilised in organising higher order chromatin structure during lineage commitment. (Phillips-Cremins, Sauria et al. 2013). This demonstrates the importance of architectural proteins such as CTCF in chromatin organisation and cell development.

The importance of CTCF has been shown though the demonstration of its role in human disease. Alterations in CTCF-associated boundary regions resulting in the disruption of TADs can result in drastic effects. This has been shown in mouse models of human genetic disorders resulting in limb malformation syndromes caused by deletions, inversions, or duplications of the TAD that spans the region of chromosome 2 containing the genes WNT6/IHH/EPHA4/PAX3 (Lupianez, Kraft et al. 2015). In addition, several *de novo* point mutations in CTCF have been shown to be responsible for causing intellectual disability, microcephaly and growth retardation, suggesting a role for CTCF in neuro-development (Gregor, Oti et al. 2013). In addition, recent work has shown that CTCF is important in the early development of the mouse brain. Watson *et al* have shown that CTCF is important for NPC survival and the loss of CTCF can lead to the premature differentiation of NPC cells and therefore depletion of the NPC pool of cells (Watson, Wang et al. 2014).

Human CTCF was characterised by Filipova *et al* whose early work on defining the binding sequences of CTCF showed that CTCF could bind to divergent sequences in the promoter of c-myc by utilising different combinations of its zinc fingers, hence it was termed a 'multivalent protein' (Filippova, Fagerlie et al. 1996). Several CTCF binding motifs have been derived from different human cell types (Kim, Abdullaev et al. 2007; Xie, Mikkelsen et al. 2007; Bao, Zhou et al. 2008; Jothi, Cuddapah et al. 2008) and recent genome-wide studies have revealed a high level of complexity in the binding of CTCF to DNA. Nakahashi *et al* (Nakahashi, Kwon et al. 2013) studied the sequence diversity, genome-wide, of a similar number of CTCF binding motifs across primary lymphocytes and defined groups of zinc fingers that bind specifically to different components of the binding motif; the conserved core and the non-conserved flanking sequences. More detailed studies of CTCF binding motifs has shown that there are, on average 55,000 binding motifs across a wide variety of human cell types, but that binding site utilisation is cell type specific (Wang, Maurano et al. 2012). These recent genome-wide studies support the early work suggesting that CTCF binding is complex; through variation in the DNA sequences bound and via the differential use of its zinc fingers. In addition, there is evidence that CTCF binding sites might have different binding affinities suggesting the occurrence of stronger binding at ubiquitous CTCF sites and that weaker binding occurs at cell type specific sites (Chen, Tian et al. 2012; Plasschaert, Vigneau et al. 2014).

Taken together, the complexity of CTCF binding would allow maximum flexibility in its putative binding locations in the genome, perhaps providing mechanisms of cell type specificity in binding, dictating lineage specification.

1.15.2 tRNA genes

Classically tRNA genes (tDNAs) are known for their role in translation in protein synthesis. They are transcribed by RNA polymerase III generating transfer RNA molecules that assemble amino acids during protein synthesis. However, in *S. cerevisiae* it has been known for some time that these tRNA genes have another role; they can act as insulators, for example silencing the HMR locus (Donze and Kamakaka 2001). This function of tDNAs depends on the binding of RNA polymerase III (RNAP) (Donze and Kamakaka 2001) the transcription factors TFIIIB and TFIIIC (Simms, Dugas et al. 2008). It was shown in *S. cerevisiae* that TFIIIC binds to the B-box in tRNA genes or to sites elsewhere known as ETC sites (extra TFIIIC sites), these are TFIIIC binding sites that are transcription independent (Moqtaderi and Struhl 2004). In addition, TFIIIC-mediated barriers require chromatin remodellers such as the RSC complex (Valenzuela, Dhillon et al. 2009).

Recently tRNA genes have been shown to act as enhancer blocking insulators in mouse cells (Ebersole, Kim et al. 2011) and both enhancer blockers and barrier insulators in human cells (Raab, Chiu et al. 2012). Similarly the presence of ETC's has been shown in humans and these sites, enriched in TFIIIC, have associated CTCF binding (Oler, Alla et al. 2010) (Moqtaderi, Wang et al. 2010). Whether cohesin is recruited at this type of insulator in a similar manner to how it is recruited at CTCF binding sites, is not yet clear. However, this conserved role of tRNA genes as insulators may be an important mechanism for genome organisation and therefore cell type-specific gene regulation in human cells. (Van Bortle and Corces 2012)

1.16 Cell differentiation

1.16.1 Embryonic stem cells ESCs.

Embryonic stem (ESCs) exist in the inner cell mass of a pre-implantation embryo. They have the capacity to differentiate into any of the three germ layers; endoderm, mesoderm or ectoderm, this property is known as 'pluripotency'. Pluripotent cells are characterised by having globally reduced DNA methylation, depletion of the post-translational histone modification H3K27me3 at the promoters of developmental genes and female cells retain two activated X-chromosomes. In addition, early work showed that particular transcription factors are expressed in pluripotent cells, defining their 'stemness'. For example, Nanog (Mitsui, Tokuzawa et al. 2003), Oct4 (Niwa, Miyazaki et al. 2000) and Sox2 which co-occupy a number

of their target genes that encode developmentally important transcription factors (Boyer, Lee et al. 2005), in addition, they co-occupy the promoters of microRNAs that are involved in the maintenance of pluripotency (Marson, Levine et al. 2008).

1.16.2 Mouse embryonic stem cells (mESCs).

Embryonic stem cells were first established in cell culture in the mouse, using mitotically inactive fibroblasts or 'feeder cells' (Evans and Kaufman 1981; Martin 1981). This opened the gateway to looking for the mechanisms involved in the control of both pluripotency and differentiation. Subsequently, several key signalling factors were found to be important in the maintenance of the undifferentiated state in mouse ESCs; Leukaemia inhibitory factor (LIF) (Kurisaki, Hamazaki et al. 2005) and bone morphogenetic protein 4 (BMP4) (Ying, Nichols et al. 2003) . Hence it was possible to culture mouse ESCs in culture using LIF and without the need for feeder cells (Smith, Heath et al. 1988). More recent work established the production of mESC by stimulating cells with LIF and by inhibition of ERK1/ERK2 and GSK3β signalling. These cell culture conditions are known as 2i/LIF (Ying, Wray et al. 2008) and generate cultured mouse ESC cells in the 'naïve' state, similar to primary cells derived from the inner cell mass.

Another category of pluripotent cells known as epiblast stem cells (mEpiSCs) has been derived and characterised. (Tesar, Chenoweth et al. 2007). These cells are derived from the post-implantation epiblast and represent a pluripotent state that is distinct from that in ESC cells, the X chromosome is inactivated and these cells are poised to differentiate into progenitor cells. This state is known as the 'primed' pluripotent state (Nichols and Smith 2009).

1.16.3 Human (hESCs).

Human ESCs were established initially using irradiated mouse embryonic fibroblasts as 'feeder 'cells (Reubinoff, Pera et al. 2000) (Thomson, Itskovitz-Eldor et al. 1998) However, it was difficult to establish clonal populations from these cells. Later, clonal populations of ES cells were derived by using serum-free medium and the addition of recombinant basic fibroblast growth factor (bFGF) (Amit, Carpenter et al. 2000) and later feeder-free cultures were established (Xu, Inokuma et al. 2001). More recently, studies of the growth and maintenance of human ESCs in culture have shown that human ESCs differ from mouse ESCs; self-renewal in human cells is dependent on different signalling pathways than in mouse. The maintenance of pluripotency in human cells is dependent on the inhibition of the TGFβ/Activin/nodal and the FGF signalling pathways (Vallier, Alexander et al. 2005). In addition, the human pluripotent state has been described as 'primed', whereas that in mouse is described as 'naive' (Ying, Wray et al. 2008) therefore human ESCs differ from mouse ESCs and are thought to be most similar

to mouse EpiESCs. (Hanna, Cheng et al. 2010). Recently, Gafni *et al* (Gafni, Weinberger et al. 2013) have shown that it is possible to derive human 'naïve' cells using defined culture conditions (NHSM medium) which include LIF/2i, TGF beta, FGF2, JNKi, p38i and ROCKi (Y27632). Manipulation of growth factor signalling in defined culture conditions has led to the ability to define pluripotent states in human cells and study the mechanisms involved in their maintenance.

1.16.4 Induced pluripotent stem cells.

Induced pluripotent cells (iPSC) are pluripotent cells derived from somatic cells**.** These cells provide a convenient source of pluripotent cells without using embryos, facilitating research into the mechanisms involved in both the maintenance of pluripotency and in bringing about differentiation. These cells can be produced for example, from fibroblasts, by express**ing** transcription factors that are involved in the maintenance of pluripotency such as KLF4, SOX2 and OCT4 and c-Myc; iPSC have been generated in both mouse (Takahashi and Yamanaka 2006) and human cells (Takahashi, Tanabe et al. 2007; Yu, Vodyanik et al. 2007). More recently, cell culture methods have been refined to generate naïve hiPSC cells with characteristics close to those of mouse iPS cells. (Hanna, Cheng et al. 2010).

1.16.5 Neural Cell Differentiation.

Neural progenitor cells are multipotent cells in the nervous system that can differentiate into neurons, astrocytes and oligodendrocytes and have a limited capacity to self-renew (Carpenter, Cui et al. 1999). It is possible to generate mammalian NPCs *in vitro* from embryonic stem cells, this has been done in human embryonic stem cells and from induced pluripotent cells (Chambers, Fasano et al. 2009). In the embryo, neural cells are produced from the ectoderm. Generation of NPCs *in vitro* is undertaken by culturing pluripotent cells in medium containing two factors that inhibit SMAD signalling, generating early neural ectoderm cells. Dual SMAD inhibition utilises the small molecule inhibitor SB431542 to inhibit the ALK4, ALK5, and ALK7 receptors (Inman, Nicolas et al. 2002) and block the TGFβ pathway which act via SMAD 2 and 3, which controls cell proliferation and differentiation to mesoderm, and inducing neural differentiation (Smith, Vallier et al. 2008). A second small molecule inhibitor LDN-193189 inhibits the transcription of the BMP type I receptors ALK2 and ALK3 which act via SMADs 1, 5 and 8 in the Bone morphogenetic protein (BMP) pathway. This pathway is involved in regulation of cell fate decisions during embryogenesis.

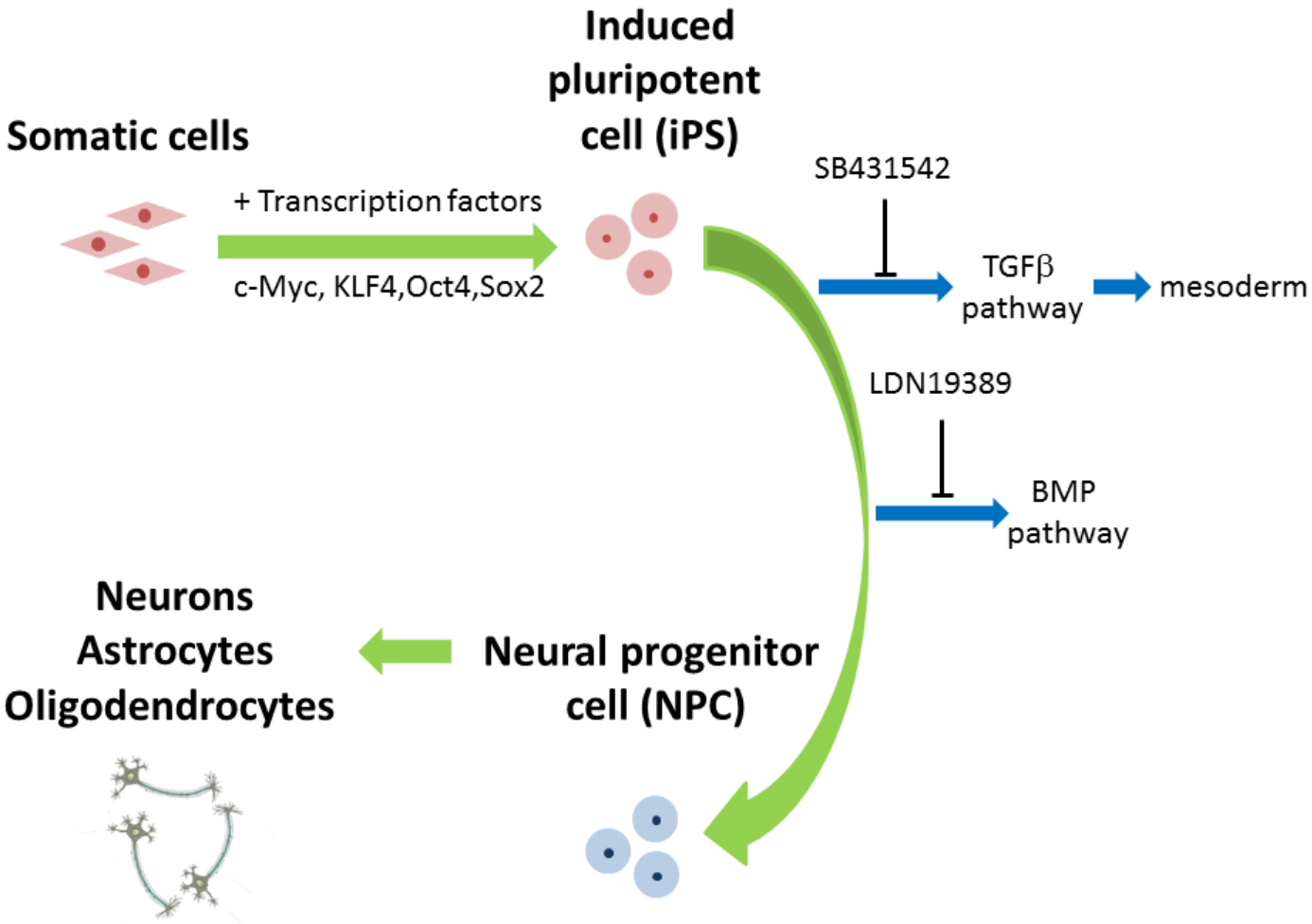


Fig 1.8 See next page for legend

Figure 1.8 Generation of induced pluripotent cells and derived neural progenitor cells from somatic cells.

iPS cells are generated by expression of transcription factors responsible for maintaining pluripotency; cMyc, KLF4, Oct4 and Sox2. NPC cells are generated from iPS cells after the formation of embryoid bodies, by dual SMAD inhibition (Chambers, Fasano et al. 2009) using the inhibitors SB431542 and LDN19389 that inhibit the TGFβ and BMP pathways respectively.

1.17 Chromatin structure during cell differentiation

Observations of chromatin structure in ES cells showed that chromatin tends to be globally decondensed (Ahmed, Dehghani et al. 2010). Some chromatin structural proteins, for example HP1 (Cheutin, McNairn et al. 2003) are less tightly bound to chromatin than in differentiated cells (Meshorer, Yellajoshula et al. 2006) In contrast, in mouse neural progenitor cells (NPC) derived from embryonic stem cells (ESC), the chromatin structure becomes more compact (Meshorer, Yellajoshula et al. 2006). In addition, there is a higher expression of chromatin remodellers in ES cells (Kurisaki, Hamazaki et al. 2005) and chromatin remodellers are necessary for differentiation (Kaji, Caballero et al. 2006). These observations suggested that the chromatin architecture in pluripotent cells is both accessible and dynamic leading to the idea that pluripotent cell chromatin architecture is 'plastic' allowing the access of cell machinery to key regions of the genome involved in the regulation of pathways involved in specifying differentiation.

Gross chromatin structure at the level of chromatin domains changes during differentiation of ES cells, these changes occur between chromatin domains, not within them (Dixon, Jung et al. 2015). Recent genome wide studies of replication timing, in both pluripotent cells and cells differentiated from them, have shown that replication timing can alter during differentiation. In addition, replication domains with similar timing are located in close proximity. Thus, replication domains are co-ordinated both in time and space and form both structural and functional units of regulation (Dixon, Jung et al. 2015).

In addition to gross changes in chromatin compaction, 'chromatin marks' have been shown to change during cell differentiation. For example, ES cells have been shown to have fewer histone modifications associated with heterochromatin than differentiating cells.(Meshorer and Misteli 2006). A large body of work has shown that enzymes involved in bringing about post-translational chromatin modifications are important in neural development (Lilja, Heldring et al. 2013).

Recent studies of changes in nucleosome positions in mouse and human cells have shown that changes in positions of single nucleosomes tend to occur at regulatory regions; at the binding sites of transcription factors that are involved in the maintenance of pluripotency and that there is a depletion of nucleosomes at enhancers in pluripotent cells (West, Cook et al. 2014).

1.18 Transcription factors involved in neural development.

Activatory transcription factors are often DNA binding proteins which act by binding to enhancers or promoter proximal elements. They usually bind to a specific binding site that is adjacent to the promoter of a gene and they recruit the transcription machinery. Repressive transcription factors act by blocking the binding of RNA polymerase to the promoter and they may recruit co-repressors to do so. Some transcription factors have specific roles in neural development and/or have associated chromatin remodelling activity. These are discussed in the next section. The chromatin structure at and surrounding the binding motifs of a selection of these transcription factors has been undertaken in the work of this thesis, these were ATF2, YY1 and PAX6 and REST.

1.18.1 ATF2 (activating transcription factor 2).

ATF2 is a global transcription factor that binds to the cyclic AMP response element (CRE). It is involved in DNA damage repair and has histone deacetylase activity (Kawasaki, Taira et al. 2000). It is ubiquitously expressed, most highly in the brain. In mice it has been shown to be involved in neuronal development and migration. (Reimold, Grusby et al. 1996) In humans it is expressed in neurons and has been found to be down-regulated in the hippocampus of Alzheimer's, Parkinson's and Huntington's patients (Pearson, Curtis et al. 2005)

1.18.2 YY1 (yin-yang) transcription factor.

YY1 is a GLI -Kruppel zinc finger protein that has roles in development and differentiation (Beketaev, Zhang et al. 2015) and can act as a transcriptional activator or repressor . It is a DNA-binding protein that can recruit the polycomb repressive complex PRC2 (Basu, Wilkinson et al. 2014) and it is also involved in transcriptional activation by recruiting the chromatin -remodelling complex INO80 (Cai, Jin et al. 2007). It has been shown that its activity is regulated through histone acetylation and deacetylation through the action of HDACs and HATs. (Yao, Yang et al. 2001) and recently it has been shown that YY1 is involved in targeting chromatin to the nuclear periphery (Harr, Luperchio et al. 2015). Thus, it might be expected that changes in chromatin structure might be detected at YY1 binding sites during neural development and differentiation.

1.18.3 PAX6 (paired box 6).

The transcription factor PAX6 is a DNA binding protein that is expressed in developing eye (Hanson and Van Heyningen 1995), telencephalon (Stoykova, Treichel et al. 2000), thalamus,

spinal cord and pancreas and mutations in PAX6 cause aniridia in humans. (Glaser, Walton et al. 1992). In addition, PAX6 is important in embryonic development, specifically in the development of the central nervous system. It has been shown that knockdown of PAX6 in human embryonic stem cells affects neuro-endoderm development. (Zhang, Huang et al. 2010).

1.18.4 EGR1/Zif268/Knox-24/NGIF-A

Is a nuclear transcription factor is encoded by an immediate early gene which may be involved in neuronal plasticity (Knapska and Kaczmarek 2004). It is expressed in both adult and embryonic brain and in (Wells, Rough et al. 2011). EGR-1 activates the CACNA1H gene that encodes the Cav3.2, a component of the voltage-dependent calcium channel complex, though the EGR1 sites in the promoters of the CACNA1H gene. This activation is counteracted by REST (Van loo et al). Mutations in the CACNA1H gene have been linked to epilepsy (Chen, Lu et al. 2003)

1.18.5 The REST complex

REST (repressor element 1 silencing transcription factor/neuron-restrictive silencing factor) is a zinc-finger protein that binds to the neuron-restrictive silencer element (NRSE) (Mori, Schoenherr et al. 1992) also known as RE1 element (Kraner, Chong et al. 1992). The 21 bp NRSE element was first characterised as a selective repressor of the rat superior cervical ganglion 10 (SCG10) gene in non-neuronal cells and this element was shown to have homology to the silencer in the voltage-gated type II sodium channel (NaV1.2) gene. It was shown that a sequence-specific protein (NRSBF) binds to the NRSE/RE1 element and that this factor is present in nuclear extracts from non-neuronal, but not neuronal cells. This suggested that the neuron-specific expression of the SCG10 gene might be regulated by the activity of NRSBF (Mori, Schoenherr et al. 1992). At the same time Kraner *et al* (Kraner, Chong et al. 1992) identified a 28 bp element, named RE1, as a repressor element in the 5' flanking region of the type II gene by using transient expression assays. Later studies characterised the sequence, genome-wide locations and conservation of its binding motif in both human, mouse and rat cells (Bruce, Donaldson et al. 2004; Johnson, Gamblin et al. 2006; Zhang, Xuan et al. 2006; Jorgensen, Terry et al. 2009).

Using *in situ* hybridisation techniques in mouse cells, Schoenherr *et al* showed that REST is expressed in undifferentiated neuronal progenitors but not in differentiated neurons *in vivo* (Schoenherr and Anderson 1995). This and more recent research has shown that REST is

involved in the regulation of neurogenesis by repressing neuronal gene expression in non-neuronal cells (Wu and Xie 2006; Johnson, Teh et al. 2008; Jorgensen, Terry et al. 2009).

REST binds to the RE1 motif and recruits co-repressors (Ballas, Grunseich et al. 2005).The co-repressors of REST mediate histone modifications and chromatin remodelling bringing about repressive changes in chromatin. REST has three domains through which it recruits co-repressors, the N-terminal domain (NTD), through which it recruits the SIN3 histone deacetylase (HDAC) complex (Huang, Myers et al. 1999), the C-terminal domain where the coREST complex, that contains BRG1, HDAC1 and HDAC2, LSD1 and G9a, binds (Andres, Burger et al. 1999; Hakimi, Bochar et al. 2002), and the DNA-binding domain which is involved in the recruitment of a further complex comprising Mediator in combination with G9a (Ding, Tomomori-Sato et al. 2009). During gene repression REST is recruited to RE1 sites. BRG1 (Ooi, Belyaev et al. 2006), an ATP-dependent chromatin remodeller, recognises acetylated H4K8, leading to an increase in REST recruitment to RE1 sites. REST recruits the co-repressor complexes CoREST and SIN3 complexes, both of which contain the histone deacetylases HDAC1 and HDAC2 (Ballas, Battaglioli et al. 2001). The C-terminal domain of REST binds the H3K9 methytransferase G9a (Roopra, Qazi et al. 2004). After H3K9 deacetylation, G9a methylates H3K9. LSD1, a histone demethylase (Tachibana, Sugimoto et al. 2001; Shi, Lan et al. 2004), removes the mono and di-methylation chromatin marks that are associated with gene activation from H3K4 (Shi, Lan et al. 2004).

Studies of REST target genes have shown that many of their functions are neuronal; they are genes involved in for example, ion channels, ion conductance, axon guidance and synaptic transmission (Paquette, Perez et al. 2000; Ballas and Mandel 2005; Johnson, Gamblin et al. 2006). However, REST also has been shown to be involved in the regulation of other more diverse cellular processes, such as heart development (Kuwahara, Saito et al. 2003) , hematopoietic development (Scholl, Stevens et al. 1996), and the cytoskeleton (Schoenherr, Paquette et al. 1996).

Changes in the levels of expression of REST can have catastrophic effects, causing disease. REST may function as an tumour suppressor in human epithelial cells, (Westbrook, Martin et al. 2005) high levels of are REST found in epilepsy (McClelland, Flynn et al. 2011), (Roopra, Dingledine et al. 2012) and recently REST has been shown to be involved in Alzheimer's disease (Lu, Aron et al. 2014). REST appears to have a role as a 'master regulator' role in neural cell development. The key to how REST regulates neural genes appears to lie in the associated chromatin remodelling and modification activities of its co-repressors that resulting in repressive changes in chromatin.

1.19 Chromatin re-modelling and human health.

Mutations in many of the genes that encode chromatin re-modelling proteins are involved in human disease, for example in cancer (Wang, Allis et al. 2007; Wang, Allis et al. 2007; Vogelstein, Papadopoulos et al. 2013). Mutations in genes encoding proteins in the H3.3-ATRX-DAXX chromatin remodelling pathway are responsible for Glioblastoma multiforme (GBM) which is a lethal brain tumour in adults and children (Schwartzentruber, Korshunov et al. 2012) and CHD5 is strong candidate a tumour suppressor gene that is deleted in neuroblastoma (Fujita, Igarashi et al. 2008).

Strikingly, a significant number of the genes that encode chromatin re-modelling proteins are implicated in genetic risk for human neuro-psychiatric disease. In particular, the CHD re-modellers, that can move and evict nucleosomes, have neuro-psychiatric disease associations in humans.

1.19.1 Neuro-psychiatric disorders

Neuro-psychiatric disorders span the clinical disciplines of neurology and psychiatry. These disorders include Autism Spectrum Disorder (ASD) and Attention deficit hyperactivity disorder (ADHD), which are prevalent in children. Also they include schizophrenia and bipolar disorder that often begin in adolescence and early adulthood. They overlap with epilepsy and the dementias arising in the ageing brain. There is also a wide-range of rarer syndromes, such as Rett's and Kleefstra's syndromes, which carry a high risk for one or more of the more common disorders. Taken together, these disorders are estimated to account for 14% of the global burden of disease, greater than that borne by cancer and cardiovascular disease.

1.19.2 Genetics and Neuro-psychiatric disorders

Recent advances in genetics has identified over 100 genes in which mutation confers elevated risk for neuro-psychiatric disorders such as Autism Spectrum Disorder (ASD), epilepsy and schizophrenia (De Rubeis, He et al. 2014). This has been done through large genome-wide association studies (GWAS) of copy number variant (CNV) data, duplications or large deletions of sections of chromosomes, and small nucleotide polymorphism data (SNP) data in many human cases to look for a common trait. These studies have generated lists of candidate disease genes and information about the types and nature of the mutations involved in these neuro-psychiatric disorders. Candidate disease genes include the voltage-gated calcium channel subunits, CACNA1C and CACNB2 (PGC 2013) and N-methyl-D-aspartate receptor (NMDAR) complexes (Fromer, Pocklington et al. 2014) and proteins that interact with, or are subunits of chromatin remodelling complexes.

1.19.3 Epigenetics and Neuro-psychiatric disorders.

GWAS has been useful in identifying the genes responsible for disease where simple inherited mutations of a single gene are responsible, for example fragile X syndrome (Willemsen, Levenga et al. 2011). However, most of these neuro-psychiatric disorders occur through more than one genetic change, which may occur as *de novo* mutations, making the task of finding candidate disease genes difficult. A further complication is that many disorders cannot be explained purely by genetic risk. The relationship between common genetic variation and human disease is complex. It is becoming clear that genome regulation is important in many human diseases. Mis-regulation of the epi-genome can lead to congenital defects affecting pre-natal and early childhood development and is associated with neuro-psychiatric and disorders. Mis-regulation of epigenetic regulators occurs in rare neurodevelopmental disorders such as Rett's, Kleefstra's and CHARGE syndromes. These may be due to mutations in genes encoding proteins that recognise modified DNA. For example, Rett's syndrome is caused by a mutation in the MECP2 gene (Amir, Van den Veyver et al. 1999). MECP2 encodes methyl CpG binding protein 2 which is essential in mature nerve cells and regulates gene expression by binding to methylated CpG in DNA. Alternatively, mutations can occur in genes encoding proteins that covalently modify histone proteins; Kleefstra's syndrome is caused by mutation or deletion of the histone methyl transferase, Histone-lysine N-methyltransferase (EHMT1) which modifies histone H3 by methylating lysine 9 (Kleefstra, van Zelst-Stams et al. 2009).

1.19.4 CHD proteins and human neuro-psychiatric disorders.

Recent research has shown that approximately half of the members of the CHD group of proteins (CHD2, 5, 7 and 8) have neuro-psychiatric disease associations in humans. CHD2 has been shown to modify chromatin structure by catalysing the assembly of nucleosomes into arrays (Liu, Ferreira et al. 2015). It is a strong candidate gene for ASD (O'Roak, Stessman et al. 2014) and it is also associated with epilepsy (Chenier, Yoon et al. 2014). CHD8 is a DNA helicase that is involved in chromatin remodelling at chromatin boundaries by enhancer-blocking (Thompson, Tremblay et al. 2008) and mutations in CHD8 have been implicated in ASD (Neale, Kou et al. 2012) . CHD5 is associated with neuroblastoma (Fujita, Igarashi et al. 2008) and CHD7 is associated with the neuro-developmental disorder CHARGE syndrome (Vissers, van Ravenswaaij et al. 2004). CHARGE syndrome is a congenital pleiotropic disease and 50% of CHARGE syndrome patients have mutations in the CHD7 gene.

1.20 Methods for determining the positions of nucleosomes.

Micrococcal nuclease (MNase) is both an endo and an exonuclease that was isolated from the bacterium *Staphylococcus aureus*. It digests single-stranded, double-stranded DNA and RNA

(Tucker, Hazen et al. 1978). Early work using MNase to digest rat liver chromatin showed that limiting digestion generated 185 bp DNA fragments and further digestion generated 150-160 bp fragments (Axel 1975). This suggested that limiting MNase digestion of chromatin cleaves DNA in the linker region between nucleosomes and that further digestion is at the boundary of protection of the DNA protected by the core nucleosome. In addition it showed that it is possible to detect positioned nucleosomes in higher eukaryotic cells.

Mapping the positions of nucleosomes by indirect end-labelling was developed by Nedospasov *et al* (Nedospasov and Georgiev 1980) using MNase in SV40. In this method, MNase and a restriction enzyme are used to digest chromatin. The ladder of DNA fragments that are protected by the nucleosome core are separated from the histone proteins that form the nucleosome core and then size fractionated by electrophoresis. The region of interest in the genome is detected using a radio-labelled probe. When the cleavage patterns of chromatin are compared with that of naked DNA, the regions where proteins are bound to DNA can be mapped. The region of DNA protected by the nucleosome generates a nucleosome 'footprint'. Thus genomic foot-printing allows the one-dimensional analysis of a complex three-dimensional structure by mapping nuclease cleavages in DNA. This method has been further developed in studies of *S. cerevisiae* chromatin to map nucleosomes at individual genes (Kent and Mellor 1995).

There are several methods available for mapping the genomic position of chromatin protein complexes. ChIP-seq (Chromatin immuno-precipitation followed by sequencing) is a method whereby DNA-binding proteins are crosslinked to DNA *in vivo*, the DNA is sheared by sonication and then the DNA fragments of interest are purified by immuno-precipitation and the purified DNA fragments are sequenced (Jiang and Pugh 2009; Park 2009).

Recent developments in DNA sequencing technology have made it possible to study the chromatin structure of entire genomes. Now it is possible to sequence all of the fragments of DNA in the genome that are protected by binding chromatin proteins using MNase-seq (Fig 1.9). After sequencing these fragments using high throughput sequencing techniques, the sequencing reads of the DNA fragments protected by chromatin particles can be aligned back to the positions in the genome from whence they came. This generates genome-wide information about the distribution and localization of chromatin species. Next-generation nucleosome-sequencing methods have been developed using either cleavage of chromatin with micrococcal nuclease (Floer, Wang et al. 2010; Kent, Adams et al. 2011; Zhang and Pugh 2011), or by using hydroxyl radicals (Brogaard, Xi et al. 2012) followed by next-generation sequencing, either by single or paired-end sequencing (Lieleg, Krietenstein et al. 2015).

When the DNA fragments generated by limiting MNase digestion of chromatin are separated by electrophoresis, a 'nucleosome ladder' is visible; each band represents a multiple of 147 bp, the amount of DNA protected by a core nucleosome. Hence, when chromatin is digested using MNase *in vivo*, it is possible to create a DNA ladder that contains all of the DNA fragments in the genome that are protected by binding chromatin proteins. By separating the data into size categories of DNA fragments, it is possible to construct several maps of the locations of chromatin species of different sizes. For example, DNA fragments in the 50-100 bp size range might represent fragments protected by transcription factors and DNA fragments in the 150-200 bp size range might represent fragments protected by mono-nucleosomes (Henikoff, Belsky et al. 2011; Kent, Adams et al. 2011). Thus, high resolution genome-wide maps of the positions of nucleosomes can be constructed.

Much of the information from these large scale studies is available in central repositories such as the UCSC genome browser (genome.ucsc.edu), ENCODE (The Encyclopaedia of DNA Elements) and mouse ENCODE (www.mouseencode.org) and GENCODE (Encyclopaedia of genes and gene variants www.gencodegenes.org). As more data is being generated, experimental protocols are being standardised and more sophisticated methods of data analysis are being developed. The epigenome roadmap project (www.roadmapepigenomics.org) is a catalogue of human epigenomic data with the aim of aiding biology and disease-oriented research.

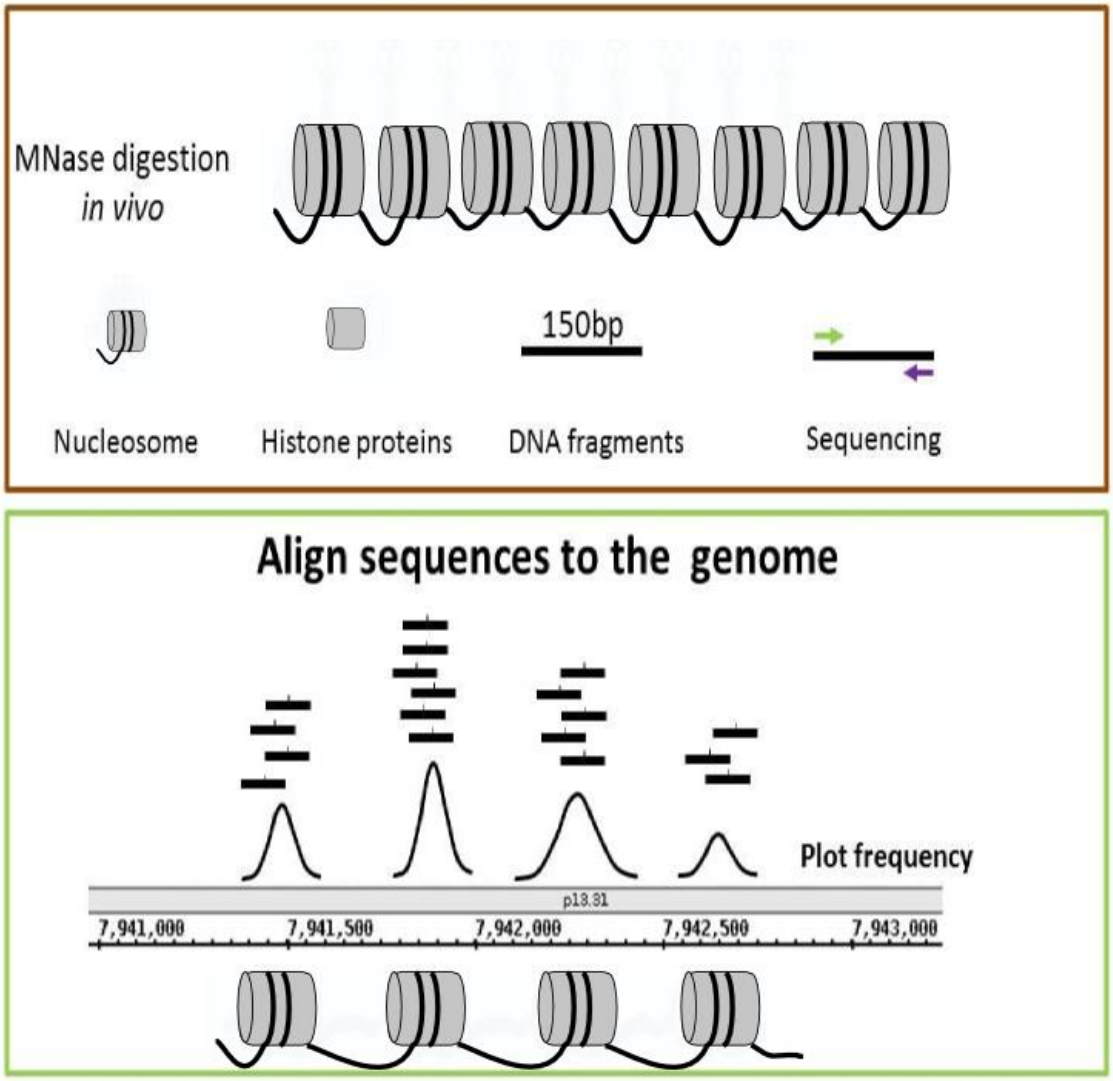


Figure 1.9 MNase-seq methodology

Chromatin is digested using micrococcal nuclease (MNase). The resulting DNA fragments that were protected by chromatin particles are purified and sequenced using an Illumina paired-end HiSeq protocol. Reads are aligned to the genome and histograms of the read frequency distributions are constructed. The peaks obtained in the histogram represent positioned nucleosomes.

1.21 Aims of this thesis

To date, genome-wide changes in chromatin structure that occur during development in human cells have not been investigated widely. The advent of high throughput genome-wide sequencing has greatly enhanced the study of both genetics and of gene regulation. Now it has become possible to use next generation sequencing to map nucleosome positions at high resolution on a genome wide scale using MNase-seq in the higher eukaryotic genomes from any cell type. Therefore it is possible to construct and compare genome-wide chromatin maps from stem cells and differentiated cells derived from them. Deciphering the relationship between changes in chromatin structure and functional changes in genome regulation will reveal mechanisms of regulation during development. This will lead to developing a greater understanding of how mis-regulation of chromatin organisation causes human disease.

MNase-seq defines both the genomic position and the size of nuclease-resistant chromatin species, to map chromatin structure (Kent, Adams et al. 2011). This methodology has been used in the work of this thesis to construct nucleosome maps from human cells. The aim of this thesis was to construct and analyse genome-wide chromatin maps from both undifferentiated human iPS cells (hiPS) and iPS cells differentiated to neural progenitor cells (NPC). By comparing the nucleosome maps from pluripotent cells and cells developed to the early stage of neuronal differentiation, those areas of chromatin structure that might be involved in the regulation of neurodevelopment would be identified. This may identify the regulatory regions of the genome that are targeted by chromatin remodelling proteins during neurodevelopment.

The work from this thesis will answer the following questions:

1. Are nucleosomes randomly distributed or strategically placed in the human genome?
2. Does chromatin structure change during neural development?
3. Does chromatin structure change at regulatory sites during neural development?
4. Can I identify genes that are targeted for regulation during neural development?

Chapter 2 : Materials and Methods

2.1 Biological sample preparation

Cell cultures were prepared by Shona Joy and Prof. Nick Allen.

2.1.1 Cell culture

The 34D6, male, human iPS cell line (a gift from Prof Chandran, Edinburgh) was cultured in mTeSR1 (Stem Cell Technologies) following the manufacturer's instructions. Briefly, tissue culture plates (Nunclon, Invitrogen) were coated for at least 2 hours with MatrigelTM (BD Biosciences) diluted 1:75 with DMEM (Invitrogen). 34D6 iPS cells were plated at a density of ~10^6 cells/10cm plate in complete mTeSR1TM medium containing 10µM Y27632 (Tocris) and incubated at 37^{o}C in a standard 5% CO_2, humidified incubator (Binder). To passage iPS cells, cells were first treated with 10µM Y27632 for 2hrs, cells were then washed with Ca^{2+}/Mg^{2+}-free PBS and cell colonies lifted from the plate by incubation with Dispase (Stem cell technologies) containing 10µM Y27632 (Tocris). Colonies were then fragmented by gentle trituration, collected by centrifugation (180 x g) and resuspended in medium for re-plating.

2.1.2 Dual SMAD inhibition

For neural differentiation a dual SMAD inhibition protocol was used (Chambers, Fasano et al. 2009). 34D6 iPS cells were harvested as described above for iPS cell passaging. 34D6 iPS cell colony fragments were plated in non-adherent bacteriological grade culture dishes in ADF differentiation medium to allow for embryoid body formation. ADF differentiation medium comprised advanced DMEM/ F12™ media (Invitrogen) supplemented with penicillin/streptomycin (5µg/l, Invitrogen), L-glutamine (200mM, Invitrogen), 1x Lipid concentrate (Invitrogen), 7.5µg/ml holo-transferrin (Sigma), 14µg/ml Insulin (Roche), 10µM β-mercaptoethanol (Sigma). Medium was supplemented with 10µM Y27632 (Tocris) for the first 2 days, with 10µM SB-431542 (Tocris) until day 4 of differentiation and 0.5µM LDN193189 (Miltenyi) until day 8 of differentiation. Medium was changed every 2 days. At day8, neuralised embryoid bodies were washed with Ca^{2+}/Mg^{2+}-free PBS and then dissociated by incubation at 37^{o}C with accutase (PAA laboratories). A single cell suspension was obtained by gentle trituration and cells washed with ADF medium and harvested by centrifugation at 1000rpm. Neural progenitors were then plated onto tissue culture plates coated with 0.1µg/ml Poly-L-Lysine (Sigma) and 10µg/ml Laminin (Sigma) in ADF medium + 5ng/ml FGF2. 34D6-derived neural progenitors were grown to sub-confluency and passaged once by dissociated with accutase and re-plating. The 34D6 iPS cells derived neural progenitors were harvested for nucleosome preparation on day 16 of differentiation.

The NPC cell population was validated using immunocytochemistry by checking for the presence of the NPC-specific markers, such as Nestin and the loss of pluripotency markers, for example Oct-4. Greater than 90% of the differentiated population contained neural progenitors.

2.2 DNA extraction and sequencing

2.2.1 *In vivo* MNase digestion of chromatin

Chromatin was prepared from human iPS cells and NPC cells as described previously for the *S. cerevisiae* genome (Kent, Adams et al. 2011). *In vivo* MNase digestion of chromatin and subsequent DNA extraction from human iPS cells and the same cells differentiated to NPC cells was undertaken by Dr. Nicholas Kent. Three bio-replicates for each cell type, induced pluripotent (iPS) and derived neural progenitor (NPC) cells were utilised for MNase-seq. The samples of cultured cells, grown independently under the same conditions were also treated independently with MNase. The cell membranes and nuclei were made permeable to MNase using the detergent NP-40 (Stewart, Reik et al. 1991) (Kent and Mellor 1995) and treated by *in vivo* digestion with 300U/ml MNase at room temperature for 4 minutes. DNA fragments were purified from the MNase treated samples and then the samples for each cell type were pooled in equimolar amounts. DNA extracted from chromatin samples was size-fractionated on agarose gels. 25-35ug of DNA less than 300 bp was size-selected for each cell type (Fig 3.1).

2.2.2 Illumina HiSeq 2000 platform (HiSeq) paired end mode sequencing.

Paired-end mode sequencing was undertaken by Source BioScience (http://www.sourcebioscience.com/). Paired end mode sequencing of size-fractionated DNA was performed on a Illumina HiSeq 2000 platform (HiSeq) using a read-length of 50 bp. Eight flow cells were used for each cell type to obtain a sufficient depth of coverage of the human genome. A standard Illumina paired-end mode sequencing protocol was used, apart from the omission of the nebulisation step and the addition of a further gel purification step to eliminate any excess concatenated linkers after the ligation of linker DNA to the sample. Base calling and quality control of the sequencing data was performed using Real Time Analysis (RTA) 1.09, CASAVA 1.8 software.

2.3 Analysis of paired-end DNA sequence data

2.3.1 Alignment of paired-end reads to the genome

A total of 3.4 and 3.0 billion paired-end reads were obtained in fastq format from iPS and NPC cells respectively. The reads from individual flow cell lanes were aligned to human genome

assembly hg19 using Bowtie version 0.12.8 (Langmead, Trapnell et al. 2009). The command line options for bowtie were as follows: bowtie -v 3 --trim3 14 --maxins 5000 --fr -k 1 --best -p 12. The maximum insert size was set at 5000 to obtain the maximum range of available particle sizes in the chromatin sample and the read length was clipped to 36 bp to remove any adaptor sequences. The bowtie –fr command ensures that alignments are only called when mate pairs have a relative orientation of FR ("forward stand, reverse strand") and the insert size condition specified is met. A total of 2.5×10^9 aligned paired-end reads were obtained for the pluripotent genome and 2.25×10^9 aligned paired-end reads for NPC.

In this method, all of the DNA fragments less than 300 bp in size, generated after MNase digestion, were utilised for sequencing. As these data contain a variety of lengths of DNA fragments, generated by the protection from MNase digestion by chromatin species, rather than fragments generated by strict size-selection, the selection of insert sizes for further analysis was undertaken computationally. This meant that other published software packages, for example Danpos (Chen, Xi et al. 2013) were not suitable for nucleosome mapping. Hence, a bespoke data processing pipeline was developed. Subsequent data processing was undertaken in a similar manner to the method previously described for the *S. cerevisiae* genome (Kent, Adams et al. 2011). This method was adapted and automated using high performance computing to process the human genome data which is a much larger data-set than that generated from the *S. cerevisiae* genome. Data was processed according to the flowchart (Fig 2.1). In house Shell scripts were used to create chromosome and particle-specific directories and to move files into them as required. All the scripts used in this analysis are available on a DVD.

2.3.2 Data validation and normalisation.

The number of reads obtained for each human chromosome in each cell type was determined using the scripts: count_reads_5 and count_reads_6. These scripts count the number of lines in the chromosome specific .sam files, hence generating the number of aligned paired-end reads for each chromosome. The results are shown in the Appendix. Fig A.2. As the ratio of reads obtained between iPS and NPC cells in the Y chromosome was 0.6, indicating a substantial difference in read counts obtained from the Y chromosome between the two cell types, the data from the autosomal genome was utilised for further analysis.

To compensate for the slight genome-wide difference in total read counts between the two cell types, the read counts for the NPC autosomal genome were multiplied by the ratio of the total aligned reads in the autosomal genomes of iPS v NPC cells, which was 1.117. (See Appendix table A.1)

2.3.3 Analysis of the number and distribution of chromatin particle sizes.

The distribution of paired-end read insert sizes binned at single base pair resolution was determined by concatenating the chromosome-specific sam files and processing the data using the scripts: pair_read_histogram_ARCCA_scratch_multi, add_pair_histo_iPS4_autosomes and add_pair_histo_NPC2_autosomes.

Normalised data from the autosomal genome was used to construct the frequency distribution presented in (Fig 3.2).

2.3.4 Division of the paired-end read data into chromatin particle size classes

Paired-end reads obtained from Bowtie alignments in .sam format were sorted into separate chromosome-specific files using the UNIX 'grep' command in the shell script: human_NIPS_chr_grepv2.sh. Edited versions of this script were used for iPS cell data processing. The resulting chromosome-specific .sam format files were processed using the script SAM_parserJH.pl. This script calculates the mid-point of each insert and separates the reads according the insert size (ISIZE). The reads were filtered according to the insert size plus or minus a window value ($pwind). The pwind value was set as shown in Table 2.1. The pwind values that were set resulted in three non-overlapping ranges of putative chromatin particle sizes. Hence, nucleosomes are represented by the 138-161 bp chromatin particle size class. Other user-defined parameters that were set in this script were the $SAM_ID_flag variable which was specific to each .sam file and the chromosomes, specified in the array @chromosome. All of the .sam files in chromosome-specific directories were processed and the resulting particle-specific sam files were output into new chromosome-specific directories.

The script cat_part was used to concatenate the replicate chromosome-specific files for each particle size, and then the script part_sort_pwind.pl was used to move the concatenated files for each chromosome into whole genome, particle-specific directories.

Particle size class	Lower limit	Upper limit	pwind
Sub-nucleosomal	112.5	137.5	0.1
nucleosomal	138.75	161.25	0.075
Super-nucleosomal	161.875	188.125	0.075

Table 2.1 Summary of the pwind values and the resulting non-overlapping chromatin particle size classes

The value of pwind in the script SAM_parserJH.pl results in three non-overlapping chromatin particle size classes of approximately 112-137 bp, 138-161 bp and 161-188 bp herein referred to as sub-nucleosomal, nucleosomal and super-nucleosomal chromatin particles respectively.

2.3.5 Construction of normalised read mid-point frequency distributions.

To represent a unique chromatin particle position for each paired-read, the convention of marking the genomic position of the mid-point of the insert DNA was followed. This convention is designed to represent a position equivalent to the nucleosomal dyad (Luger et al., 1997). The script humanhistogramXY_2.pl was used to calculate the frequency distribution for the mid-point position of the paired read insert size values counting reads at 10 bp resolution (defined in the variable $bin_width). To reduce random variation, the data was smoothed using a 3bin moving average.

Frequency distributions were output as chromosome-specific files in .sgr format. Sgr files are tab-delimited graph files in the format: chromosome identification: chromosomal location of the start of each 10 bp bin: frequency of the paired-read midpoint values that fall within that 10 bp bin. These chromosome-specific files can be can be loaded into the Integrated Genome Browser (IGB) (Nicol, Helt et al. 2009), so that the frequency distribution of the paired-read midpoint values can be visualised as a histogram aligned with the genome. This provides information about the location and distribution of the user-selected particle sizes in the genome. The frequency distributions for all of the autosomal .sgr files for all three chromatin particle size classes in NPC cells were normalised for read depth as described in section 2.3.2 using the series of scripts: NPC_pwind_scale_factor.

2.4 Processing data from published human nucleosome maps

Maps derived by Kunjale *et al* (Kundaje et al. 2012) from the K562 chronic myelogenous leukaemia (Cml) cell line (Lozzio and Lozzio 1979) and from the B-lymphoblastoid cell line GM12878 (Corriell biorepository) were made by sequencing DNA purified from chromatin in the mono-nucleosome size range after MNase digestion. These published data-sets were converted from bigwig to bedgraph format using the script: bigWigToBedGraph (downloaded from the UCSC genome browser). Chromosome-specific bedgraph files were constructed using the shell scripts: human_chr_grep_word_boundary_stanford_Gm12878.sh and human_chr_grep_word_boundary_stanford_K562.sh. The chromosome-specific bedgraph files were then converted to .sgr files (Scripts Gm12878_bedgraph_to_sgr_v5 and K562_bedgraph_to_sgr_v5). These scripts bin the data into 10 bp bins and calculate a 3 bin moving average exactly as described for the iPS and NPC data above.

To validate and determine the quality of the final re-processed maps, the final chromosome-specific .sgr files were analysed using the scripts sgr_analyser_K562 and sgr_analyser_Gm12878. This generated the total numbers of aligned paired -end reads obtained for each chromosome and hence for each genome.

The positions of all of the peaks in the K562 and GM12878 genomes were determined using the peak algorithm (PeakFinder), script: peak_finder43.

2.5 Construction of genomic feature lists

2.5.1 TSS

The positions of transcriptional start sites for all of the full length transcripts for human genome build Feb 2009 GRCH37/hg19 were derived as follows. Track = Gencode Genes v17, table =basic, was downloaded from the UCSC genome browser generating the file: Gencode_basic_TSS_v2.txt (contains 94,151 TSS). Two files were generated from the UCSC data download: a) The file gencode_TSS_genes.txt which contains annotation information for all of the genes extracted from the UCSC browser was generated using the script extract_GENCODE_TX_start and b) a non-redundant file of strand-specific TSS (Gencode_unique_TSS_v5.txt, no of non-redundant TSS = 83,142) was generated using the script Gencode_TSS_Duplicates_extract_v5.pl. The output from this script was split into chromosome-specific files.

2.5.2 Transcription factor binding motifs.

The genome-wide binding positions for a select group of transcription factors were obtained. The sequence of the consensus binding motif used in this study for each transcription factor is shown in Table 2.2. The genomic positions of the consensus binding sequence for each transcription factor were extracted from fasta files from the hg19 genome assembly using in house perl scripts. Fasta files were downloaded from
http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/
This directory contains the Feb. 2009 assembly of the human genome (hg19,
GRCh37 Genome Reference Consortium Human Reference 37 (GCA_000001405.1))
Since the ATF2 consensus binding sequence is a palindrome, only forward strand motif positions were obtained. For the other motifs, forward and reverse strand positions were obtained and combined to generate a single file of binding positions for each transcription factor. All perl scripts used to extract the transcription factor binding positions are in the folder TF_BINDING_SITE on the DVD.

The consensus binding motif for YY1 was derived from ChIP-seq data derived from H1ESC cells in factorbook (http://www.factorbook.org) (Wang, Zhuang et al. 2013) which were derived by taking the sequences of the top 500 TF ChIP-seq peaks. In this study, the motif M1 was used from factorbook. The ATF2 CRE binding motif was taken from Hai et al (Hai, Liu et al. 1989). The PAX6 consensus binding sequence was derived using (ChIP) in ES-derived neuroectodermal cells (NECs) (Bhinge, Poschmann et al. 2014). CTCF binding positions were derived by matching the H1ESC ChIP region data from the Broad institute downloaded from the UCSC genome browser, file wgEncodeAwgTfbsBroadH1hescCtcfUniPk.narrowPeak (UCSC accession wgEncode EH000085) with the consensus binding motif derived by Ong *et al* (Ong and Corces 2014).

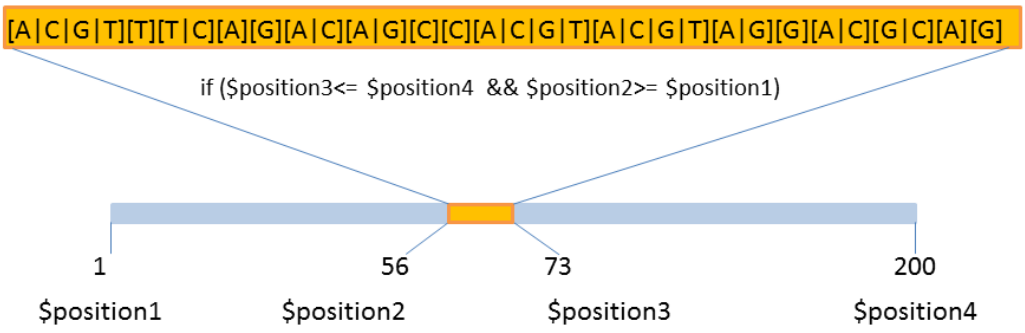
RE1 sites were derived by matching the consensus binding motif derived by Bruce *et al* (Bruce, Donaldson et al. 2004) with ChiP data that was derived from the ENCODE database (Gerstein, Kundaje et al. 2012) and downloaded using the UCSC genome browser from human genome assembly Feb 2009 GRCH37/hg19 Group=regulation, Track name = TXnFactorChiP, Table= wg EncodeRegTfbsCLusteredv2 from uniform processing of data from the Jan. 2011 ENCODE data freeze (Fig 2.1).

Transcription factor	Consensus binding motif	Motif length	No. of motif	No. consensus motifs matched with ChIP fragment.	Reference
ATF2	[T][G][A][C][G][T][C][A]	8	9881	NA	(Hai, Liu et al. 1989)
YY1	[A][A][G|A|C][A][T][G][G][C][G|C|T][G|C|A][C]	11	38,945	NA	(Wang, Zhuang et al. 2013)
PAX6	[A][T][T][C][A][T][G][C][A|C|G|T][T][G][A]	12	1432	NA	(Bhinge, Poschmann et al. 2014)
CTCF	[C][C][A|G][C|G][C|T][A][G][G|A][T|G][G][G][C|T]	12	19,299	9516	(Ong and Corces 2014)
RE1	[A|C|G|T][T][T|C][A][G][A|C][A|G][C][C][A|C|G|T] [A|C|G|T][A|G][G][A|C][G|C][A][G]	17	3889	871	(Bruce, Donaldson et al. 2004)

Table 2.2 See next page for legend.

Table 2.2 Generation of genome-wide positions of transcription factor motifs

The genome-wide binding motifs for a select group of transcription factors were determined. The binding motifs of ATF2 (Hai, Liu et al. 1989) and YY1 binding sequences from H1ESC cells in factorbook (Wang, Zhuang et al. 2013) and the PAX6 consensus binding sequence derived using (ChIP) in ES-derived NECs (Bhinge, Poschmann et al. 2014) were extracted from the hg19 genome using in house perl scripts.



Factor	Site locations	ENCODE ChIP regions[1]	Autosomal consensus RE1 motifs[2]	Autosomal sites within ChIP regions
NRSF/REST	Genome-wide	27386	3889	871

1. Gerstein, M.B et al. (2012).. Nature *489*, 91-100.
2. Bruce, A.W. et al , (2004). Proc Natl Acad Sci U S A *101*, 10458-10463.

Figure 2.1 Derivation of REST genomic binding positions.

Alignment of REST consensus binding motifs were aligned with ChIP data from the ENCODE project.

2.6 Development of a peak-finding algorithm

In order to locate and quantify the number of positioned chromatin particles in each size class, a peak-finding algorithm was developed. In order to set the parameters for the peak-finding algorithm, the quality of the final chromosome-specific .sgr files was assessed using the script: sgr_analyser_iPS_v3. The output from this script is two files; the first contains the sum of the paired-read midpoint values per 10 bp bin for each chromosome and the second contains the sum of the paired-read midpoint values per 10 bp bin for each chromosome below user-defined thresholds. In this way, the average number of paired-read midpoint values per 10 bp bin for the autosomal genome was determined for the nucleosomal chromatin particle size class (138-161 bp) in both iPS and NPC cells. The average number of reads per 10 bp bin for the iPS and NPC nucleosome maps was determined as 1.56 for the iPS map and 2.1 for the NPC map nucleosomal size class (138-161 bp). Therefore, a value of two was chosen to define the 'noise' threshold for subsequent peak searches. This analysis showed that there was a minority of 10 bp bins in which the frequency of the paired-read midpoint values was >1000 (632 bins in iPS and 806 bins in NPC) and most of the frequency distribution comprised 10 bp bins in which the frequency of the paired-read midpoint values was <50. These regions of the genome with high frequencies of paired-read midpoint values were, for example at the ends of chromosomes and at runs of repeats and at centromeres. As I could not rule out the fact that these high frequencies were an artefact of the repetitive nature of the underlying sequence,

these regions were excluded by setting the upper threshold for the centre of the peak in the peak-finding algorithm to 1000. Peaks were defined as three consecutive 10 bp bins where the value of the paired-read midpoint frequency in the central 10 bp bin was between 30 and less than 1000, given as A. The lower threshold for the paired-read midpoint frequency values for the bins either side of the central bin was set at two. The upper threshold for each of the paired-read midpoint frequency values for the bins either side of the central bin given as B, was less than the paired-read midpoint frequency in the central bin (A). The series of scripts called peak finder peak_finder43_iPS_150_pWIND2_hive and peak_finder43_NPC_150_scaled_hive were used to obtain the number and location of peaks representing positioned chromatin particles for all three classes of chromatin particle in this analysis.

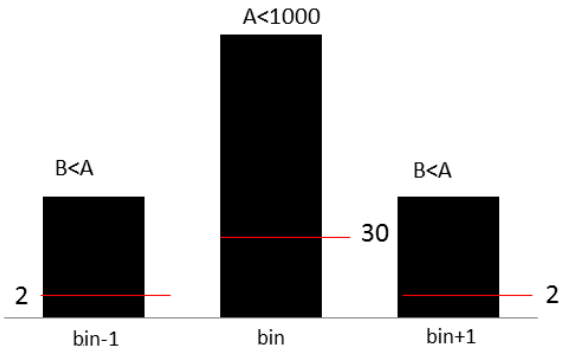


Figure 2.2 Defining a peak representing a positioned chromatin particle.

Numbering on the 10 bp bins represent the upper and lower paired-end read frequency thresholds used for each bin. The central bin has a low threshold of 30 reads and an upper threshold of <1000, defined as A. Each of the 10 bp bins either side have a low threshold of 2 and an upper threshold value defined as B, where B must be less than the value of A.

2.7 Locating and comparing patterns of positioned chromatin particles

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were given explicit genome positions using the heuristic peak-finding algorithm (peak-finder) as described in section 2.6. This was carried out for all of the chromatin particle size classes in iPS and NPC cells and the published human chromatin particle maps derived from K562 and GM12878 cells (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012). The scripts used for comparison of the locations of positioned chromatin particles are in the folder: PEAK_COMPARE on the DVD.

2.7.1 SiteWriter

The perl script SiteWriterJH_CFD_10bin_2.pl, (Kent, Adams et al. 2011) was used to construct normalised average cumulative frequency distributions of the paired-read midpoint values at and in the bins surrounding genomic feature loci within a user-defined window, for example at TSS and transcription factor binding sites. The chromosome-specific, chromatin-particle-specific .sgr files generated as described above were concatenated to form whole-genome-particle.sgr files to use in SiteWriter. The file containing the locations of the genomic features were generated as described in section 2.5.

The output from the SiteWriter script comprises two files: the first is a CFD.txt file of the normalised average cumulative frequency values that are output in tab-delimited text format. The values are normalised by dividing the cumulative frequency value in each bin by the number of bins specified in the user-defined window. The graphs derived from the CFD.txt file were plotted in excel or in R version 3.2.1 using the ggplot2 package. The second output file from SiteWriter, a C3.txt file, contains a matrix of locally-normalized dyad frequency values for every bin position. These data can be used in cluster analysis. See section 2.8.

2.7.2 Comparison of positioned chromatin particles from different size classes.

To compare the overlap in the locations of positioned chromatin particles within +/- 10 bp between all three chromatin particle size classes in the same cell type, the following scripts were used:

iPS_peak5_part_compare_v5_sitewriter_window_v6

NPC_peak5_part_compare_v5_sitewriter_window_v6_ARCCA.

The results of this analysis are shown in the Venn diagrams in Fig 3.9

The script ‘ diff_cell_peak5_part_compare_v5_sitewriter_window_v6_ARCCA’ was used to compare the overlap in the locations of positioned nucleosomal chromatin particles within +/- 10 bp in NPC, K562 and GM12878 cells. The results are shown in Fig A3 (Appendix).

2.7.3 Comparison of positioned chromatin particles in the same size class during differentiation from iPS to NPC.

To compare the overlap in the locations of positioned chromatin particles within +/- 10 bp in the same size class during differentiation from iPS to NPC the following scripts were used:

iPS150_v_NPC150_peak5_part_compare_v5_sitewriter_window_v6.
iPS175_v_NPC175_peak5_part_compare_v5_sitewriter_window_v6

2.7.4 Comparing positioned chromatin particles at TSS

The number and location of TSS with positioned chromatin particles within 300 bp of a TSS was determined using series of scripts called peak_TSS (e.g. iPS_peak5_TSS-UP or iPS_peak5_TSS-DOWN) for all three chromatin size classes. The scripts iPS_v_NPC_peak5_TSS_down_compare_v5 and iPS_v_NPC_peak5_TSS_up_compare_v5_new were used to determine, for the genes that possessed nucleosomes near the TSS, which genes were exclusive to one cell type and which were in common between iPS and NPC cells. In addition, the annotation information for these subclasses of TSS was obtained. These scripts are in the folder: TSS on the DVD.

2.8 Cluster analysis.

In order to determine whether there were common patterns of chromatin particle positioning frequency in a user-defined window, surrounding a genomic feature, cluster analysis was undertaken using data generated from the 'C3' files generated from SiteWriter. Cluster analysis was undertaken in R, using the Canberra method to generate a distance matrix from the 'C3' files generated as described in section 2.7.1. Several different methods for generating the distance matrix were utilised on test data: for example maximum, Euclidean and Manhattan. The Canberra method is similar to the Manhattan method, which sums the absolute differences between variables, but in the Canberra method the absolute difference between the variables of the two objects is divided by the sum of the absolute variable values prior to summing. The Canberra method generated the best data separation. Dendrograms generated by hierarchical agglomerative clustering were used to determine the number of groups to use in k-means clustering. Heat maps were generated from the $\log_2$ transformed data from the cluster groups using the heatmap2 function from the gplots package in R. All R scripts used for clustering and heat map generation are in the folder CLUSTER on the DVD.

2.8.1 Clustering chromatin particle positions.

As there were too many peak positions derived to use all of them in cluster analysis, randomly selected peak positions (n=20,000) were derived from the 'C3' files using the shell script 'shuf.sh' for each chromatin particle size class in both iPS and NPC cells. The UNIX function 'shuf', used in the script shuf.sh', extracts lines of text from a file at random. The output files from this are located in the folder called 'PEAK5_20K_Random_site_files' on the DVD. The peak positions for each chromatin particle size class and for both iPS and NPC were clustered using the 'R' scripts in the folder: Peak5_cluster_R_scripts on the DVD.

2.8.2 Clustering chromatin particle size class data at RE1 sites.

The iPS 150 .sgr data was processed using the script SiteWriter_Full_CFD_C3.plx and the list of RE1 sites generated in section 2.5.2 (filename: autosome_RE1_site_match_encode_chip_FR.txt). The data from the output C3.txt file was clustered using the script cluster_R_v3_RE1_iPS_150k_final.R.

411 of the RE1 sites (those in clusters C2, 5 and 7) in the iPS genome that possessed positioned nucleosomal (138-161 bp) chromatin particles surrounding them were grouped together to form a subset of the total number of RE1 sites named C257, filename: RE1_iPS-150_C257.txt. The remaining 460 sites, which in the iPS genome had less well-positioned chromatin particles surrounding them, were grouped together and named C12368, filename: RE1_iPS-150bp_C12368.txt.

Using the R script 're-cluster_R_RE1_NPC_150_v2.R'. The data for the C257 subset of RE1 sites in NPC cells was extracted from the NPC nucleosomal size class (138-161 bp) 'C3' file. These data were clustered into four groups, generating the files:

RE1_iPS150_C257_NPC_C1_sitewriter.txt
RE1_iPS150_C257_NPC_C2_sitewriter.txt
RE1_iPS150_C257_NPC_C3_sitewriter.txt
RE1_iPS150_C257_NPC_C4_sitewriter.txt

These cluster site files were re-grouped according to the chromatin structure in each cluster: C1, 3 and 4 had well-positioned chromatin particles surrounding the RE1 site in the nucleosomal chromatin particle size class (138-161 bp) from NPC cells and were grouped together in the file: RE1_iPS150_C257_NPC_C134_sitewriter.txt

2.8.2 Clustering chromatin particle size class data at CTCF sites.

The sub-nucleosomal (112-137 bp) size class data from iPS cells was processed using SiteWriter and the file of CTCF sites derived in section 2.5.2: filename: autosome_CTCF_ONG_site_match_H1ESC_chip_sitewriter.txt (n=9516). Data from the 'C3' file generated was clustered using the R script: R_v3_iPS_125_CTCF_H1ESC_ARCCA_v2. The resulting output files were used to generate average cumulative frequency distributions in the sub-nucleosomal particle size class, for each cluster, using SiteWriter.

The cluster site files were concatenated in order of the strength of the signal in each cluster follows: Cluster5: Cluster4: Cluster1: Cluster2: Cluster6: Cluster3 to form the file: iPS_125_CTCF_clus_ordered_sites.txt. This file of ordered CTCF binding sites was used to extract the data from the C3 files in the same order for the nucleosomal and super-nucleosomal particle size classes in iPS cells and for the sub-nucleosomal, nucleosomal and super-nucleosomal particle size classes in NPC cells, using the perl script: CTCF_cluster_order.pl. The output files from this are in the folder: "ORDERED _CTCF_C3_data". These ordered data sets were used to create heat maps for each of the chromatin particle size classes using the R script: CTCF_ONG_cluster_ordered_big_heatmaps_v3.R.

2.9 Extraction of REST target gene annotation.

Annotation information for RE1 sites was extracted from the annotation file 'gencode.v18.annotation.gtf' downloaded from GENCODE: www.gencodegenes.org using the series of scripts in the folder: RE1_ANNOTATION on the DVD.

Chapter 3 : Generation and analysis of chromatin maps

3.1 Introduction

The aim of this chapter was to construct and characterise genome-wide nucleosome maps using MNase-seq (Kent, Adams et al. 2011), from human induced pluripotent cells (iPS) in their undifferentiated state (Takahashi, Tanabe et al. 2007; Yu, Vodyanik et al. 2007) and iPS cells developed to the neural progenitor stage (NPC) (Chambers, Fasano et al. 2009). The analysis and comparison of these maps would reveal the genome-wide changes in patterns of chromatin structure that occur during the earliest stages of neural cell development.

Other genome-wide chromatin maps have been constructed from various human cell types (Schones, Cui et al. 2008; Gaffney, McVicker et al. 2012; Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012), but this the first study of high-resolution chromatin maps from human cells at different stages of neural cell differentiation. The work in this chapter details the construction, validation and comparison of these maps. This work shows that only a small proportion of the human genome contains highly positioned nucleosomes and that the frequency of positioned nucleosomes in the genome increases during neuro-development. This suggests that there is a high degree of genome regulation during neural differentiation.

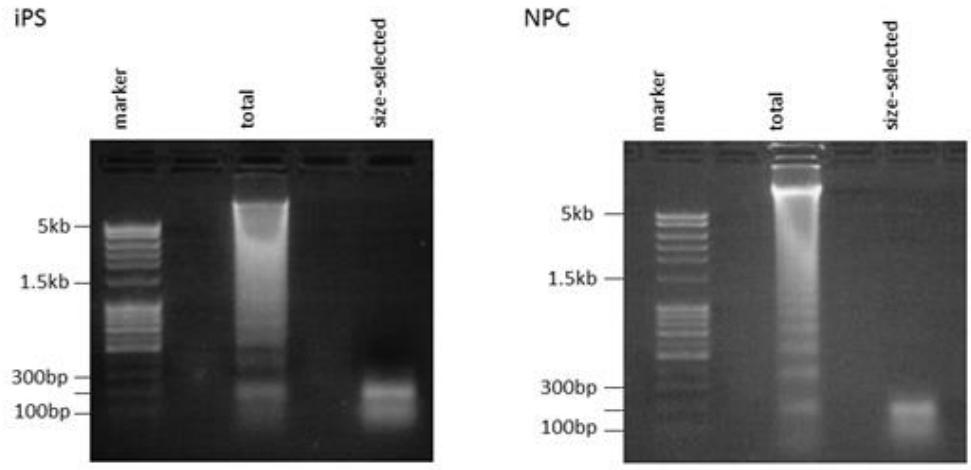
3.2 Generation of size-selected DNA samples

Undifferentiated human iPS cells, herein referred to as iPS cells and the same cells developed to NPC were used to construct genome-wide MNase-resistant chromatin particle maps. The cells were cultured and differentiated by Shona Joy and Prof. Nick Allen. Subsequent processing of the cells in order to obtain DNA samples for sequencing was undertaken by Dr. Nick Kent. In brief, three bio-replicates for each cell type were treated by partial *in vivo* digestion with MNase and DNA fragments were purified from them and pooled in equimolar amounts. The DNA fragments were separated according to size by gel electrophoresis. Figure 3.1A shows that a 'nucleosome ladder' is visible in the samples from both cell types. The nucleosome repeat length for bulk chromatin for both iPS and NPC was calculated by plotting calibration curves of the log of the molecular size of the DNA in base pairs as a function of the migration distance of the DNA in the gel (Appendix Fig A.1). In order to determine the average nucleosome repeat length, the size of the furthest dominant migrating band in the MNase ladder, representing the DNA protected by nucleosomes, was determined. A nucleosome repeat length of 200 bp was obtained in both cell types. This is similar to the human nucleosome repeat lengths obtained previously in other cell types, for example in HeLa cells, 190-200 bp (Whitlock and Simpson 1976), 193 bp in granulocytes, 203 bp in CD4+ cells (Valouev, Johnson et al. 2011) and 187 bp in lymphoblastoid cells (Gaffney, McVicker et al. 2012).

Since the human genome is large (approximately $3x10^9$ base pairs), MNase-resistant DNA fragments of less than 300 bp were size-selected from each cell type from the complete nucleosome ladders using preparative gel electrophoresis. This size-selection was necessary in order to be able to obtain a sufficient read-depth of the human genome (approximately one read per base pair of DNA) (Kent, Adams et al. 2011) in the size range of DNA that would be protected from MNase by human nucleosomes, within the cost constraints of this project. Nevertheless, this necessitated using eight HiSeq flow cell lanes for each cell type. However, by sequencing all of the fragments below the approximate size of a di-nucleosome (300 bp), this allowed the study of MNase resistant species, represented by DNA fragments, that were both smaller and larger than the size of core nucleosome that footprints approximately 150 bp of DNA. This strategy was taken so that it might be possible to a) detect the locations in the genome protected from MNase digestion through the binding of transcription factors, by analysis of the mapped MNase resistant fragments that were smaller than the size of a core nucleosome (150 bp) and b) to analyse locations in the genome protected from MNase digestion by the presence of chromatin species that were larger than a core nucleosome (approximately 150 bp). This strategy would cover the analysis of MNase resistant species in the size range of transcription factor complexes, human core-nucleosomes and those in the size range of human chromatosomes.

Fig 3.1A shows the total and size-selected DNA samples after MNase digestion of chromatin in both iPS and NPC cells. Size-selected DNA pools from each cell type were sequenced in paired-end mode using Illumina technology as described in Chapter 2. In total, the pooled iPS cell sample generated 2.5×10^9 aligned paired-end reads and the NPC sample generated 2.25×10^9 aligned paired-end reads (Fig3.1B). The resulting DNA fragments purified from both iPS and NPC represent regions of genomic DNA that are protected from MNase digestion by the binding of chromatin proteins and perhaps transcription factors and herein they are referred to as chromatin particles.

A



B

Sequencing reads(x10^9)	iPS	NPC
Total reads processed	3.4	3.0
Aligned paired-end reads	2.5	2.25

Figure 3.1 MNase digestion of chromatin in iPS cell and NPC cultures yields nuclease protected DNA suitable for size selection and analysis by paired-end mode Illumina sequencing.

DNA was extracted from 3 pooled bio-replicates of iPS cells and iPS cells developed to the neuro-progenitor stage (NPC) A. Ethidium bromide stained agarose gel showing the size distributions of MNase-resistant DNA species before and after size-selection. DNA fragments less than 300 bp in size were selected for Illumina paired-end sequencing. B. The total number of sequencing reads obtained and the number of aligned paired-end sequencing reads obtained from the Illumina sequencing for both cell types cells is indicated in the table.

3.3 Construction of MNase-resistant chromatin particle maps.

3.2.1 Data validation

The paired-end reads were aligned to the human genome (assembly hg19) using Bowtie 0.12.8 (Langmead, Trapnell et al. 2009) and the resulting .sam files of aligned reads were examined to determine the quality and comparability of the data obtained from each cell type. The number of aligned reads obtained for each human chromosome in each cell type was determined (see Appendix Fig A.2). The number of paired-end reads obtained for each chromosome in each cell type is broadly similar in most of the chromosomes. However, the ratio of reads obtained between iPS and NPC cells in the Y chromosome is only 0.6, indicating a substantial difference in read counts obtained from the Y chromosome between the two cell types. For this reason, it was decided to use the data from the autosomal genome for further analysis. Although there was only a small difference in total read counts between the two cell types, it was decided to normalise the read counts, by multiplying the reads for the NPC autosomal genome by the ratio of the total aligned reads in the autosomal genomes of iPS v NPC cells.

To assess the number and size distribution of chromatin particle sizes for the autosomal genome as a whole, the frequency distribution of the total number of aligned paired-end reads obtained from both cell types with respect to insert size was determined at single base pair resolution. These data were plotted as a frequency distribution (Fig 3.2). This generated an estimate of the number and distribution of chromatin particle sizes in both cell-types. The distributions from both cell types were broadly similar, confirming a comparable level of MNase digestion. In particular, the number of DNA fragments in the 140-160 bp range matched closely between the two cell types. This is the size range of DNA that would be protected from MNase by core nucleosomes (147 bp). However, there was a difference in the distribution of chromatin particle sizes at the extreme ends of each distribution. More chromatin particles were observed in the lower end of the distribution in NPC than in iPS cells. For example, in the 50 bp chromatin particle size class, there were seven-fold more particles in NPC than in iPS cells. In contrast, there were more chromatin particles in the upper end of the distribution in iPS cells than in NPC. For example, in the 300 bp particle size class, there were seventeen fold more particles in iPS cells than in NPC cells. This result might suggest a difference in nucleosome structure between the two cell types. If nucleosomes were more accessible to MNase in NPC cells, then there would be a greater number of chromatin particles in the smaller size range.

Hence, it was decided that fragments in the 110 bp - 190 bp range, where the difference in the read count frequency between the two cell types is minimal would be used for subsequent analysis.

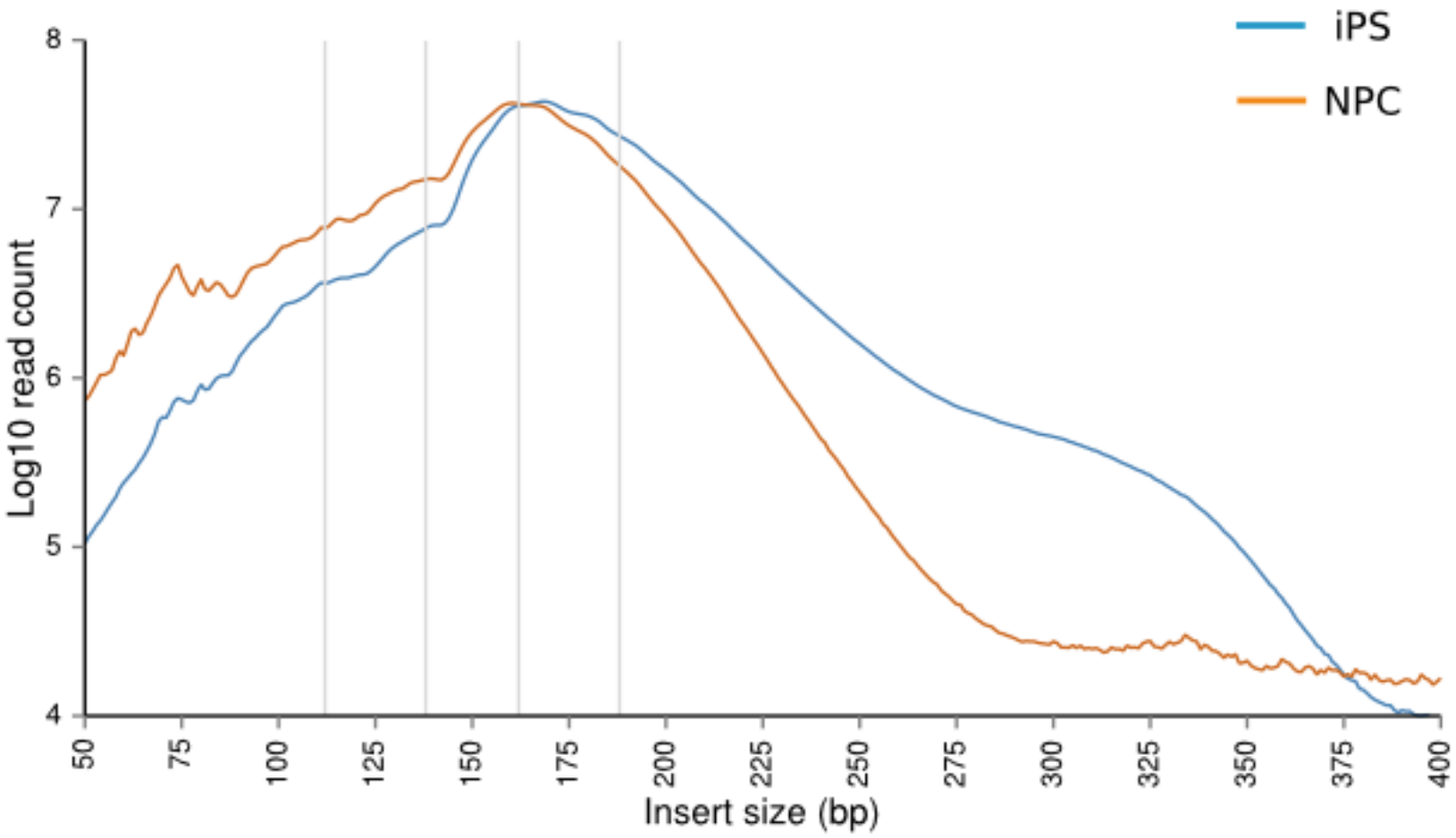


Fig 3.2 See next page for legend

Figure 3.2 Sequence reads derived from MNase resistant DNA in iPS cell and NPC cultures show broadly similar paired-read insert size distributions.

The number and size of the aligned paired-end sequencing reads from the genomes of both iPS and NPC was determined single base pair resolution. Read counts for the autosomal NPC genome were normalised (reads x 1.117.) Data from the autosomal genomes of both cell types was plotted as a frequency distribution of the $\log_{10}$ of the number of aligned paired-end reads. Data in the 110-190 bp size range was divided into three size classes (112-137 bp, 138-161 bp and 162-188 bp) for use in later analysis, shown by the vertical grey lines.

3.2.2 Division of the chromatin particle data into three size classes.

Aligned reads with insert sizes in the 110 bp-190 bp range were processed further as described in the flow chart (Fig 3.3), using a series of in-house Perl and shell scripts to generate chromosome-specific and genome-wide frequency distributions for the paired-end read mid-point positions of MNase-resistant chromatin particles. To be able to define the positions of chromatin particles of different sizes in the human genome precisely, it was decided to divide the 110-190 bp chromatin particle size range data obtained into three non-overlapping size classes. These were 1) 112-137 bp (125 bp +/- 10%), herein referred to as representing sub-nucleosomal particles 2) 138-161 bp (150 bp +/- 7.5%) herein referred to as representing core nucleosomal particles and 3) 162-188 bp (175 bp +/- 7.5%) herein referred to as representing super-nucleosomal particles. This would mean that it would be possible to detect core nucleosomal sized (approximately 150 bp) chromatin particles in the 138-161 bp size class. In addition, it might be possible to detect chromatin particles that might represent transcription factors or transcription factor complexes in the 125-137 bp size class. Finally in the 162-188 bp size class, it might be possible to detect chromatin particles that generate a DNA footprint that is larger than that of a core nucleosome, for example those generated by modified nucleosomes or chromatosomes.

Chromatin particle maps for each size class were generated by plotting the frequency distributions of the mid-point position of the paired-read insert size values on to the genome for each DNA fragment. Hence, the output from the processing pipeline was three chromatin particle maps corresponding to sub-nucleosomal, 112-137 bp, core nucleosomal, 138-161 bp and super-nucleosomal, size 162-188 bp fragments (Table 3.1).

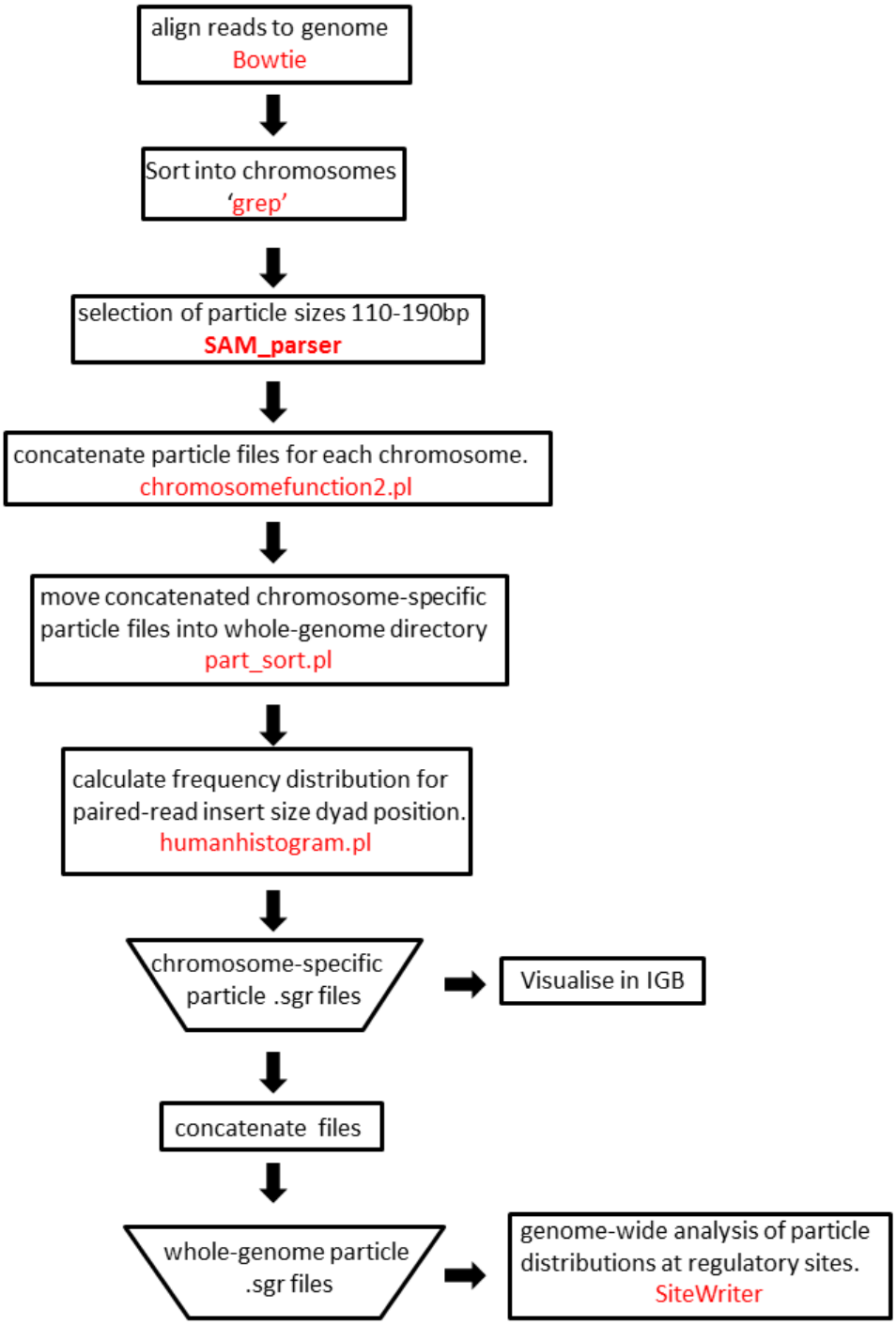


Figure 3.3 Flow chart of the data processing pipeline for paired-end sequence read data

Perl scripts or shell functions used to process the data at each step are in red text. A description of the data processing activity is shown in black text. The output from the pipeline is files containing frequency distributions of paired-read mid points for various chromatin particle size classes in Affymetrix .sgr format.

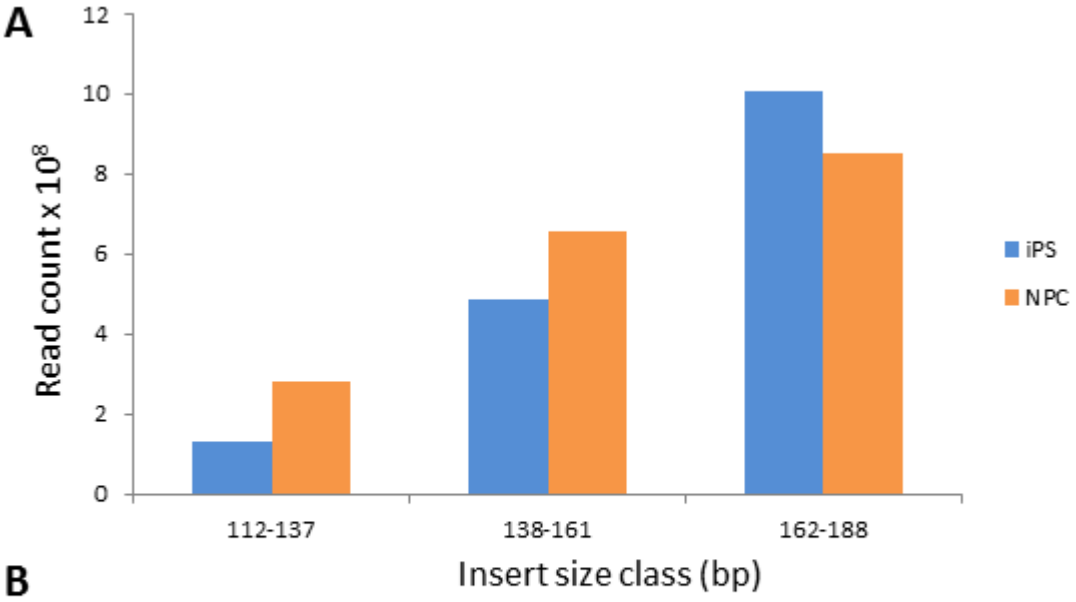
Chromatin particle size class	Name
112-137 bp	Sub-nucleosomal
138-161 bp	Nucleosomal
162-188 bp	Super-nucleosomal

Table 3.1. Division of the chromatin particle data into three size classes.

Chromatin particle data was divided computationally into three non-overlapping size classes 1) 112-137 bp: sub-nucleosomal particles 2) 138-161 bp: core nucleosomal particles and 3) 162-188 bp: super-nucleosomal particles.

The number of reads obtained in each of the three chromatin particle size classes in each cell type was calculated. This was possible as the total number of paired-end reads obtained from both cell types with respect to insert size had been determined at single base pair resolution. The frequency of chromatin particles in each of the three size classes is similar in each cell type. The ratio of NPC/iPS particles in each size class is shown in Fig 3.5

A



B

	112-137 bp	138-161 bp	162-188 bp
Ratio NPC/iPS	2.2	1.4	0.8

Figure 3.4 Analysis of paired read data obtained from iPS and NPC cells.

Chromatin particle data was divided into three size classes; i) 122-137 bp (sub- nucleosomal) ii) 138-161 bp (nucleosomal) iii) 162-188 bp (super-nucleosomal. A. Histogram of the number of aligned paired-end reads in each chromatin particle size class determined at single base pair resolution for the autosomal genome in each cell type. Read counts for the autosomal NPC genome were normalised (reads x 1.117.) B. The ratio of reads (NPC/iPS) obtained in each cell type and each insert size class

.

3.4 Characterisation of the MNase-resistant chromatin particle maps

The chromatin particle positioning maps from both iPS and NPC cells were visualised using the Integrated Genome Browser (IGB) (Nicol, Helt et al. 2009). IGB is a visualization tool for the interactive exploration of large, integrated genomic datasets. Using this tool it is possible to view chromatin particle maps from each of the human chromosomes, plotted as histograms and aligned to the genome. Hence, is possible to scan a human chromosome for regions of interest in the genome to gain an overall impression of the patterns in chromatin particle positioning in the human genome.

Fig 3.5 shows histograms of the frequency of paired-read mid points for each of the three chromatin particle maps from a section of chromosome 12 for both iPS and NPC cells. The frequency distribution of paired-read mid points in each size class resolve into discrete peaks, suggesting that individual positioned MNase-protected chromatin species have been mapped successfully.

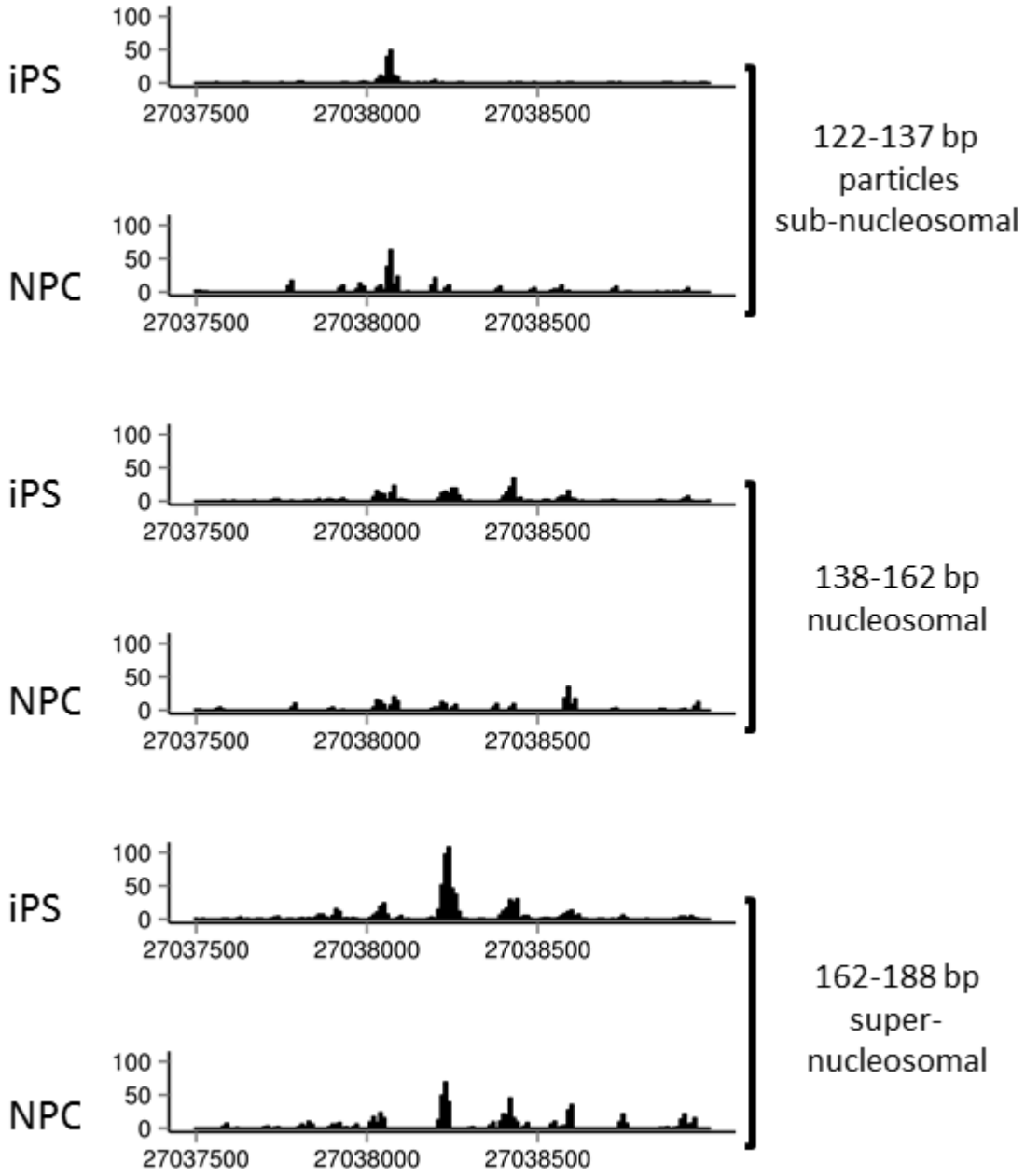


Figure 3.5 Histograms of paired-read midpoint positions derived from MNase resistant DNA in iPS cell and NPC.

Histograms were constructed by counting sequencing reads in 10 bp bins for all three chromatin particle size classes in both iPS and NPC cells. The histograms were visualised using the Integrated Genome Browser (IGB) at an intergenic region of chromosome 12 and plotted in R. Distinct patterns of peaks are detected in all three chromatin particle size classes, suggesting that the mapping of MNase-resistant chromatin particles of varying sizes in the human genome was successful.

Initial viewing of sections of the human chromatin particle maps using the IGB suggested that the patterns of chromatin particle positioning in the human genome were different from that of the *S. cerevisiae* genome. 85% of the *S. cerevisiae* genome possesses positioned nucleosomes (Mavrich, Ioshikhes et al. 2008; Kent, Adams et al. 2011) and gene bodies tend to possess nucleosome arrays (Lee, Tillo et al. 2007) whereas gene promoters and the 3' ends of genes tend to be nucleosome depleted. Examination of a section of human chromosome 12 revealed that nucleosomes do not tend to be strongly positioned at high density across the human genome, nor strongly positioned across whole gene bodies. Fig 3.6 shows an example of this using a region of chromosome 12 from the iPS chromatin particle map in the super-nucleosomal size class. An array of positioned nucleosomes is visible that is within an intron of the SOX5 gene but it is also just upstream of a splice variant of SOX5. This illustrates that it is possible to detect arrays of strongly positioned nucleosomes in the human genome, but they do not always occur across entire gene bodies. In addition, large parts of the human genome appeared to possess weakly positioned or 'fuzzy' nucleosomes. These observations suggested that much of the human genome does not possess strongly positioned nucleosomes and that strongly positioned nucleosomes might occur in low numbers at regulatory regions in the human genome, rather than across whole gene bodies.

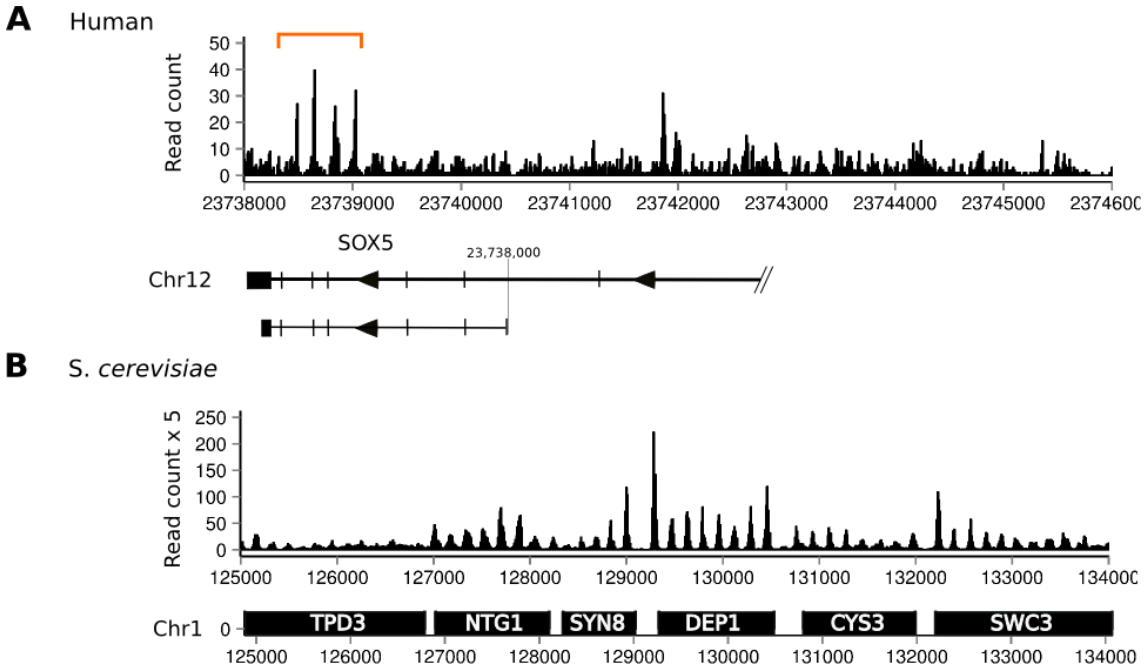


Figure 3.6 Comparison of the histograms of paired-read midpoint positions derived from MNase resistant DNA in the human and *S. cerevisiae* genomes

A. Detection of an array of peaks in the paired-read midpoint positions derived from MNase resistant DNA derived from the super- nucleosomal (162-188 bp) chromatin particle size class in iPS cells in a section of human chromosome 12. The peaks represent an array of positioned super-nucleosomal chromatin particles and the array is located just upstream of a splice variant derived from the SOX5 gene. Few well-positioned chromatin particles are detected in the rest of the gene body. A map representing the structure of the model of the SOX5 gene and the relevant splice variant is shown. **B**. A section of the histograms of paired-read midpoint positions derived from the *S. cerevisiae* chromosome 1. The peaks represent arrays of well positioned nucleosomes across all of the gene bodies. Nucleosomes are not positioned in the intervening regions, known as nucleosome free regions.

3.5 PeakFinder

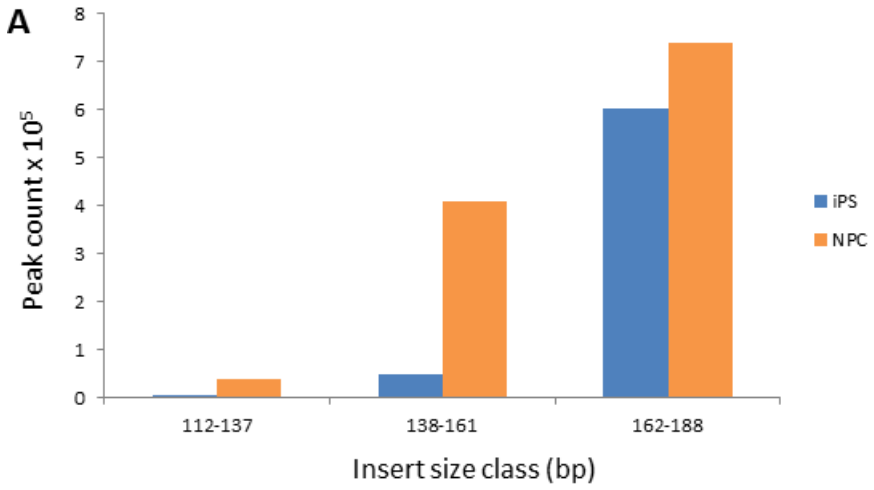
PeakFinder is a peak marking algorithm that I derived in order to investigate the number and distribution of well-positioned nucleosomal chromatin particles in iPS and NPC cells. This tool was designed to locate and report the number of genomic positions of peaks in the chromatin particle mid-point frequency distribution that might represent positioned chromatin particles. Chromatin particles that are positioned at a particular genomic locus are usually represented by a Gaussian distribution of the sequencing reads at that locus. This distribution may occur if MNase trims the DNA protected by the chromatin particle imprecisely, or it could occur because of variation in the exact nucleosome positioning across a population of cells. In a Gaussian distribution, the standard deviation determines the width of the 'bell'. In the case of chromatin particle positioning data, the shape of the distribution of aligned sequencing reads at a given locus in the genome reflects the variation in positioning; a narrow peak represents well positioned chromatin particles (See Chapter 1 Fig 1.7). Therefore it was decided to derive the algorithm to detect the locations of Gaussian distributions with a narrow bell in the frequency distributions of the mid-point position of the paired-read insert size values.

Hence, PeakFinder was a simple heuristic peak marker that detected Gaussian peaks derived from the cumulative read count in each of three consecutive 10 bp bins. The lower, 'noise' threshold of the cumulative read count in the first and third bins was derived from the average number of reads in the genome per 10 bp bin in the 138-161 bp chromatin particle size class. The upper threshold set for the centre of the three consecutive 10 bp bins was set to eliminate any peaks with artificially high values caused by sequencing artefacts. The lower threshold set for the centre of the three consecutive 10 bp bins was empirically determined. The final parameters used to define this peak were set stringently to detect high numbers of reads in the central bin of the peak, so as to detect only very strongly positioned chromatin particles. The PeakFinder tool was tested empirically by visualising the results in the IGB.

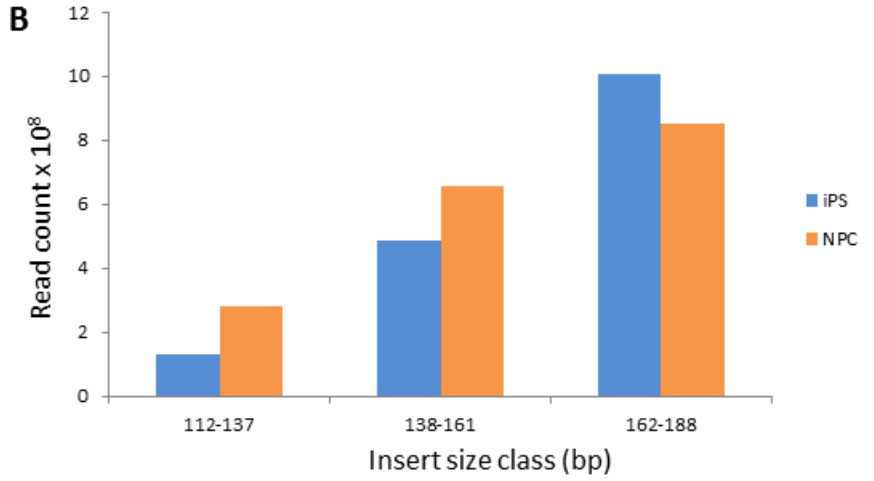
3.6 The number of positioned nucleosomes increases during neural cell development.

The PeakFinder tool was used to investigate the number of strongly positioned nucleosomes in each chromatin particle size class in both iPS and NPC cells. This would yield information about changes in chromatin particle positioning during neural cell differentiation. Using the PeakFinder tool, approximately 48,840 positioned nucleosomal chromatin particles were identified in iPS cells. Assuming a nucleosome repeat length of 200 bp in iPS cells, this corresponds to 0.32% of the genome and suggests that there are few well positioned chromatin particles in the nucleosomal chromatin particle size class in iPS cells. Following

differentiation to NPC, the number of peaks representing positioned nucleosomal chromatin particles increased to 408,152; an 8.4 increase in peaks in NPC compared with iPS cells. For the sub -nucleosomal size class, 6,026 and 39,315 peaks were found in iPS and NPC cells respectively; a 6.5 fold increase in the number of positioned particles during differentiation from iPS to NPC cells. In addition, there was only a small increase in the number of positioned super-nucleosomal chromatin particles during differentiation from iPS to NPC cells (Fig 3.7A). This suggests that chromatin is more organised in NPC cells. The major change in numbers of peaks observed during differentiation from iPS to NPC cells occurred in the nucleosomal (138-161 bp chromatin particle size) class. This suggests that the chromatin particle size representing core nucleosomes may be more dynamic in the pluripotent state and then they become more strongly positioned during differentiation from pluripotent to neural progenitor cells. Fig 3.7 B shows the same graph shown previously in Fig 3.4. When compared with Fig 3.7A, this shows that although the number of reads obtained in each chromatin particle size class is similar, indicating that the total number of chromatin particles in each genome is similar, the number of strongly positioned nucleosomes changes during differentiation from iPS to NPC.



	112-137 bp	138-161 bp	162-188bp
iPS	6,026	48,840	603,825
NPC	39,315	408,152	740,197
Ratio NPC/iPS	6.52	8.36	1.23



	112-137 bp	138-161 bp	162-188 bp
Ratio NPC/iPS	2.2	1.4	0.8

Figure 3.7 See next page for legend.

Figure 3.7 Comparison of the total number of positioned chromatin particles in each size class in iPS and NPC cells.

The number and position of peaks representing positioned chromatin particles in the iPS and NPC genomes was determined for each chromatin particle size class; i) 122-137 bp (sub- nucleosomal) ii) 138-161 bp (nucleosomal) iii) 162-188 bp (super-nucleosomal) using the heuristic peak calling method (PeakFinder). A. The frequency distribution of positioned chromatin particles in each insert size class. B. The total numbers of positioned chromatin particles in each insert size class for each cell type and the fold change in the numbers of positioned chromatin particles between iPS and NPC. C. The frequency of chromatin particles in each of the three size classes as shown in previously in Fig 3.4, when compared with the number of positioned particles in A. this illustrates that there is a change in chromatin particle positioning during differentiation to NPC, but little change in the total number of chromatin particles in each genome.

To validate the chromatin particle maps constructed in this study, data from two published high-resolution chromatin maps constructed using MNase-seq from two other human cell types, K562 and GM12878 by Kundaje *et al* (Kundaje et al. 2012) were processed so that a comparison between the data sets could be made. K562 is a chronic myelogenous leukaemia (Cml) cell line (Lozzio and Lozzio 1979) and GM12878 is a B-lymphoblastoid cell line GM12878 (Corriell Biorepository :https://catalog.coriell.org/1/Browse/Biorepositories). These maps were constructed by sequencing the DNA fragments that were purified from agarose gels in the mono-nucleosome (150 bp) size range after MNase digestion according to the method by Valouev *et al* (Valouev, Johnson et al. 2011) GEO: GSE35586. In the study presented in this thesis, all of the DNA fragments that were less than 300 bp in length were purified for DNA sequencing and then fragments were divided into three size classes for further analysis. Hence, the fragments, in the work presented here, of 138-161 bp, referred to as nucleosomal chromatin particles were used for comparison with the published K562 and GM 12878 maps. The 138-161 bp chromatin particles are the most similar in size to those in the K562 and GM 12878 mono-nucleosome maps. The data from the K562 and GM 12878 nucleosome maps was converted into the same format as the iPS and NPC data sets derived in this thesis. The data was converted from bedgraph to sgr format, re-binned into 10 bp bins and smoothed using a three bin moving average. Hence, the two data sets could be compared directly.

The number of positioned chromatin particles in each of the K562 and GM12878 re-processed nucleosome maps from K562 and GM12878 cells was determined using the PeakFinder tool and the results are shown in Table 3.2 along with the previously obtained results for iPS and NPC cells. The number of positioned nucleosomal size chromatin particles data obtained from NPC cells is most similar to that found in the K562 and GM12878 nucleosome maps. This result suggests that the number of positioned nucleosomal chromatin particles is similar in the differentiated cell lines, NPC, K562 and GM12878. Interestingly, an analysis of the overlap in the locations (within 10bp) of the positioned nucleosomal chromatin particles between these three cell types (Fig A3, Appendix), showed that only approximately 1.4% and 0.5% of the positioned nucleosomal chromatin particles in NPC cells were in the same location as those in K562 and GM12878 cells respectively and very few (682), nucleosomal chromatin particles were located in the same position in all three cell types. However, approximately 17% of the positioned nucleosomal chromatin particles in K562 cells were located in the same place as those in GM12878 cells.

Cell type	No of peaks in 150bp chromatin particle size class
hiPS	48,840
NPC	408,152
K562	363,784
GM12878	241,064

Table 3.2 Putative positioned chromatin particles in different cell types.

The frequency of positioned chromatin particles as determined by a heuristic peak calling method for the entire genomes in iPS, NPC, K562 and GM12878 cells.

Assuming a nucleosome repeat length of 200 bp in the NPC, K562, GM12878 cell types and that the size distribution of the DNA fragments isolated and sequenced in the mono nucleosome size range by Kundaje *et al* is the same as those in the nucleosomal (138-161 bp) chromatin particle size class in this study, it is possible to estimate and compare the fraction of the genome that possesses positioned nucleosomal chromatin particles in each of these cell types. 2.7% of the NPC genome contains chromatin particles in the nucleosomal (138-161 bp) size class. In the K562 and GM12878 cells, it was determined here, using PeakFinder on the re-processed published data, that 2.45% and 1.6% of these genomes contain positioned nucleosomes respectively. This analysis suggests that the quality of the chromatin particle maps derived in this thesis is similar to that of the published maps generated by Kundaje *et al*.

Gaffney *et al* and Valouev *et al* generated nucleosome maps from differentiated human cell lines using MNase seq and estimated the fraction of the genome that possesses positioned nucleosomes. Valouev *et al* (Valouev, Johnson et al. 2011) used a positioning stringency metric to estimate the percentage of the genome that is occupied by positioned nucleosomes in granulocytes, CD4+, and CD8+ T-cells. They determined that, at the lowest stringency they utilised, a maximum of 20% of the genome possesses positioned nucleosomes. Similarly, Gaffney *et al* (Gaffney, McVicker et al. 2012) analysed a sample of a million regions of 200 bp in length from their nucleosome maps generated by *in vivo* MNase digestion in lymphoblastoid cells. They estimated nucleosome positioning by counting the number of midpoints of DNA fragments of 142-152 bp in size within 15 bp of a genomic position and excluded positions with less than 50 midpoints. In this way they estimated that 81% of the human genome has weak nucleosome positioning, 8.4 % has moderately strong positioning and only 0.3% has very strong positioning.

Although there are minor differences between the results obtained in this study and in the published studies, these may be accounted for by cell type specific differences in positioning, in the methods of DNA fragment isolation or analysis and variation in the algorithms used to determine nucleosome positions. Taken together, the results presented here are in general agreement with the conclusion reached by Gaffney *et al* and Valouev *et al*, that most of the human genome does not possess strongly positioned nucleosomes.

3.7 Investigating the organization of positioned chromatin particles.

Having determined the number of positioned nucleosomal size chromatin particles in the iPS NPC, K562 and GM12878 genomes, the next question was whether the chromatin particles might be positioned in functional groups such as phased arrays. To answer this question, the average chromatin structure at each of the peak positions obtained using the peak finding

algorithm (PeakFinder) in the genomes of all four cell types was determined in the nucleosomal chromatin particle size class. The positions of the peaks were used to construct an average cumulative frequency distribution of the sequence read mid-points at and surrounding (+/- 300 bp) each nucleosomal (138-161 bp) chromatin particle position for each of the cell types. Fig 3.8 shows a comparison of the average cumulative frequency distribution of the nucleosomal chromatin particle size class surrounding a peak position in the nucleosomal chromatin size class in all four cell types, iPS, NPC, K562 and GM12878 cells. This result shows that the pattern of nucleosome positioning in all four cell types is broadly similar. Nucleosomal chromatin particles in these cell types do not form phased arrays; rather, they are on average flanked by one or two nucleosomes. Hence, the results presented here suggest that the human genome possesses few strongly positioned nucleosomes and that they are located in discrete regions of the genome.

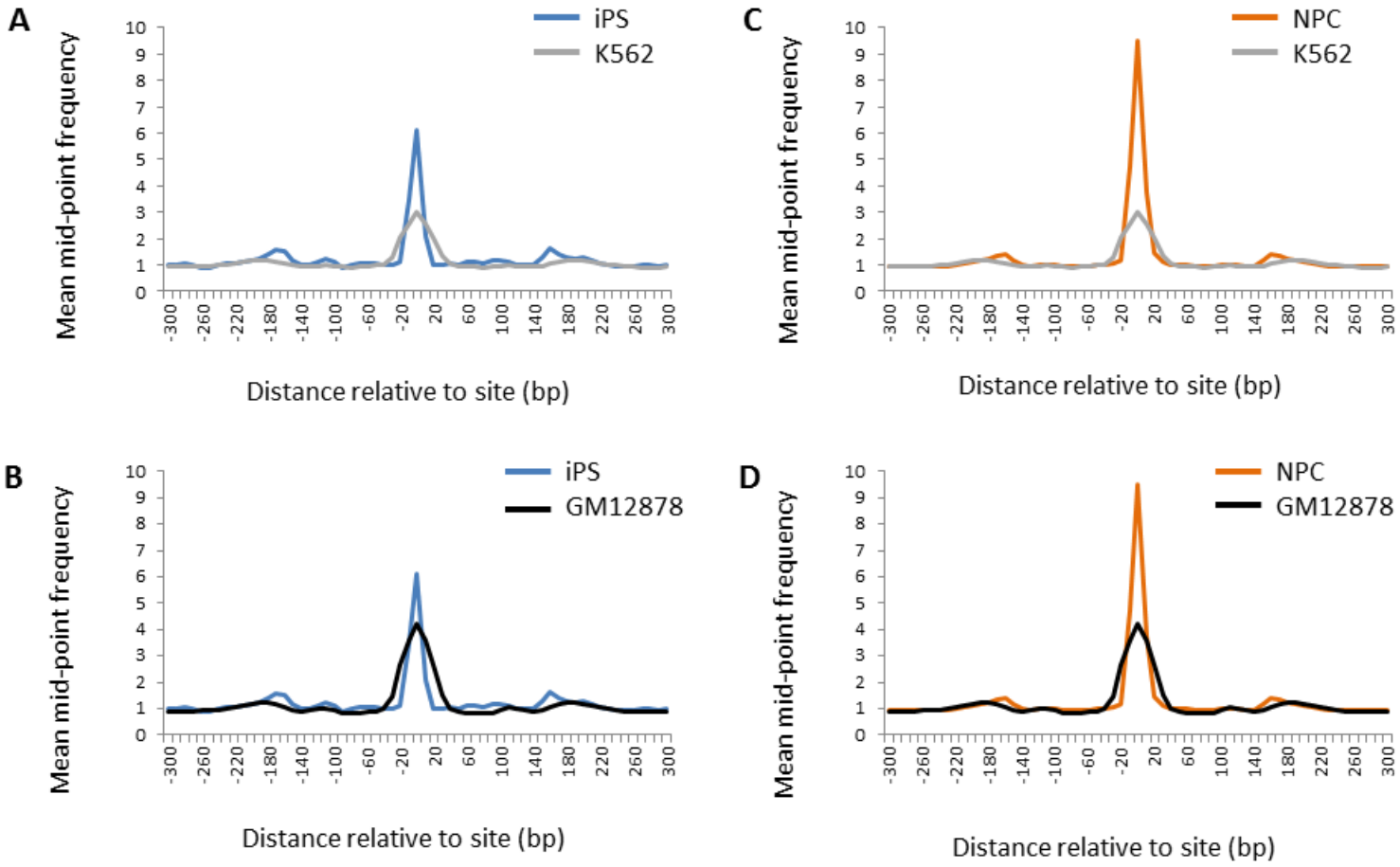


Fig 3.8 See next page for legend

Figure 3.8 Comparisons of the positions of chromatin particles in the nucleosomal (138-161 bp) chromatin particle size range mapped in iPS and NPC cells by MNase-seq with those mapped in K562 and GM12878 cells.

The data from the nucleosome maps generated for the K562 and GM12878 cell lines (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012) was converted from bedgraph to sgr format and the data was re-binned into 10 bp bins, calculating a 3bin moving average. The K562 and GM12878 maps were constructed by sequencing DNA purified from chromatin in the mono-nucleosome (approx. 150 bp) size range after MNase digestion. These data were compared with the iPS and NPC chromatin particle maps generated in the 138-161 bp size range since this is most similar to the size range of the chromatin particles in the K562 and GM12878 maps. The positions of all of the peaks in the K562, GM12878, iPS and NPC genomes were determined using the peak detecting algorithm (PeakFinder). The positions of the peaks from each genome were used to construct an average cumulative frequency distribution of the chromatin particle distribution at and surrounding (+/- 300 bp) each 150 bp chromatin particle position in the respective genomes of each cell type. Comparison of the average cumulative frequency distribution of the peak positions are shown in A. iPS and K562, B. iPS and GM12878, C. NPC and K562, D. NPC and GM12878.

3.8 Chromatin particles from different size classes in the same cell type are largely discrete.

The chromatin particle size data was divided into three non-overlapping size classes; sub-nucleosomal particles, nucleosomal particles and super-nucleosomal particles. This was in order to detect genomic loci that might be protected by chromatin particles with differing structure and /or function, for example transcription factors, transcription factor complexes, modified nucleosomes or chromatosomes. However, it was possible that the chromatin particle size classes created here might have been artificial and rather than represent discrete particles, that they were derived from a common binding protein. This could occur through variation in the extent of MNase digestion, generating extended footprints and/or variation in positioning across the cell population. Alternatively, these chromatin particles might interconvert. As nucleosomes can change their structure *in situ* through the action of chromatin remodelling complexes, it was decided to investigate whether there might be any relationship between the positions of chromatin particles in the same cell type but in different size classes. Nucleosomes can be post-translationally modified, for example by the exchange histone variants. In addition, how tightly the DNA wraps around the nucleosome core can vary and in the human genome, the presence or absence of histone H1 associating with the nucleosome core can vary (See Chapter 1 section 1.7). All of these changes have the potential to affect the amount of DNA protected by chromatin proteins when it is digested with MNase.

A comparison of the locations of positioned of chromatin particles across different size classes within each cell type was determined by comparing the location of the peaks, representing well-positioned chromatin particles, obtained using PeakFinder. This was undertaken by determining how many chromatin particles in one size class were located in the same genomic position as a chromatin particle in each of the other two size classes within a window of +/- 10 bp Fig 3.9A shows the results of this analysis in iPS cells. 20% of the chromatin particles in the sub-nucleosomal (112-137 bp) size class overlap in position with those in the nucleosomal (138-161 bp) size class in iPS cells. Similarly, in iPS cells, 24% of the chromatin particles in the nucleosomal size class (138-161 bp) overlap in position with those in the super nucleosomal (162-188 bp) size class. In NPC cells 6.6% of the sub-nucleosomal chromatin particles (112-137 bp) overlap in position with those in the nucleosomal (162-188 bp) chromatin particle size class. The result is similar (5.4%) for the overlap of the nucleosomal (162-188 bp) and the super nucleosomal (162-188 bp) chromatin particle size class in NPC cells. (Fig3.9 B). These data show that the chromatin particles in the three different chromatin particle sub-classes are mostly discrete.

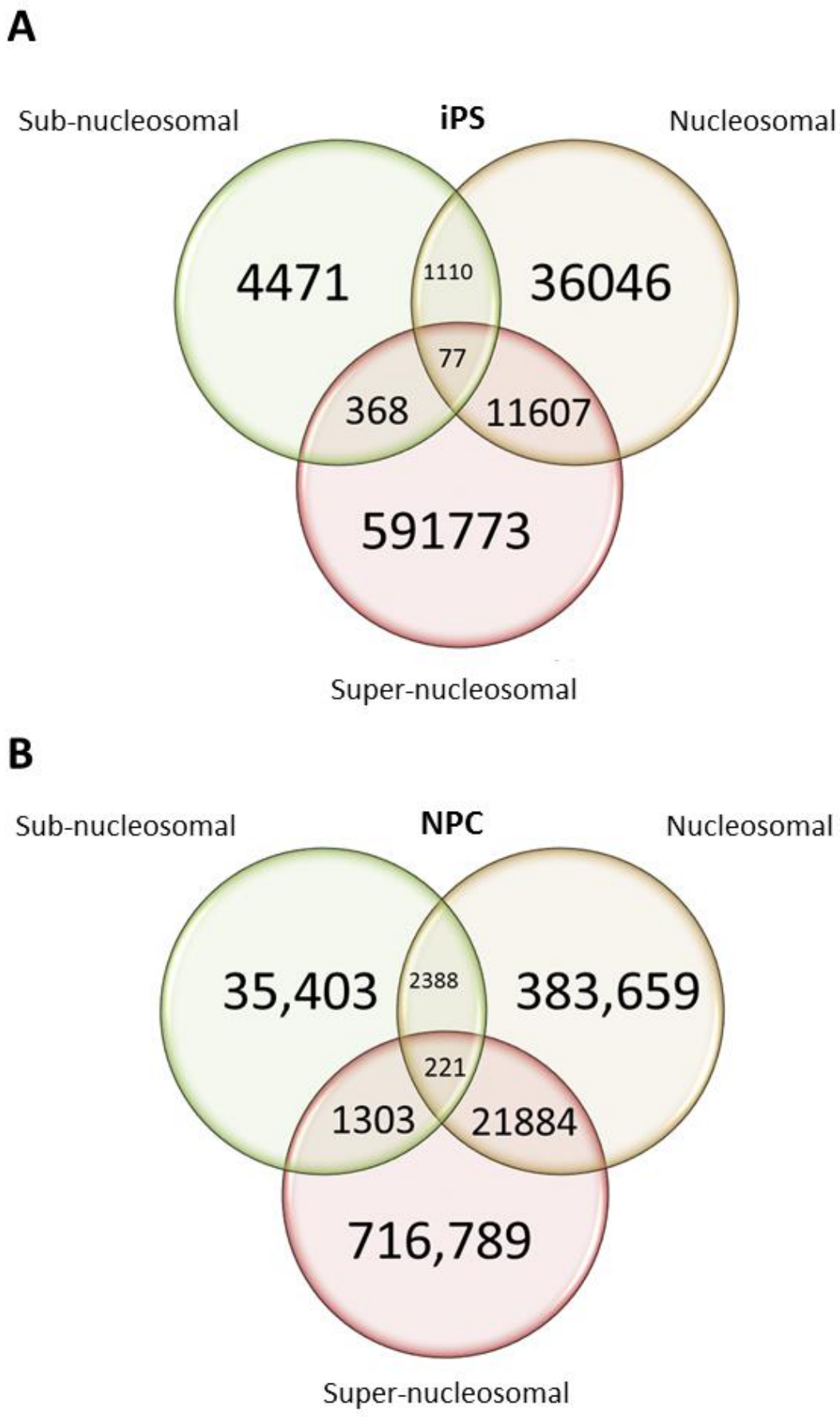


Fig 3.9 See next page for legend

Figure 3.9 Chromatin particles mapped by MNase-seq in the same cell type, exhibit largely discrete MNase-resistant DNA size classes

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were given explicit genome positions using the heuristic peak-finding algorithm (PeakFinder). The Venn diagram shows the number of particle positions determined for each size class of sequence read; i) 122-137 bp (sub- nucleosomal) ii) 138-161 bp (nucleosomal) iii) 162-188 bp (super-nucleosomal) and the frequency of the overlap of the positions within +/- 10 bp for each size class for A. iPS and B. NPC.

It was decided to test this further using a different method. Hence, the chromatin particle data for 20,000 randomly selected peak positions in iPS cells were clustered into four groups in the nucleosomal chromatin particle size class (Fig 3.10). These groups are shown as heat maps on the left side of the panel. For each group, the mean chromatin particle frequency values centred at and surrounding the nucleosomal chromatin particle size class positions were plotted for both the nucleosomal (138-161 bp) and the super-nucleosomal (162-188 bp) size class data. The results are shown in the average cumulative frequency distributions on the right-hand side of each panel. The peaks in mean chromatin particle mid-point frequency in the nucleosomal (138-161 bp) size class data are accompanied by peaks that occur at a much lower frequency in the super-nucleosomal (162-188 bp) size class data. This result suggests that MNase protected particles in the nucleosomal size class are not frequently observed as super-nucleosomal-size MNase protected particles in the same genomic positions. This confirms the observation shown in the Venn diagrams in Fig 3.9.

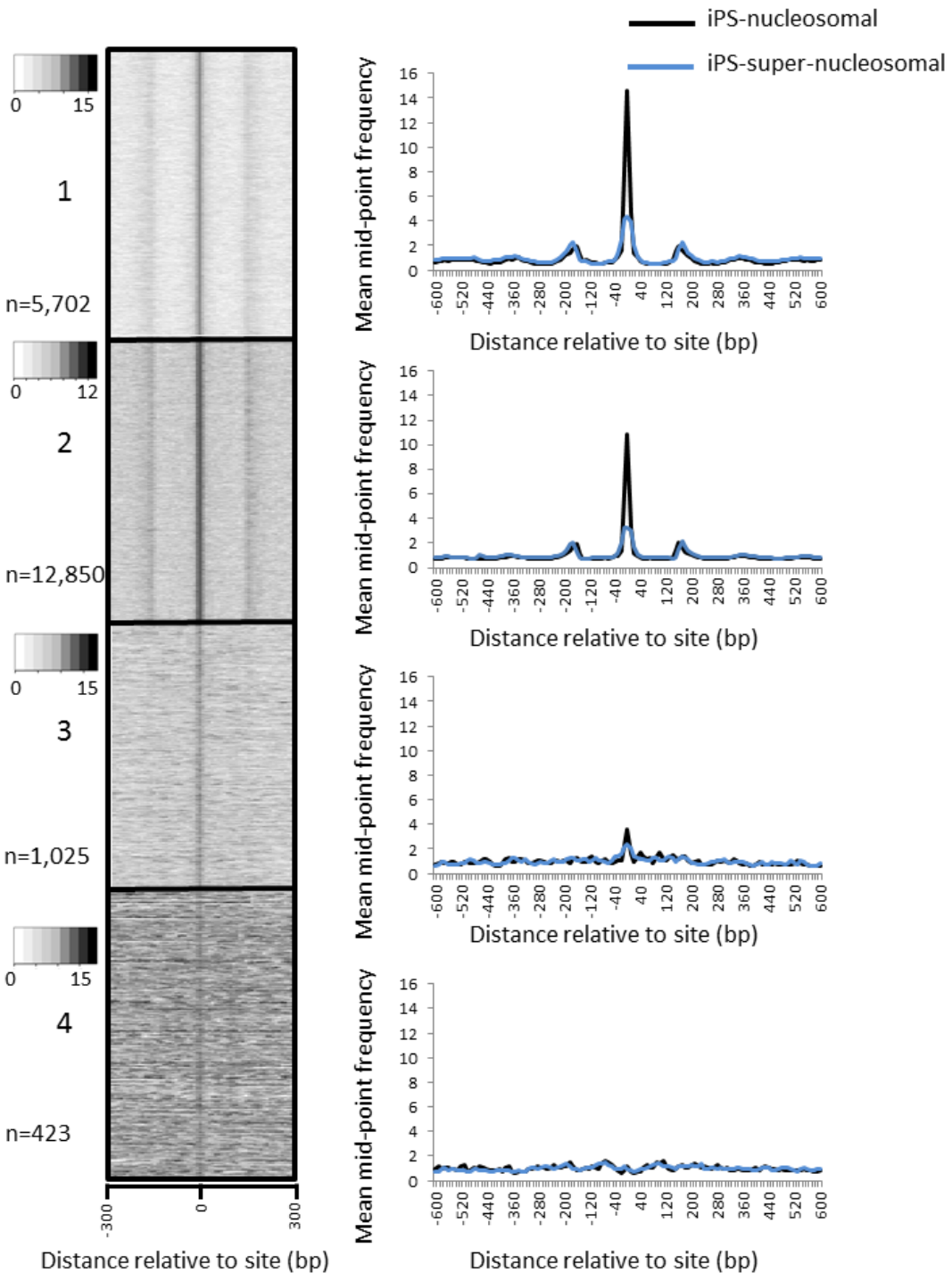


Fig 3.10 see next page for legend

Figure 3.10 Chromatin particles mapped in iPS cells by MNase-seq exhibit largely discrete MNase-resistant DNA size classes.

20,000 randomly selected peak positions in the nucleosomal (138-161 bp) chromatin particle size class in iPS cells described above were clustered into four groups as shown in the left-hand panel. For each group, the mean chromatin particle frequency values centred on and surrounding the nucleosomal (138-161 bp) particle positions were plotted for both nucleosomal (138-161 bp) and the super-nucleosomal (162-188 bp) size class data. The graphs are shown on the right-hand side. The peaks in mean chromatin particle frequency in the nucleosomal (138-161 bp) size class data are accompanied by peaks that occur at a much lower frequency in the super-nucleosomal (162-188 bp) size class data, again suggesting that nucleosomal (138-161 bp) MNase protected particles in chromatin are not frequently observed as super-nucleosomal (162-188 bp) MNase protected particles in the same genomic positions.

A similar analysis was carried out for 20,000 randomly selected peak positions in the nucleosomal chromatin particle size class in NPC cells (Fig 3.11). These results show that core nucleosomal size MNase protected particles in chromatin are not observed frequently as super-nucleosomal-size MNase protected particles in the same genomic positions in NPC cells.

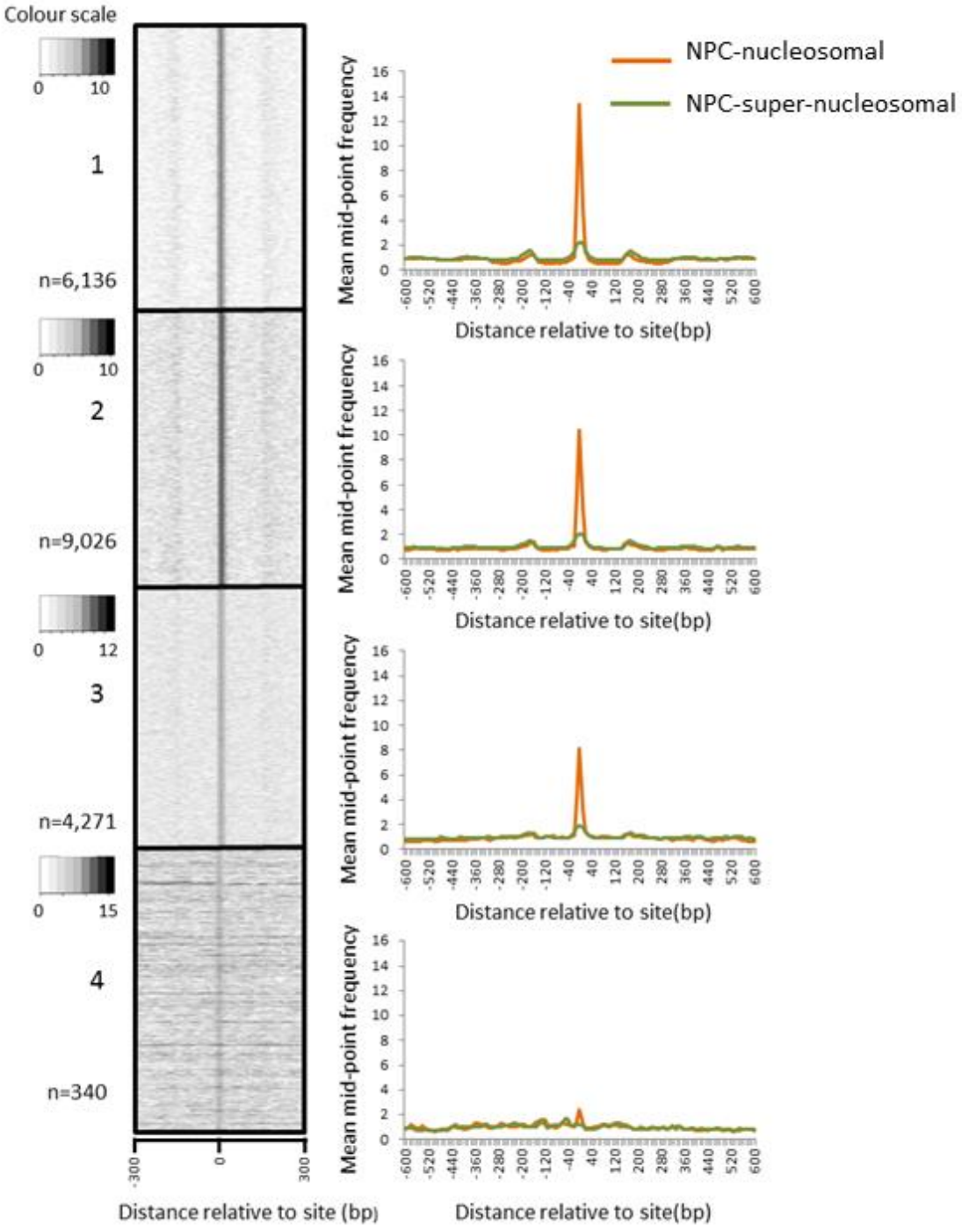


Fig 3.11. See next page for legend

Figure 3.11 Chromatin particles mapped in NPC cells by MNase-seq exhibit largely discrete MNase-resistant DNA size classes.

20,000 randomly selected peak positions in the nucleosomal (138-161 bp) chromatin particle size class in NPC cells described above were clustered into four groups as shown in the left-hand panel. For each group, the mean chromatin particle frequency values centred on and surrounding the nucleosomal (138-161 bp) particle positions were plotted for both nucleosomal (138-161 bp) and the super-nucleosomal (162-188 bp) size class data. The graphs are shown on the right-hand side. The peaks in mean chromatin particle frequency in the nucleosomal (138-161 bp) size class data are accompanied by peaks that occur at a much lower frequency in the super-nucleosomal (162-188 bp) size class data, again suggesting that nucleosomal (138-161 bp) MNase protected particles in chromatin are not frequently observed as super-nucleosomal (162-188 bp) MNase protected particles in the same genomic positions.

3.9 Defining the relationship between chromatin particles during neural cell differentiation.

Next, the relationship between chromatin particles in the same size class was investigated during differentiation from iPS to NPC as an indication of developmental change. The location of the peaks, representing well-positioned chromatin particles, obtained using PeakFinder was quantified in each of the nucleosomal and super-nucleosomal size classes in iPS and NPC. The degree of overlap in the positions of particles during neural development is shown in the Venn diagrams Fig 3.12 A (nucleosomal particles, 138-161 bp) and Fig 3.13 A (super-nucleosomal particles, 162-188 bp). This shows how many positioned chromatin particles of the same size class were in the same position, within a window of +/- 10 bp after differentiation of iPS to NPC cells. These data show that 32% of the iPS nucleosomal chromatin particles overlap in position with the NPC nucleosomal chromatin particles, whereas 17% of the iPS super-nucleosomal chromatin particles overlap with the NPC super-nucleosomal chromatin particles.

The four cluster groups generated in iPS cells in both the nucleosomal chromatin particle size class (138-161 bp) and the super-nucleosomal chromatin particle size class (162-188 bp) were used to generate the average cumulative frequency distributions shown in Fig 3.12B and Fig 3.13B respectively, for both iPS and NPC cells. These results show that where a nucleosomal chromatin particle is positioned in iPS cells, this tends to remain positioned in NPC cells. Similarly, where a positioned super- nucleosomal chromatin particle is positioned in iPS cells, this tends to remain positioned in NPC cells. Hence, chromatin particles of the same size class tend to occur in the same locations during neural cell differentiation, confirming the results in the Venn diagrams in each case.

These analyses, taken together with the fact that the number of positioned chromatin particles increases during differentiation, suggest that during differentiation, most of the additional nucleosomes are positioned *de novo*, i.e. generally they do not occur due to the inter-conversion of existing particles of another size class. These results suggest that the different chromatin particles are discrete entities that are positioned independently from each other during differentiation.

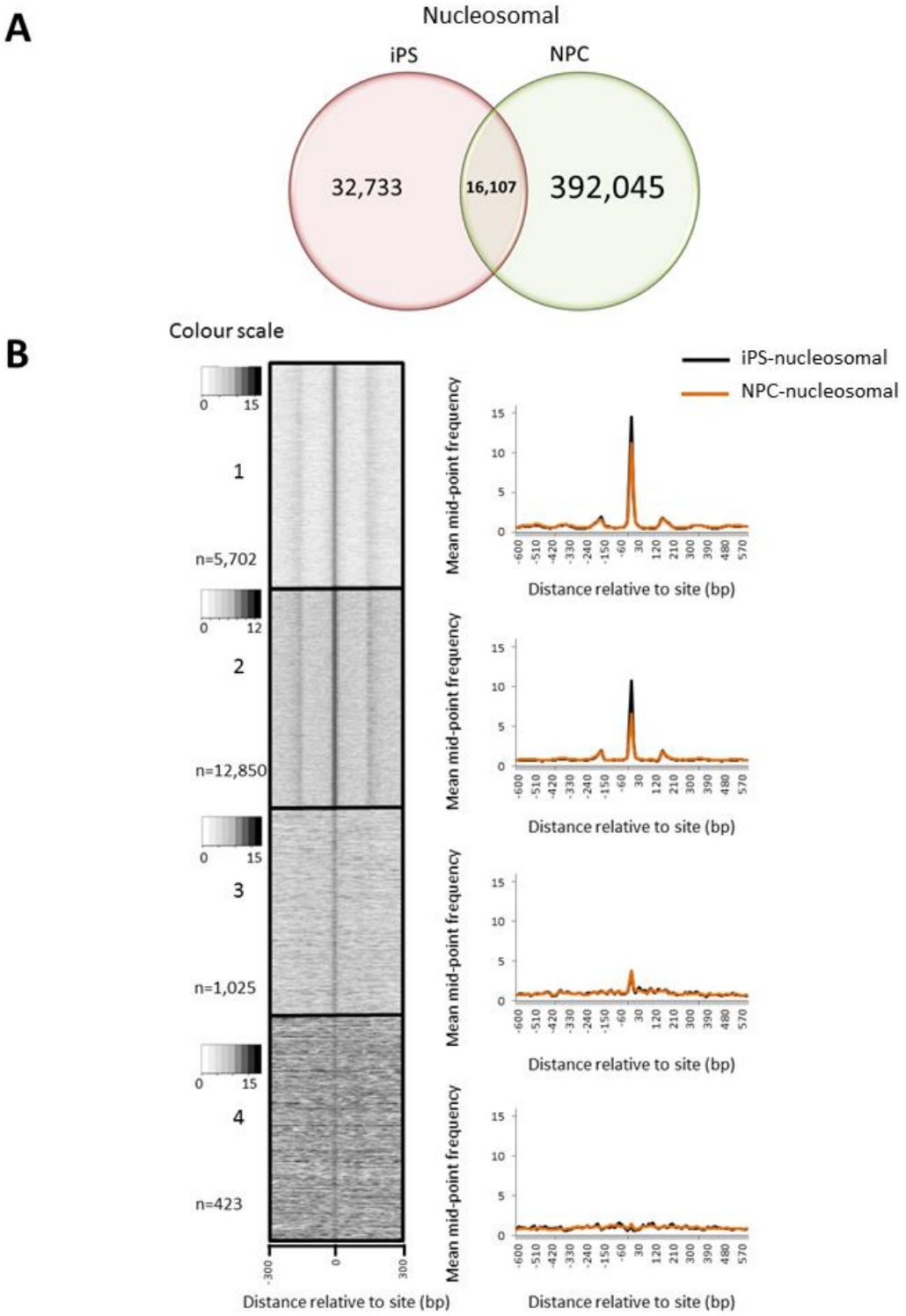


Figure 3.12 Distribution of nucleosomal chromatin particle positions during neural cell development.

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were given explicit genome positions using the heuristic peak-finding algorithm (PeakFinder). A. The Venn diagram shows the number of particle positions determined for the nucleosomal (138-161 bp) size class of sequence read in iPS and NPC cells and the frequency of the overlap of the positions in each cell type within +/- 10 bp for this size class. B. The mean chromatin particle frequency values centred on and surrounding the nucleosomal (138-161 bp) size class particle positions were plotted for the nucleosomal (138-161 bp) size class data in iPS and NPC cells and they are shown in the graphs.

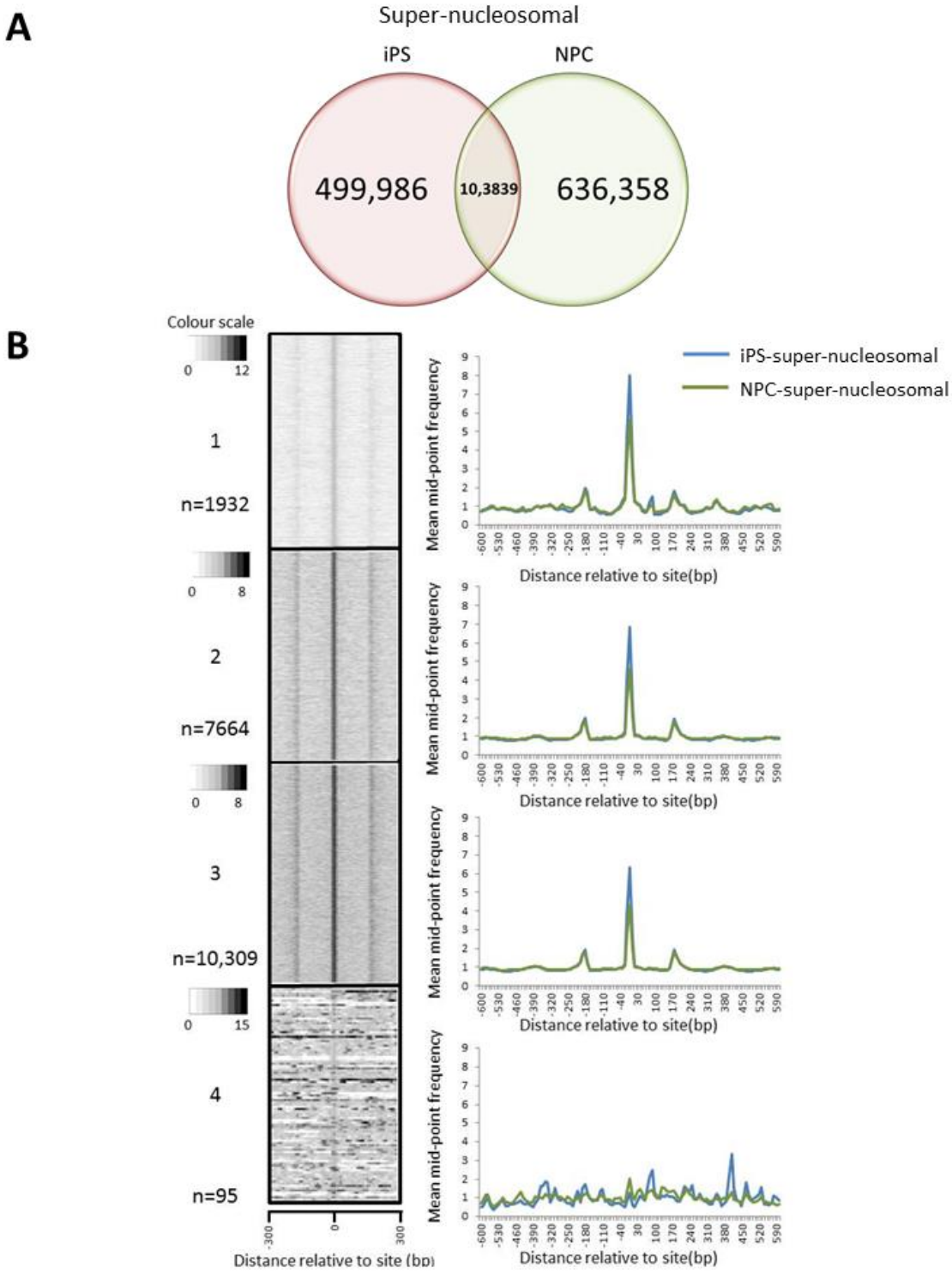


Fig 3.13 See next page for legend.

Figure 3.13 Distribution of super-nucleosomal chromatin particle positions during neural cell development.

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were given explicit genome positions using the heuristic peak-finding algorithm (PeakFinder). A. The Venn diagram shows the number of particle positions determined for the super-nucleosomal (162-188 bp) size class of sequence read in iPS and NPC cells and the frequency of the overlap of the positions in each cell type within +/- 10 bp for this size class. B. The mean chromatin particle frequency values centred on and surrounding the super-nucleosomal (162-188 bp) size class particle positions were plotted for the super-nucleosomal (162-188 bp) size class data in iPS and NPC cells and they are shown in the graphs.

3.10 Discussion.

In this chapter, high-resolution genome-wide chromatin particle maps were constructed from both iPS and NPC cells, from which individual positioned chromatin particles were detected. The chromatin particles between approximately 110 bp and 190 bp in size were divided into three size classes and it was shown that the frequency of positioned chromatin particles is greater in NPC cells than in iPS cells in all of the chromatin particle size classes. This suggests that the frequency of positioned chromatin particles in the genome increases during development from pluripotent to neural progenitor cells. The eight-fold increase in the frequency of chromatin particles that occurs in neural progenitor cells compared with pluripotent cells in the chromatin particle size class that represents core nucleosomes (138-161 bp) is the largest and most notable. Although the vast majority of the fragments in the 138-161 bp size range will be derived from sequences protected by nucleosomes, for any given loci the possibility that protected fragments could be derived from the binding of other proteins, whose footprints fall into the nucleosomal size range, cannot be excluded. However, this result suggests that it is the core nucleosomes that are the most dynamic of the chromatin particles analysed here.

Hence, in pluripotent cells few of the nucleosomes are positioned and then, during differentiation to NPC cells, chromatin is re-organised, resulting in positioning of nucleosomes (Fig 3.14). This global increase in the number of positioned nucleosomes in differentiated cells compared with pluripotent cells is in agreement with West *et al* (West, Cook et al. 2014) who have shown in both mouse and human cells that nucleosome occupancy increases during differentiation from iPS to fibroblast cells. This is also illustrated in this study (Table 3.2) by comparison of the number of positioned nucleosomes in iPS cells with the number detected in the published K562 and GM12878 chromatin maps (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012). Patterns of organised nucleosomes have been implicated in gene regulation both in yeast (Jiang and Pugh 2009) and more recently in the human genome Schones *et al* (Schones, Cui et al. 2008). The high frequency of positioned nucleosomes in NPC cells that we detect, indicative of more organised chromatin, suggests a high degree of genome regulation in these cells.

Mapping the locations of changes in patterns of nucleosome positioning to regulatory regions of the genome is important in determining the role of nucleosome positioning in the mechanisms of regulation of neural cell development. As there is evidence that nucleosomes might limit the access of polymerases to genes (Lorch, LaPointe et al. 1987) and deletion of histone proteins can affect gene expression (Han and Grunstein 1988), the work in the rest of

this thesis focusses on nucleosome positioning at transcriptional regulatory regions in the human genome. To date, much of the work to date in many different eukaryotic organisms (see Chapter 1, section 1.14) has focussed on nucleosome positioning at transcriptional start sites (TSS) and at transcription factor binding sites, therefore the work in the next chapter investigated nucleosome positioning at TSS and at the binding sites of transcription factors relevant to neural cell differentiation

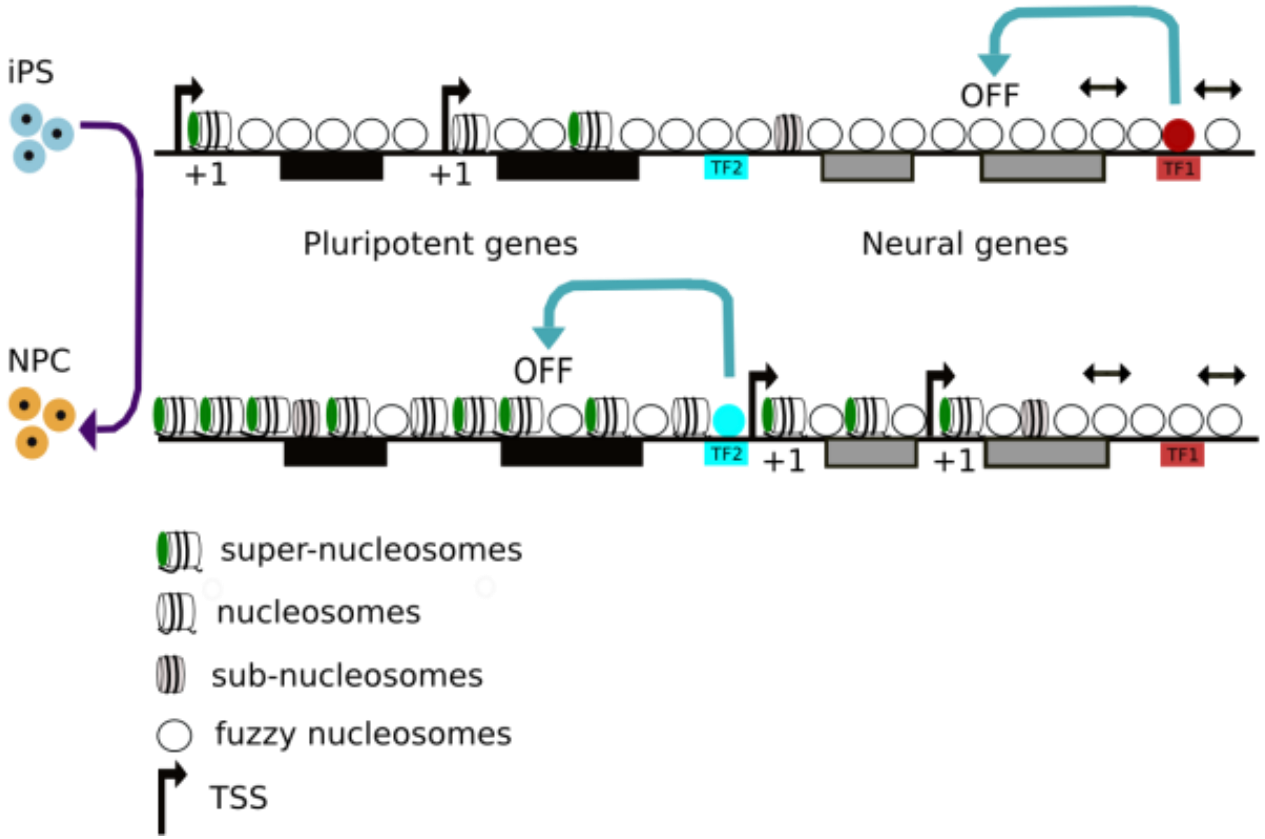


Fig 3.14 See next page for legend

Figure 3.14 Summary of nucleosome positioning during human neural cell development.

Discrete chromatin particles of different size classes can be detected in both iPS and NPC, represented as sub-nucleosomes (112-137 bp), nucleosomes (138-161 bp) and super-nucleosomes (162-188 bp). Chromatin particles that are not strongly positioned exist in all three size classes in both the iPS and NPC genomes; they are represented by 'fuzzy' nucleosomes in the diagram. During neural cell development, the number of positioned chromatin particles increases in all chromatin particle size classes, but most notably in the nucleosomal chromatin particle size class.

Chapter 4 : Nucleosome positioning at TSS and selected transcription factor binding sites during neural cell differentiation.

4.1 Introduction

Patterns of nucleosome positioning at TSS appear to be mainly conserved across the eukaryotes. Patterns of nucleosome positioning at transcriptional start sites have been well characterised in a variety of eukaryotic organisms from yeast to man. Many of these studies have examined the correlation between the patterns of nucleosome positioning at TSS and gene expression. This has led to the idea that the positioning of the +1 and +2 nucleosomes in gene promoters is important in the regulation of gene expression (Lorch, LaPointe et al. 1987; Boeger, Griesenbeck et al. 2003; Henikoff 2008). However, this is not as simple as it first appeared. Zaugg *et al* showed that patterns of nucleosome positioning at TSS in *S. cerevisiae* can vary and that there is a group of highly expressed *S. cerevisiae* genes that do not have a strongly positioned +1 nucleosome (Zaugg and Luscombe 2012). They suggested that the patterns of +1 and -1 nucleosome positioning can be grouped into four categories, reflecting different expression states, determining whether a gene is 'on' or 'off', rather than using the expression level of a gene. In *S. pombe* it has been shown that the presence of a -1 nucleosome negatively correlates with gene expression (Moyle-Heyrman, Zaichuk et al. 2013).

Studies of genome-wide patterns of nucleosome positioning in human cells have shown that the human genome appears to have a higher density of nucleosomes at regulatory regions. This may generate cell-type specific chromatin structure, specifying cellular differentiation (Schones, Cui et al. 2008; Tillo, Kaplan et al. 2010). These studies have shown that human nucleosomes tend to be positioned around the TSS of expressed genes, whereas unexpressed genes have a +1 nucleosome, but no other positioned nucleosomes near the TSS. In addition, West *et al* have shown that changes in the positions of nucleosomes in the human genome tend to be single changes, often at transcription factor binding sites and enhancers (West, Cook et al. 2014). It has been shown that the average pattern of nucleosome positioning for all TSS is not a strong one. This is most likely because the patterns of nucleosome positioning at human TSS are variable; there is not a canonical pattern. By clustering the patterns of positioned nucleosomes at TSS in human K562 cells, Kundaje *et al* have shown that variable positioning patterns exist. In addition, 20% the TSS had high gene expression levels that correlated with a strongly positioned -1 nucleosome, but not a strongly positioned +1 nucleosome (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012).

The aim of this chapter was to determine whether there are patterns of chromatin particle positioning at transcription regulatory sites in the genome and whether they change during neural cell differentiation. This would indicate which genes might be targeted for regulation during early neural development.

4.1 Sub-nucleosomal particles are positioned upstream of protein - coding gene TSS in iPS and NPC

In order to determine whether there is a canonical pattern of chromatin particle positioning at TSS in iPS cells and NPC and whether there is a general change in any pattern at TSS during neural cell development, the positions of all known human protein coding TSS (n=67,047) were extracted from the Gencode annotation file. Fig 4.1 shows the average cumulative frequency distribution of MNase protected DNA species for all three chromatin particle size classes; sub-nucleosomal, nucleosomal and super-nucleosomal at and surrounding the TSS of all known human protein coding genes. This result shows that, on average, a canonical chromatin particle positioning pattern cannot be detected in the nucleosomal and super-nucleosomal chromatin particle size classes in iPS cells or in NPC. However, a peak in the frequency distribution of MNase protected DNA species in the sub-nucleosomal chromatin particle size class approximately 60 bp upstream of the TSS in both cell types was detected.

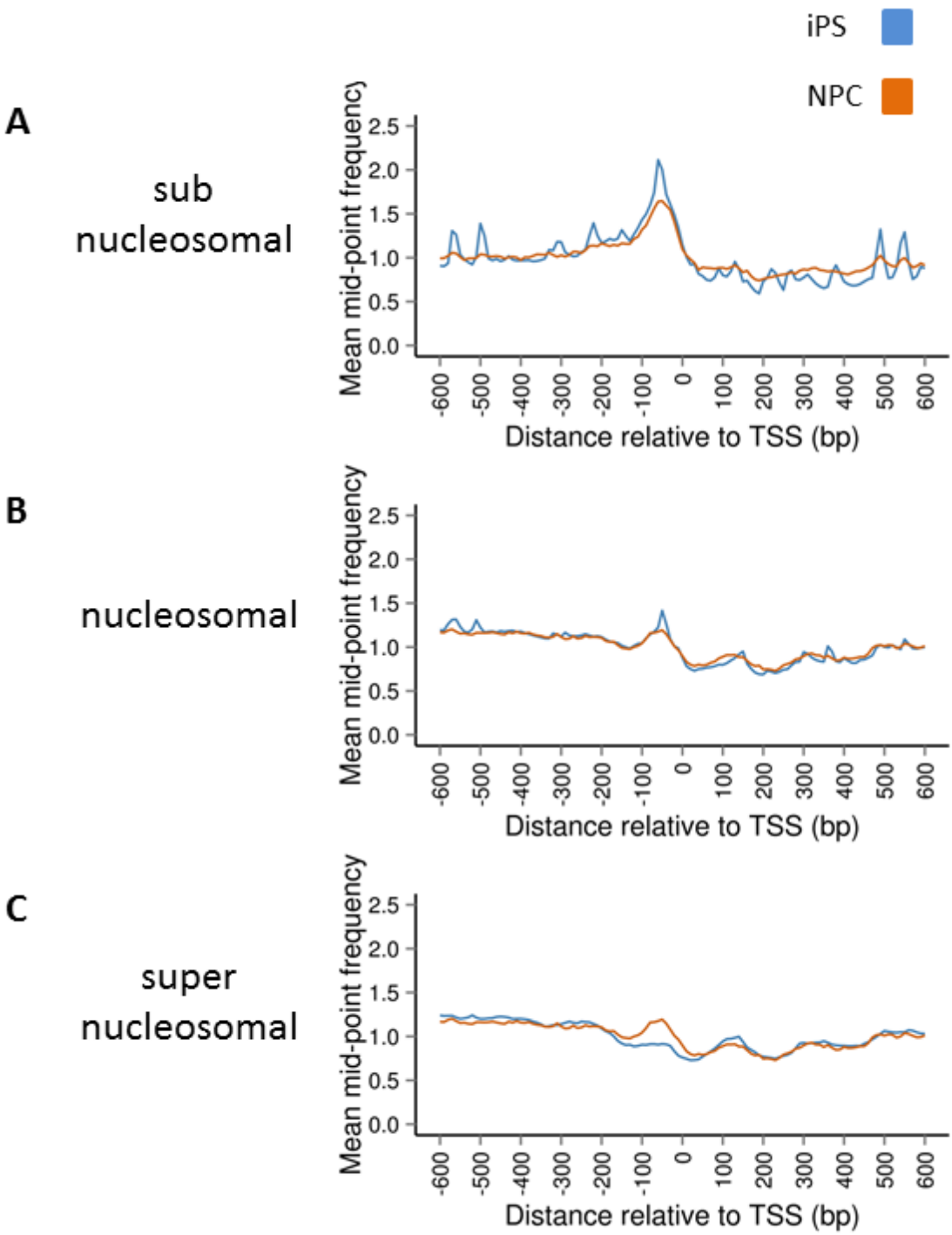


Figure 4.1 Chromatin structure at known protein coding TSS in iPS and NPC cells

Average cumulative frequency distributions of chromatin particle size classes A. 112-137 bp (sub-nucleosomal), B. 138-161 bp (nucleosomal) and C. 162-188 bp (super-nucleosomal) show the chromatin structure at and surrounding (+/-600 bp) of known protein coding TSS in both iPS and NPC cells (n=67,047).

4.2 Detection of changes in chromatin structure at transcriptional start sites during neural cell differentiation.

To detect changes in chromatin particle positioning at TSS during neural cell differentiation, the number and location of peaks in the frequency distribution of MNase protected DNA species representing positioned chromatin particles from all three chromatin particle size classes was investigated. In chapter 3 the PeakFinder tool was used to locate the positions of all of the strongly positioned chromatin particles in all three chromatin particle size classes; sub-nucleosomal, nucleosomal and super-nucleosomal. These data were used together with the positions from a non-redundant list of all full length human TSS (n=84,142) to determine the number and location of peaks present within 300 bp upstream or downstream of a transcriptional start site (TSS) in one cell type or in common between iPS and NPC cells. The results of this analysis are shown in the Venn diagrams in Fig 4.2.

These results suggest firstly, that in all three chromatin particle size classes, there are more positioned chromatin particles at TSS in NPC cells than in iPS cells. This is not particularly surprising, as the magnitude of the increase in positioned chromatin particles at TSS is broadly in line with the overall increase in number of positioned particles that occurs in the whole genome during neural cell differentiation. Secondly, these results show that a) in the sub-nucleosomal chromatin particle size class 49% of the TSS with peaks in iPS cells overlap with those in NPC, b) in the nucleosomal chromatin particle size class 55% of the TSS with peaks in the nucleosomal chromatin size class iPS overlap with those in NPC and c) in the super-nucleosomal chromatin particle size class, 38% of TSS with peaks in iPS overlap with those in NPC.

Taken together, these results suggest that nucleosomes can be positioned both upstream and downstream of a TSS in both cell types, but the number of positioned nucleosomes within 300 bp of a TSS increases during neural cell development. In addition, of the genes with chromatin structure within 300 bp of a TSS in pluripotent cells, approximately 50% are specific to these cells.

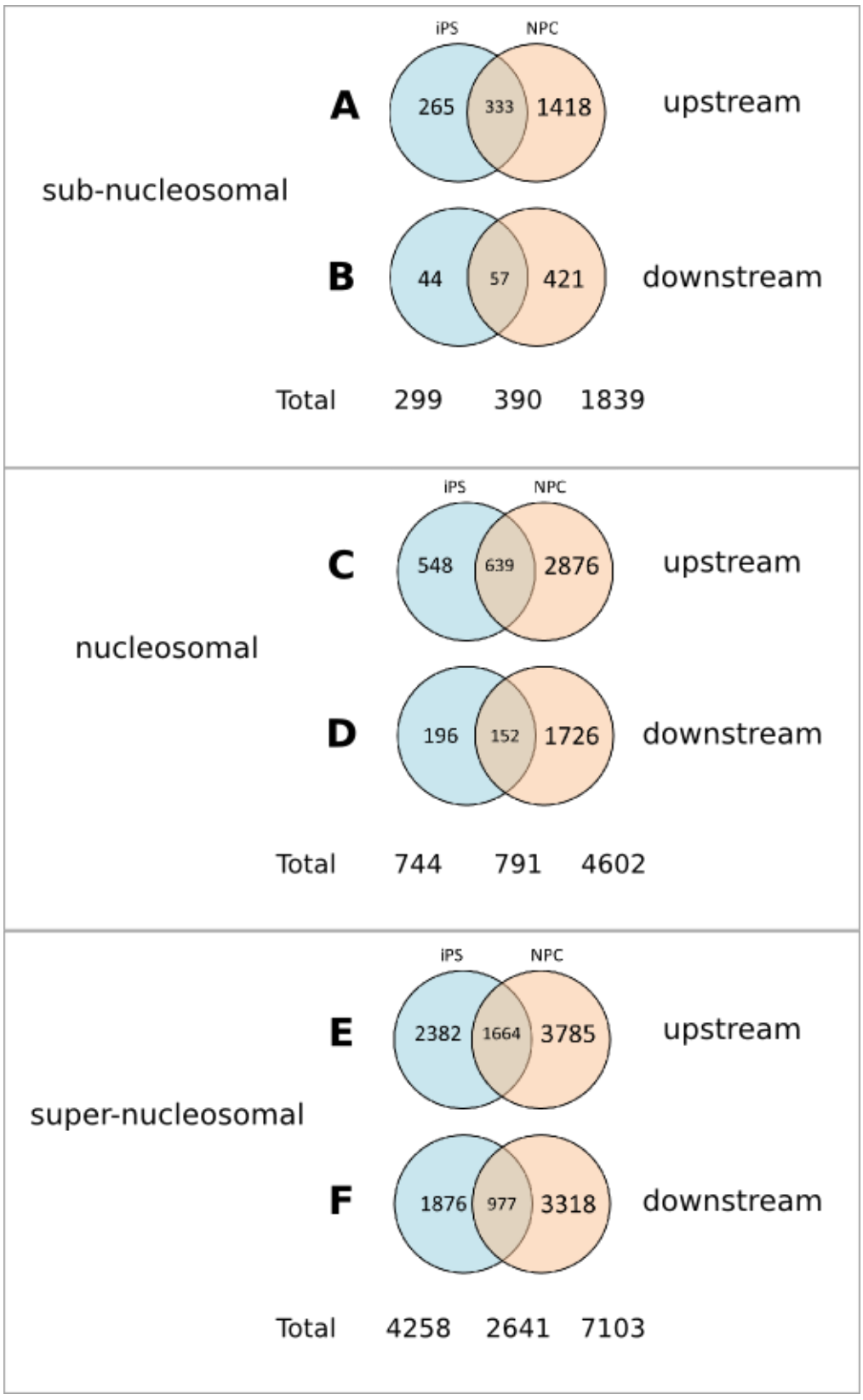


Fig 4.2 for legend see next page.

Figure 4.2 Nucleosome positioning at transcriptional start sites (TSS) in iPS and NPC cells.

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were given explicit genome positions using the heuristic peak-finding algorithm (PeakFinder). Venn diagrams show the numbers of TSS that have at least one positioned chromatin particle within + 300 bp "upstream" (A, C and E) and -300 bp, "downstream" (B, D and F) from a non-redundant list of all human TSS (n=84,142) exclusively in one cell type and in common between iPS and NPC cells in all three chromatin particle size classes: sub-nucleosomal (112-137 bp), nucleosomal (138-161 bp) and super-nucleosomal (162-188 bp). At the bottom of each panel the "total" figure represents the total number of positioned chromatin particles (+/-300 bp from a TSS) exclusively in one cell type and in common between iPS and NPC cells for each chromatin particle size class.

4.3 Detection of changes in chromatin particle positioning at individual genes during neural cell differentiation.

As the results from in earlier work in this chapter suggested that the nucleosomal chromatin particle size class is the most dynamic during neural cell development, it was decided to examine the patterns of chromatin particle positioning in the nucleosomal chromatin particle size class at individual genes. This would validate whether it was possible to detect loss and gain of individual nucleosomal chromatin particles at TSS during early neural development.

The paired-read midpoint position frequency data at and surrounding the TSS of individual genes at which loss and gain of nucleosomal chromatin particles nucleosomes was detected were visualised using the IGB. Then data was extracted from the .sgr files containing the paired-read midpoint position frequency data in 10 bp bins, at and surrounding representative TSS and plotted as histograms for all three chromatin particle size classes in both iPS and NPC cells.

Fig 4.3 and Fig 4.4 shows examples of individual genes where there is a positioned nucleosomal chromatin particle at a TSS in iPS cells, but not in NPC cells. Fig 4.3 shows a section of NANOG gene, in which there is a peak representing a positioned nucleosomal chromatin particle near the TSS that is present only in iPS cells. Interestingly, this peak overlaps with a positioned sub-nucleosomal chromatin particle in iPS cells, but it is not seen in the sup-nucleosomal particles. This suggests that the predominant peak at the NANOG TSS in iPS cells is smaller than the size of a human super-nucleosome. NANOG is a transcriptional regulator whose expression is involved in the maintenance of pluripotency (Boyer, Lee et al. 2005) and it is expressed in iPS but not in NPC cells. A positioned nucleosome was detected just upstream (-1) of the NANOG gene in iPS cells that is completely absent in NPC cells. The location of the nucleosome detected in iPS cells is in an important regulatory region of the NANOG gene that may be involved in expression of alternative NANOG transcripts (Das, Jena et al. 2011).This result is consistent with the idea that a nucleosome positioned in the promoter of the NANOG gene is necessary for the expression of NANOG in pluripotent cells.

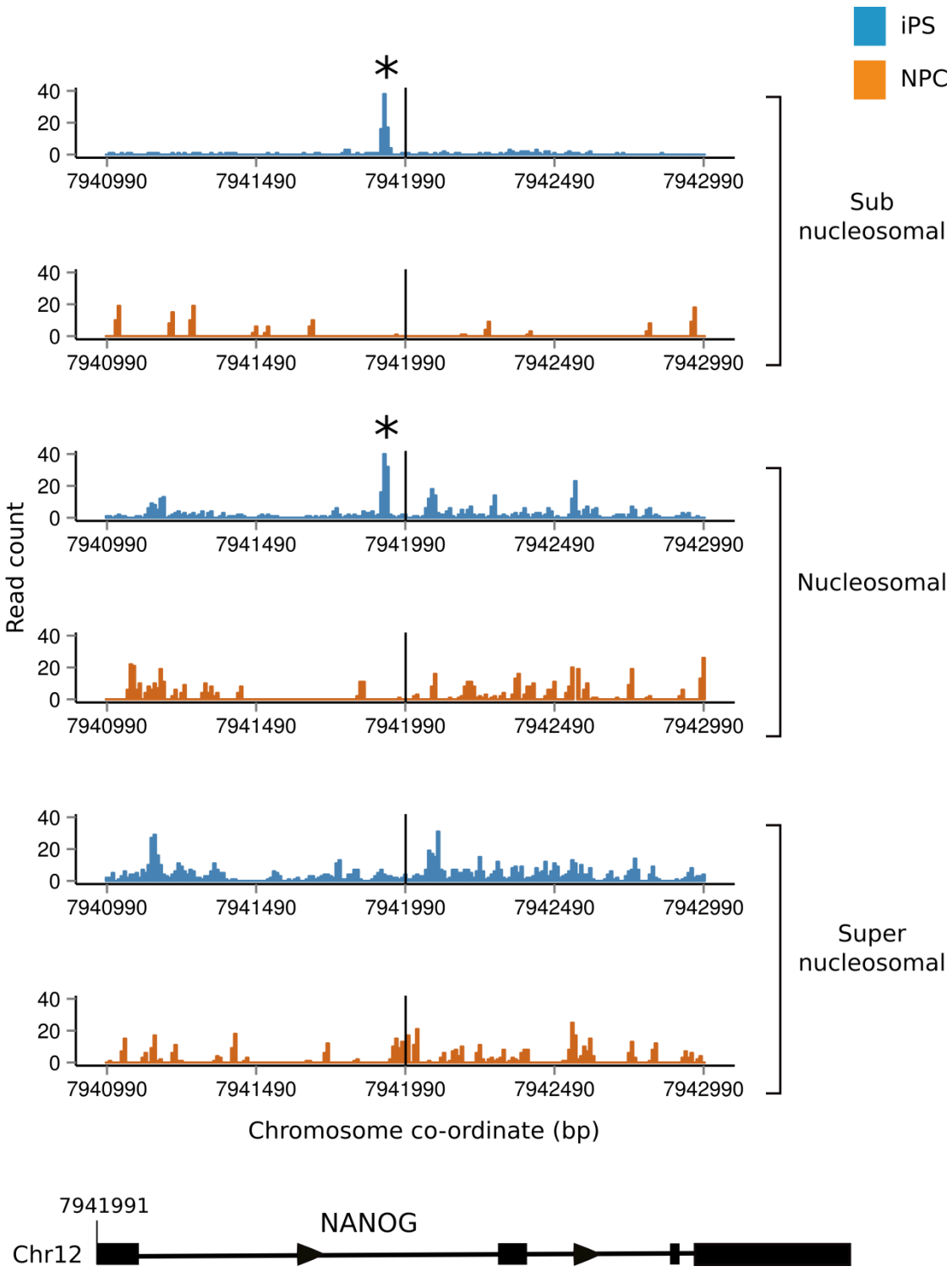


Fig 4.3 see next page for legend

Figure 4.3 Positioned Chromatin particles are detected upstream of the TSS in the NANOG gene in iPS cells.

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were detected at and surrounding TSS using the heuristic peak-finding algorithm (PeakFinder). Histograms were plotted of counts of paired-read midpoint positions in 10 bp bins from a section of chromosome 12 for all three chromatin particle size classes: sub-nucleosomal (112-137 bp), nucleosomal (138-161 bp) and super-nucleosomal (162-188 bp) for both iPS and NPC cells. A strongly positioned chromatin particle is visible just upstream of the NANOG gene TSS exclusively in iPS cells (*), in the sub-nucleosomal and nucleosomal chromatin particle size classes. A map of the 6,661 bp NANOG gene is depicted at the bottom of the panel. Exons are represented by filled black boxes and the direction of transcription by a black arrow head. The position of the TSS is represented by a black vertical line.

Fig 4.4 shows a section of ELF3 gene, in which there is a peak representing a positioned nucleosomal chromatin particle at the TSS that is present only in iPS cells. Interestingly, this peak overlaps with both a positioned sub-nucleosomal chromatin particle and a positioned super-nucleosomal chromatin particle in iPS cells, but it is not seen in any of the three chromatin particle size classes in NPC. ELF3 is a member of the ETS transcription factors that is thought to be expressed specifically in epithelial tissues but is also thought to play a role in early development (Oliver, Kushwah et al. 2012). This protein seems to have many and various roles including repressing androgen receptor activity (Shatnawi, Norris et al. 2014), but no specific role in neural development. Hence it would not be expected that this gene would be expressed in NPC cells and the lack of chromatin structure near the TSS of this gene in NPC cells may reflect this.

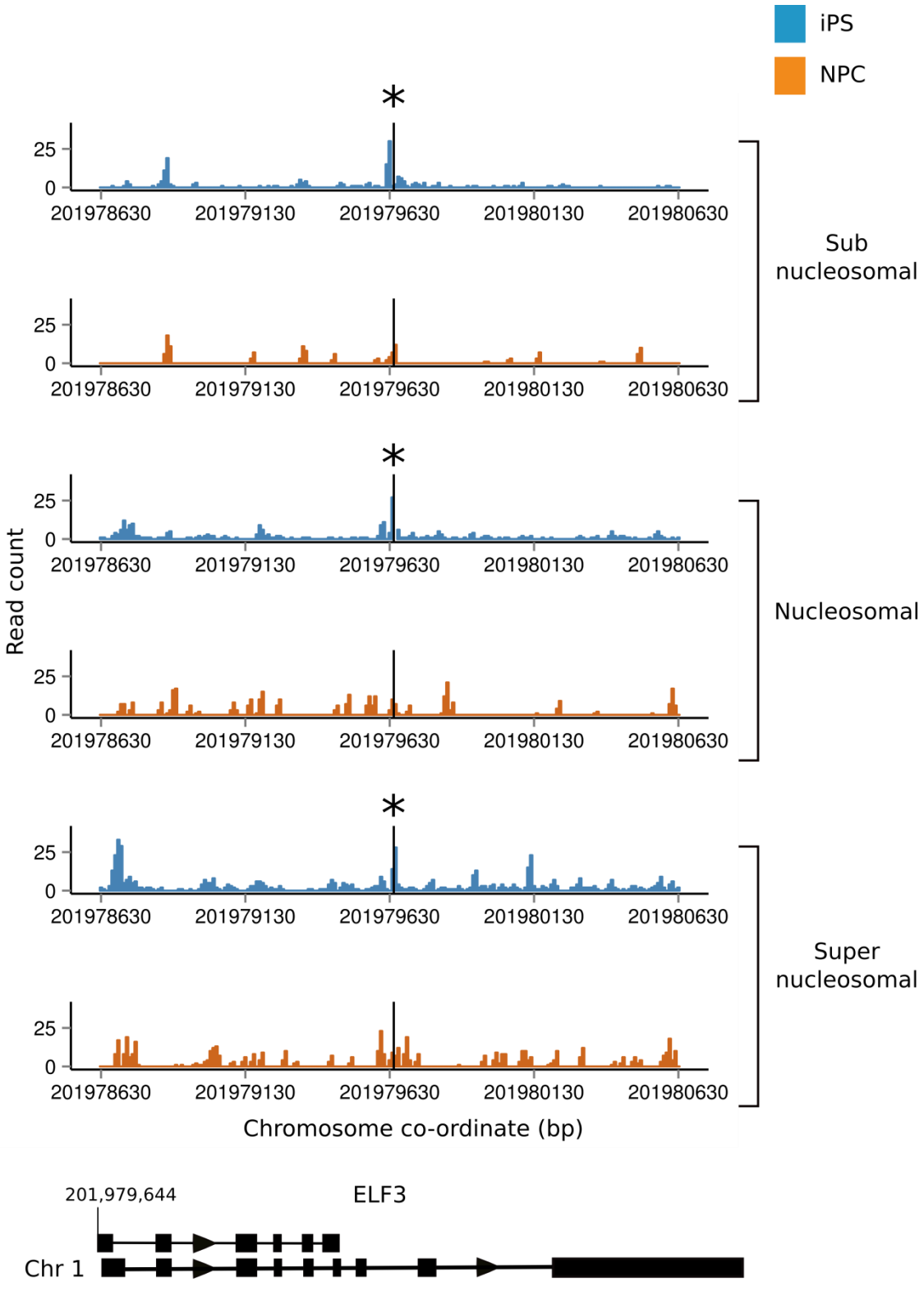


Fig 4.4 See next page for legend

Figure 4.4 Positioned Chromatin particles are detected at the TSS in the ELF3 gene in iPS cells.

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were detected at and surrounding TSS using the heuristic peak-finding algorithm (PeakFinder). Histograms were plotted of counts of paired-read midpoint positions in 10 bp bins from a section of chromosome 1 for all three chromatin particle size classes: sub-nucleosomal (112-137 bp), nucleosomal (138-161 bp) and super-nucleosomal (162-188 bp) for both iPS and NPC cells. A strongly positioned chromatin particle is visible at the TSS of the short form of the ELF3 gene TSS in iPS cells (*), but not in NPC cells, in all three chromatin particle size classes. A map of the 6,670 bp ELF3 gene is depicted at the bottom of the panel. Exons are represented by filled black boxes and the direction of transcription by a black arrow head. The position of the TSS is represented by a black vertical line.

A section of the GRIA1 (GluR-1) and NEGR1 genes are shown in Fig 4.5 and Fig 4.6 respectively. In both genes there is a nucleosomal chromatin particle positioned downstream of the TSS (+1 nucleosome), only in NPC cells. In the case of the NEGR1 gene, there is a small peak in the super-nucleosomal chromatin particle size class in the same place. However, for both of these genes there is a dominant peak in the nucleosomal chromatin particle size class only in NPC. The GRIA1 gene encodes an AMPA receptor subunit. These glutamate receptors are receptors for neurotransmitters in the brain. NEGR1 (neuronal growth regulator 1) is a neural growth promoting factor that is involved in dendrite growth (Pischedda, Szczurkowska et al. 2014). The NPC-specific chromatin structure detected here may be important in the neural-specific regulation of these genes.

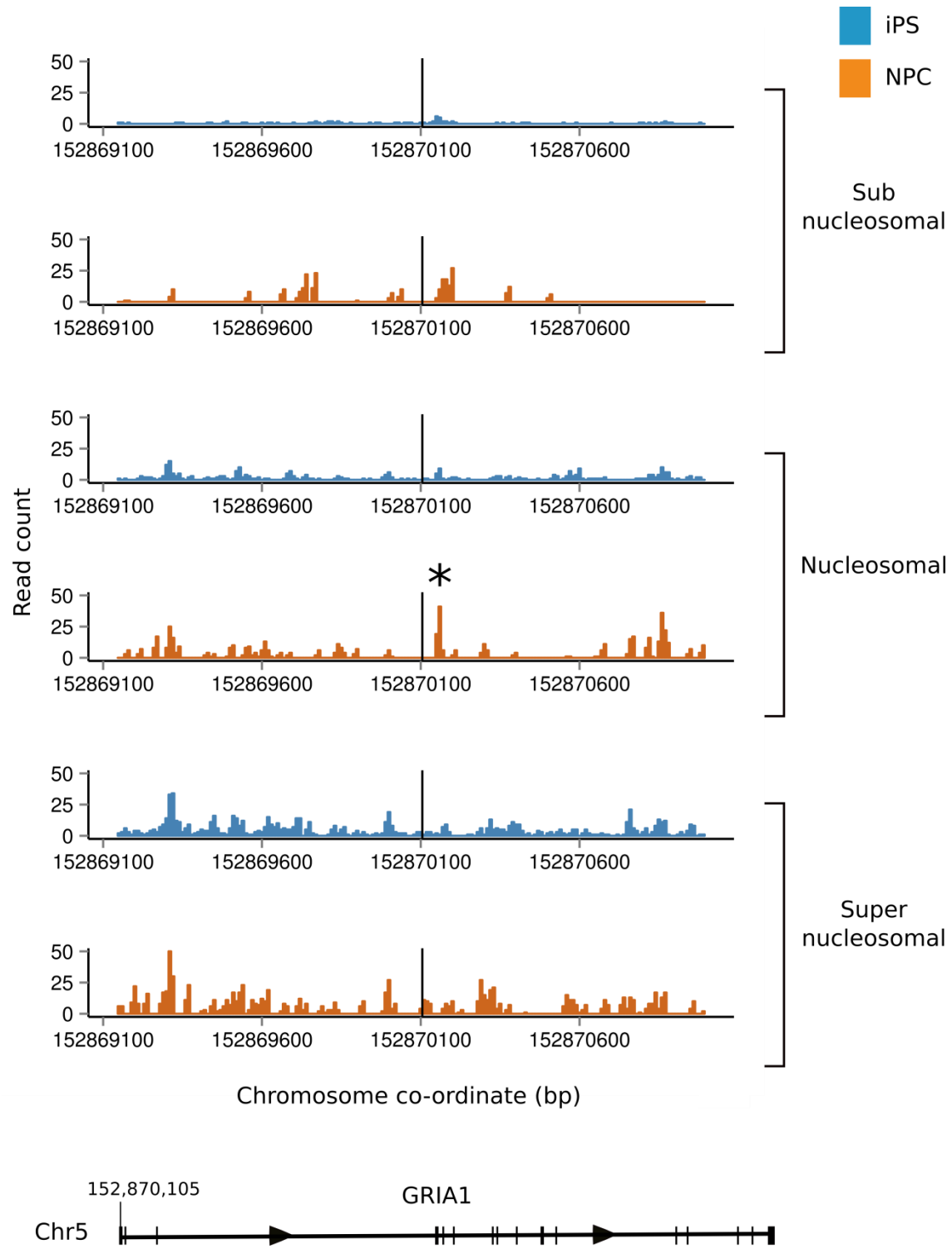


Fig 4.5 See next page for legend.

Figure 4.5 Positioned Chromatin particles are detected at the TSS in the GRIA1 gene in NPC cells.

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were detected at and surrounding TSS using the heuristic peak-finding algorithm (PeakFinder). Histograms were plotted of counts of paired-read midpoint positions in 10 bp bins from a section of chromosome 5 for all three chromatin particle size classes: sub-nucleosomal (112-137 bp), nucleosomal (138-161 bp) and super-nucleosomal (162-188 bp) for both iPS and NPC cells. A strongly positioned chromatin particle is visible downstream of the TSS of the GRIA1 gene TSS in NPC cells (*), but not in iPS cells, in the nucleosomal chromatin particle size class. A map of the 323 kb GRIA1 gene is depicted at the bottom of the panel. Exons are represented by filled black boxes and the direction of transcription by a black arrow head. The position of the TSS is represented by a black vertical line.

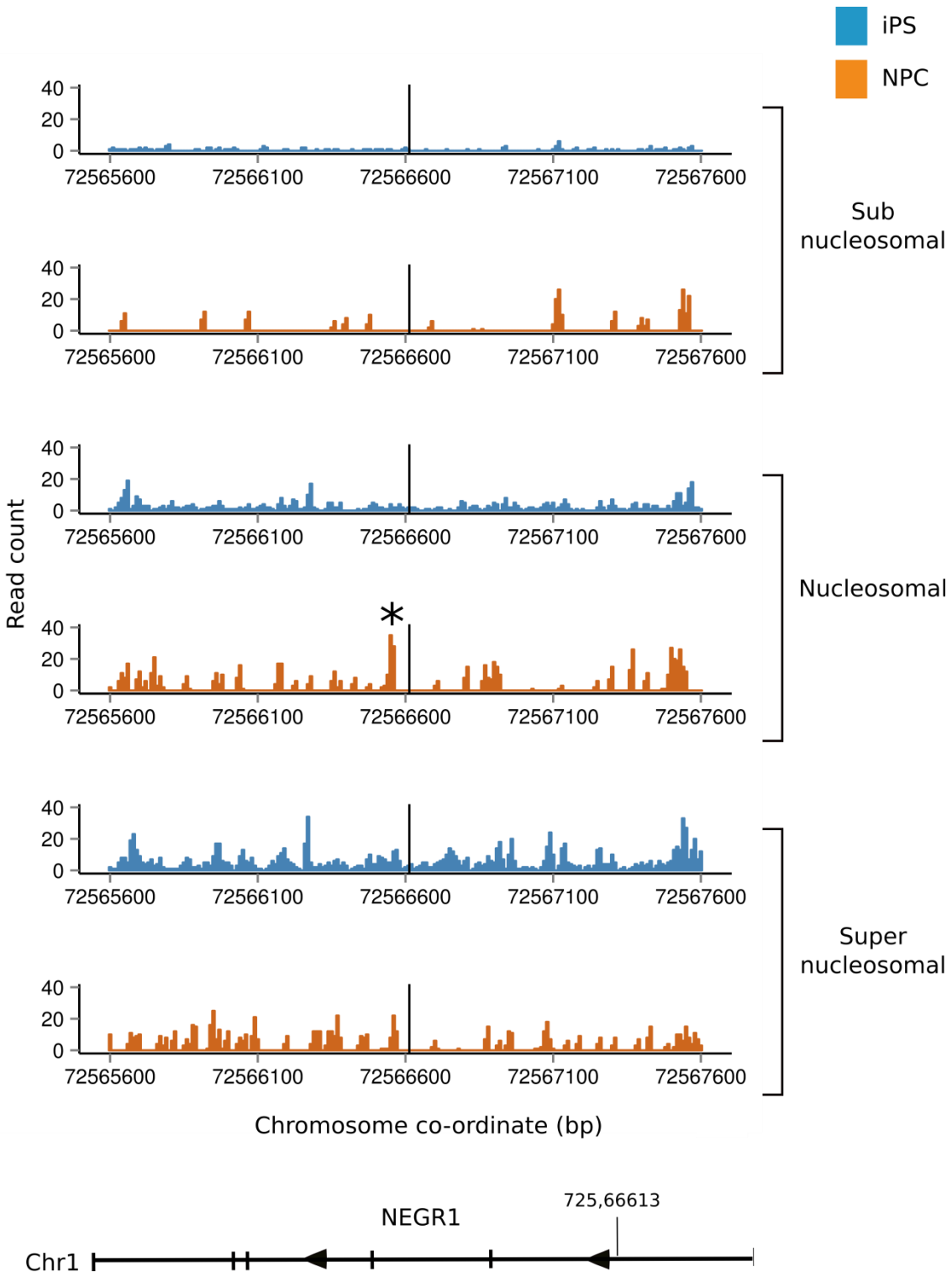


Fig 4.6 See next page for legend.

Figure 4.6 Positioned Chromatin particles are detected at the TSS in the NEGR1 gene in NPC cells.

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were detected at and surrounding TSS using the heuristic peak-finding algorithm (PeakFinder). Histograms were plotted of counts of paired-read midpoint positions in 10 bp bins from a section of chromosome 1 for all three chromatin particle size classes: sub-nucleosomal (112-137 bp), nucleosomal (138-161 bp) and super-nucleosomal (162-188 bp) for both iPS and NPC cells. A strongly positioned chromatin particle is visible upstream of the TSS of the NEGR1 gene TSS in NPC cells (*), but not in iPS cells, in the nucleosomal chromatin particle size class. A map of the 879 kb NEGR1 gene is depicted at the bottom of the panel. Exons are represented by filled black boxes and the direction of transcription by a black arrow head. The position of the TSS is represented by a black vertical line.

Fig 4.7 shows a section of the PSMD4 gene. In this case the chromatin particle positioning in all three size classes is very similar in both cell types. Peaks in the paired-read midpoint position frequency in both the nucleosomal and super-nucleosomal chromatin particle size classes occur at the TSS of the PSMD4 gene in both cell types. However, sub-nucleosomal chromatin particles are not positioned at the TSS of this gene in either cell type.

Fig 4.8 shows a section of the PIK3R2 gene. At and surrounding the PIK3R2 TSS, the patterns of chromatin particle positioning are similar in each chromatin particle size class in both iPS cells and NPC.

In both cases, the pattern of chromatin particle positioning at and surrounding the TSS in these two genes does not change during differentiation from iPS to NPC. PSMD4 is a 26S proteasome regulatory subunit and PIK3R2 (PI3-kinase regulatory subunit beta) is involved many cellular functions such as cell growth, proliferation, differentiation, and glucose metabolism. Since these two proteins are involved in general cellular function, it might be expected that they are expressed in both of the cell types analysed here.

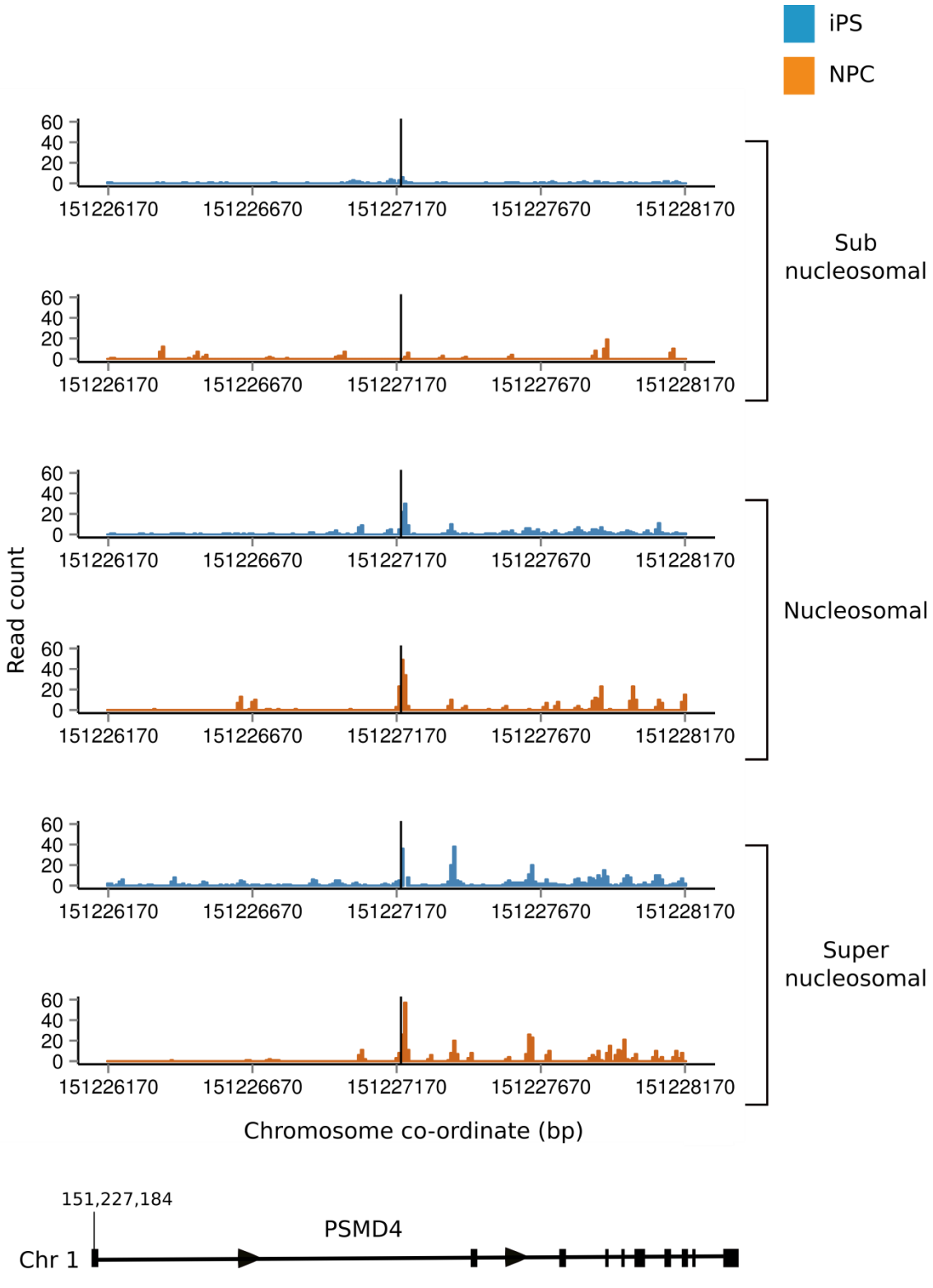


Fig 4.7 See next page for legend.

Figure 4.7 Chromatin particles positioning patterns are similar in all chromatin particle size classes in both cell types at the TSS of the PSMD4 gene

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were detected at and surrounding TSS using the heuristic peak-finding algorithm (PeakFinder). Histograms were plotted of counts of paired-read midpoint positions in 10 bp bins from a section of chromosome 1 for all three chromatin particle size classes: sub-nucleosomal (112-137 bp), nucleosomal (138-161 bp) and super-nucleosomal (162-188 bp) for both iPS and NPC cells. Chromatin particle positioning is visible at the TSS of the PSMD4 gene TSS in both iPS cells and NPC cells. A map of the 12 kb PSMD4 gene is depicted at the bottom of the panel. Exons are represented by filled black boxes and the direction of transcription by a black arrow head. The position of the TSS is represented by a black vertical line.

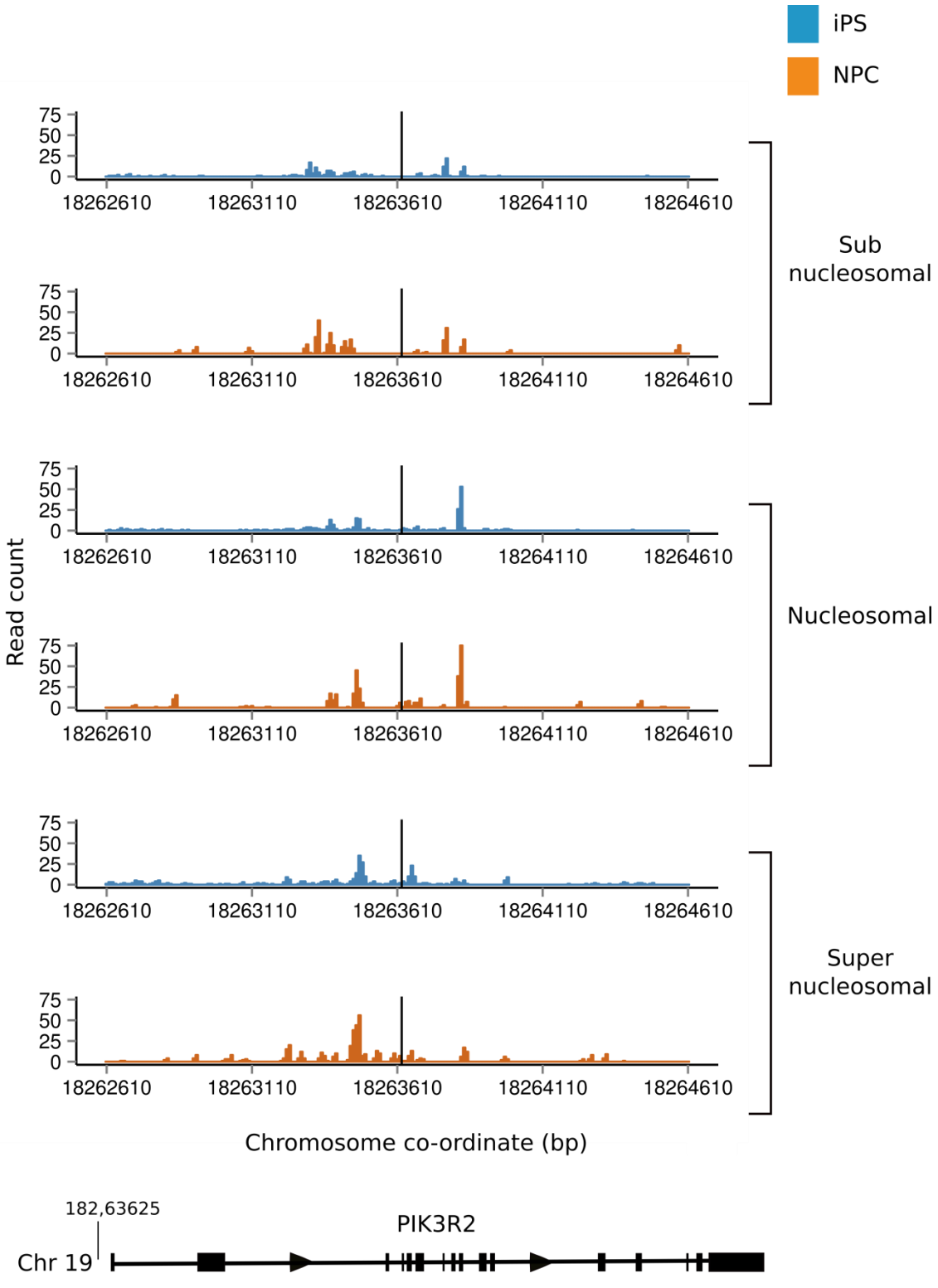


Fig 4.8 See next page for legend.

Figure 4.8 Chromatin particles positioning patterns are similar in all chromatin particle size classes in both cell types at the TSS of the PIK3R2 gene.

Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were detected at and surrounding TSS using the heuristic peak-finding algorithm (PeakFinder). Histograms were plotted of counts of paired-read midpoint positions in 10 bp bins from a section of chromosome 19 for all three chromatin particle size classes: sub-nucleosomal (112-137 bp), nucleosomal (138-161 bp) and super-nucleosomal (162-188 bp) for both iPS and NPC cells. Similar patterns of chromatin particle positioning are visible surrounding the TSS of the PIK3R2 gene within each chromatin particle size class in both iPS cells and NPC cells. A map of the 17 kb PIK3R2 gene is depicted at the bottom of the panel. Exons are represented by filled black boxes and the direction of transcription by a black arrow head. The position of the TSS is represented by a black vertical line.

4.5 Patterns of chromatin particle positioning at individual loci correlate with gene expression.

Wu *et al* have examined changes in gene expression during human neural cell development using RNA-seq (Wu, Habegger et al. 2010). Wu *et al* used hESC cells and hESC cells differentiated to three stages of differentiation; N1 (early initiation), N2 (neural progenitor), and N3 (early glial-like). The data presented in thesis was generated from hiPSC cells and hiPSC cells differentiated to NPC. The method of generating NPC cells in both cases was by embryoid body formation, therefore the hESC and hESC developed to N2 cells are similar to the cells used in the work of this thesis. The expression level of the six genes whose chromatin structure at TSS was examined in section 4.3 was determined from the data published by Wu *et al* and the results are shown in Table 4.1. Expression of the NANOG and ELF3 genes is high in hESC cells and decreases in hESC-N2 cells. This correlates with the loss of the positioned nucleosome at the TSS of these genes during differentiation from iPS to NPC in this study. Similarly, expression of the GRIA1 and NEGR1 genes is high in hESC-N2 cells and decreases in hESC cells. This correlates with the gain of the positioned nucleosome at the TSS of these genes during differentiation from iPS to NPC in this study. In the case of the PSMD4 and PIK3R2 genes, there is little change (approximately 2 fold) in gene expression during differentiation from hESC cells to hESC-N2, nor any change in chromatin particle positioning during differentiation from iPS to NPC.

These results suggest that in the case of the six genes shown here, patterns of chromatin particle positioning may correlate with gene expression in iPS and NPC. This suggests that loss and gain of positioned chromatin particles at TSS are important in gene regulation and that this regulation may influence the progression from pluripotent iPS cells to neural progenitor cells.

Gene symbol	Fold change gene expression. N2-A/hESC	Direction of change in gene expression hESC to N2-A	Positioned nucleosome at TSS in iPS cells	Positioned nucleosome at TSS in NPC cells
NANOG	0.0009	down	✓	✗
ELF3	0.0358	down	✓	✗
GRIA1	265.2	up	✗	✓
NEGR1	12.625	up	✗	✓
PSMD4	0.416	down	✓	✓
PIK3R2	1.973	up	-	-

Table 4.1 Correlation of cell type specific chromatin particle positioning in iPS and NPC cells with gene expression data derived from hESC and N2-A cells.

Six representative TSS were visualised using the IGB to determine their chromatin structure in iPS and NPC (shown in Figs 4.3-4.8). The presence (tick), absence (cross) of a positioned chromatin particle within 300 bp of the TSS is indicated. The absence of any strong chromatin particle positioning within 300 bp of the TSS is indicated by a dash. Gene expression data was taken from data published by Wu *et al* and it is shown as the ratio of the expression values for hESC and N2-A cells (Wu, Habegger et al. 2010).

4.6 Chromatin structure at transcription factor binding sites.

In addition to investigating changes in chromatin particle positioning at transcriptional start sites, changes in chromatin particle positioning at transcription factor binding sites was investigated. The genome-wide binding positions for a select group of transcription factors which are involved in either in chromatin remodelling or neural development were determined. These were, ATF2, YY1 and PAX6 (See chapter1, section 1.18). Fig 4.9 shows the average cumulative frequency distributions for all three chromatin particle size classes at and surrounding the ATF2, YY1 and PAX6 binding sites. Positioned chromatin particles were not detected at or surrounding any of these transcription factor binding sites. In addition, a change in chromatin structure during differentiation from iPS to NPC cells was not detected. It seems surprising that there are no changes in chromatin structure surrounding these sites during differentiation, since all of these transcription factors are thought to have functions related to chromatin remodelling or neural development. In the case of YY1, positioned nucleosomes have been shown to exist at YY1 sites in GM12878 at sites located distally to annotated transcripts (Wang, Zhuang et al. 2012). It is possible that only a specific subset of these transcription factor binding positions are utilised at this stage in differentiation and if the nucleosomes were offset in position, then the results would be masked by averaging the chromatin structure across all sites. This could be resolved by clustering the data into subsets of sites according to the chromatin structure surrounding the sites. Work by Kundaje *et al* (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012) showed recently, that in K562 cells at EGR1 sites, no particular pattern of chromatin structure was detected in an aggregation plot, but when the data was clustered, a specific patterns of nucleosome positioning were detected at small subsets of sites. However, in the work presented here, there are no gross global changes in chromatin structure occurring at these transcription factor binding positions during differentiation from iPS to NPC cells.

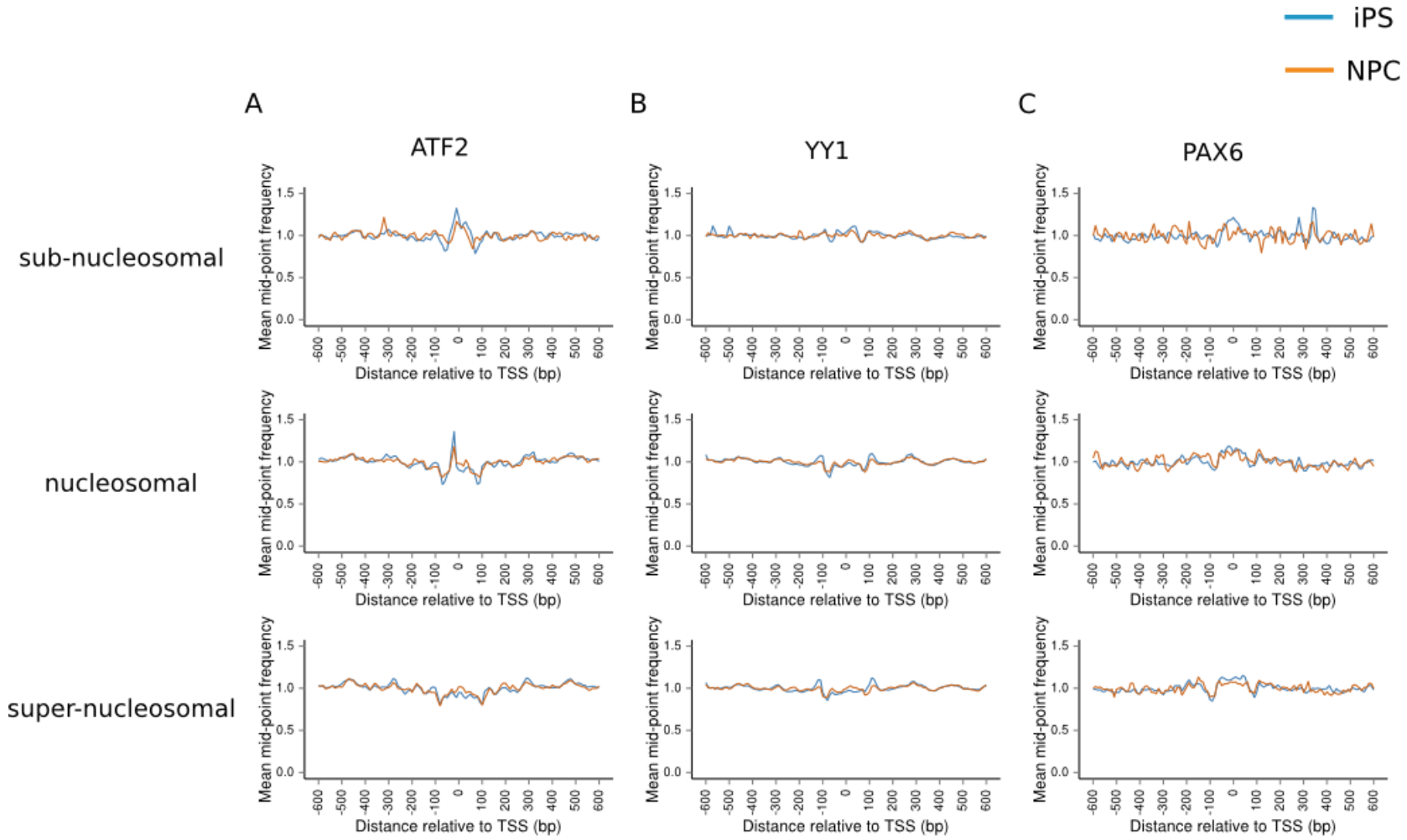


Fig 4.9 See next page for legend.

Figure 4.9 Nucleosomes are not positioned at or surrounding several transcription factor binding sites that are involved in remodelling or neural development.

Average cumulative frequency distributions at and surrounding transcription factor binding sites (+/-600 bp) for all three chromatin particle size classes: 112-137 bp (sub-nucleosomal), 138-161 bp (nucleosomal) and 162-188 bp (super-nucleosomal) at A. ATF2 (n=9,881), B.YY1(n=39,945) and C. PAX6 (n=1,432) binding sites.

Discussion.

In this chapter, unbiased searching has revealed a spectrum of examples of cell-type-specific nucleosome positioning at and surrounding TSS. Nucleosome positioning surrounding TSS at the level of individual genes has shown that loss and gain of individual nucleosomes between the two cell types, iPS and NPC, can be detected. In addition, specific patterns of chromatin structure at the TSS of genes associated with pluripotency (NANOG) and neural-specific genes (GRIA1 and NEGR1) have been shown. This variation in chromatin structure at TSS may be important in gene regulation during the progression from pluripotent to neural progenitor cells.

In this chapter it was shown that 1.3% of the positioned nucleosomes in the NPC genome are positioned in close proximity to transcriptional start sites, therefore a large proportion of the nucleosomes in the human genome are positioned in locations other than at TSS, for example at transcription factor binding sites, enhancers and insulator binding sites.

An initial study of a selected set of transcription factor binding sites that are involved either in chromatin remodelling or neural development, where it might be expected that chromatin structure would change during neural cell differentiation, showed that there was little change in chromatin structure at these regulatory sites during differentiation from iPS to NPC cells. This suggested that changes in chromatin structure do not influence the activity of these factors during neural cell development. However, it has been shown that positioned nucleosomes exist at RE1/NRSF binding sites in the human genome in CD4+ cells (Valouev, Johnson et al. 2011). REST is involved in the regulation of neurogenesis by repressing neuronal gene expression in non-neuronal cells (Wu and Xie 2006) (Johnson, Teh et al. 2008) (Jorgensen, Terry et al. 2009), acting through binding to RE1 elements and then recruiting co-repressors (Ballas, Grunseich et al. 2005). Therefore, the work in the next chapter examines whether there are changes in the positions of chromatin particles during differentiation from iPS to NPC cells at RE1/NRSF binding sites.

:

Chapter 5 : Chromatin structure at REST binding motifs during neural cell differentiation

5.1 Introduction

It has been shown that nucleosomes are positioned at and surrounding RE1 sites, where REST binds, in human lymphoblastoid cells (Valouev, Johnson et al. 2011) and recently in K562 and GM12858 cells (Nie, Cheng et al. 2014). However, changes in nucleosome positioning during neural cell development in human cells has not been investigated previously. The work in this chapter investigated whether it was possible to detect changes in i) MNase-protected species at RE1 sites, ii) chromatin structure flanking RE1 sites and iii) whether it was possible, through the analysis of changes in chromatin structure, to identify genes that are targeted by REST during neural cell development.

Putative mammalian targets of REST, initially defined by the presence of a REST binding motif, and more recently defined using ChIP-seq data, have been tested by using gene targeting and RNA interference experiments to alter the expression of REST and measure any resulting changes in the transcriptional activity of these putative target genes (Sun, Cooper et al. 2008; Bruce, Lopez-Contreras et al. 2009). REST target genes have been defined from studies in mouse, rat and human cells. Much of the early work was done in mouse cells but recently studies have been done in human cells (Rockowitz, Lien et al. 2014). A comparison of the REST target genes in mouse and human ES cells suggests that there are twice as many REST targets in human than mouse but that there is conservation of REST binding at a core, but not all, REST target genes (Rockowitz and Zheng 2015).

REST was reviewed in Chapter 1 section 1.18.5 and its role as a transcriptional repressor of neural genes in non-neuronal cells is summarised in Fig 5.1.

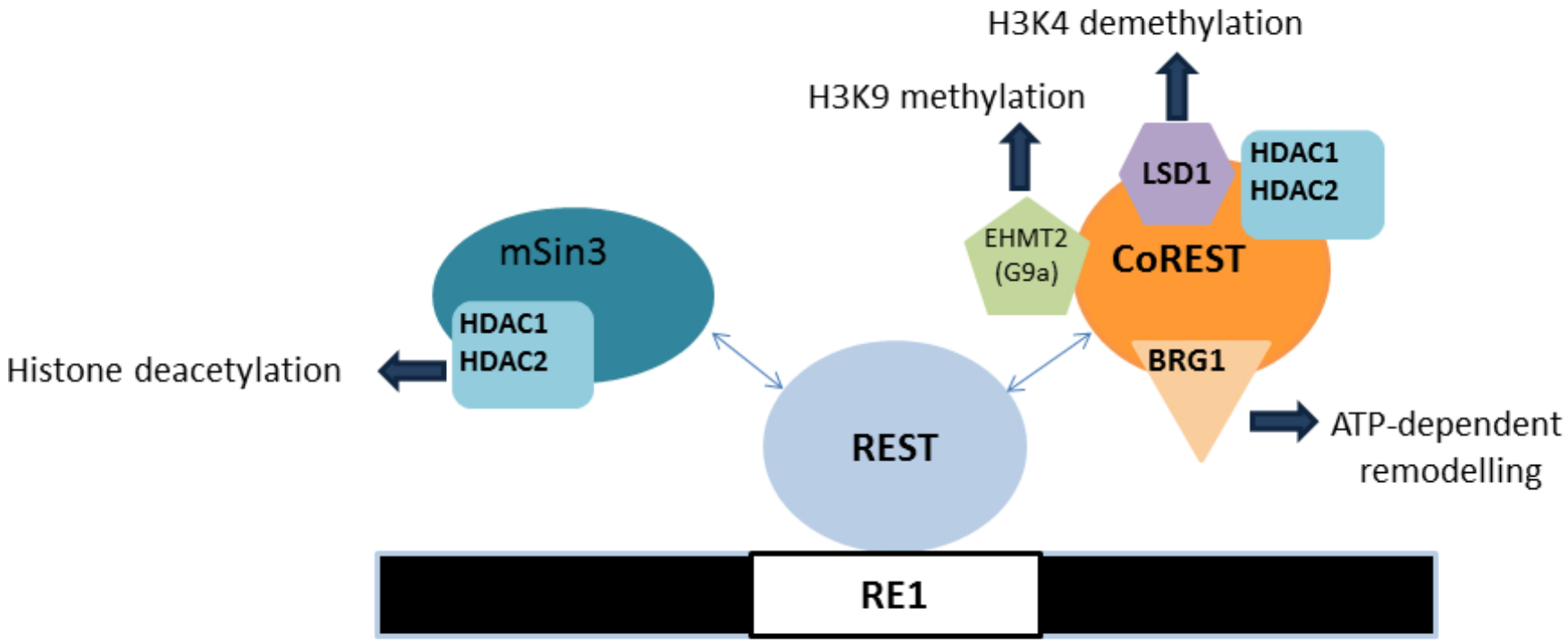


Figure 5.1 See next page for legend.

Figure 5.1 REST is a transcriptional repressor of neural genes in non-neuronal cells

REST mediates epigenetic change through the removal of 'activatory' chromatin marks and the addition of 'repressive 'chromatin marks. REST binds to RE1 elements and recruits co-repressors that have chromatin remodelling activity. BRG1, an ATP-dependent chromatin remodeller, recognises acetylated H4K8, leading to an increase in REST recruitment to RE1 sites. REST recruits the co-repressor complexes CoREST and mSin3, both of which contain the histone deacetylases HDAC1 and HDAC2. After H3K9 deacetylation, the H3K9 methytransferase EHMT2 (G9a) methylates H3K9, repressing transcription. LSD1, also a member of the CoREST complex, removes the mono and di-methylation chromatin marks that are associated with gene activation from H3K4.

5.2 Derivation of REST binding motifs

Traditionally the sequences of transcription factor binding motifs have been determined by using DNase foot-printing, Electrophoretic Mobility Shift Assay (EMSA) or reporter construct assays at individual genes. From this information, consensus binding sequences were derived and then, as whole genome sequences became known, the locations of these sequences across whole genomes were derived computationally. More accurate methods of deriving motifs utilise a position weight matrix (PWM), or position-specific scoring matrix (PSSM), using an algorithm that constructs a matrix of weights that can distinguish between true binding sites and other non-functional sites with similar sequences (Stormo, Schneider et al. 1982). However, it has become clear that not all computationally predicted binding locations are necessarily real binding sites or they may not be utilised in all cell types. Recently, large-scale studies of the genome-wide binding of transcription factors using ChIP-chip (Ren, Robert et al. 2000) and more recently, ChIP-seq (Johnson, Mortazavi et al. 2007) were undertaken in many different cell types to gain more accurate transcription factor binding site information (Gerstein, Kundaje et al. 2012). This information, combined with consensus binding sequences or position-specific scoring matrix information can be used to derive accurate, cell type specific locations of transcription factor binding motifs.

In this study, a set of positions in the genome where REST binds to DNA were derived by matching the 17 bp consensus binding motif NTYAGMTCCNNRGMSAG (Bruce, Donaldson et al. 2004) from the hg19 human genome autosomal .fasta files, with transcription factor binding ChiP-seq data from the ENCODE project (Gerstein, Kundaje et al. 2012). These Chip-seq data were derived from multiple cell lines (See Chapter 2 section 2.5.2). This provided a set of precise co-ordinates at the centre of each consensus NRSF/RE1 binding sequence at which it has been shown by ChiP-seq that REST binds in multiple cell types. The derived set of 871 REST binding elements used in this thesis herein are referred to as RE1 sites.

5.3 Detection of sub-nucleosomal particles at RE1 sites in iPS cells.

Genome-wide, cell-type specific changes in patterns of chromatin structure surrounding RE1 sites during neural cell development were analysed. The average cumulative frequency distributions of all three chromatin particle size classes (sub-nucleosomal, nucleosomal and super-nucleosomal) derived in this study were plotted at and surrounding RE1 sites. Fig 5.2A shows the average cumulative frequency distribution of the sub-nucleosomal (112-137 bp) chromatin particle size class at and surrounding RE1 sites. This analysis shows that there is a large peak in the frequency of these chromatin particles centred on the RE1 site in iPS cells. This peak is absent in NPC cells.

It has been shown in mouse ES cells that REST protein levels are high in ES cells and then decrease during the early stages of neuronal differentiation (Ballas, Grunseich et al. 2005; Singh, Kagalwala et al. 2008). This downregulation of REST may be mediated through degradation of REST through the ubiquitination pathway (Huang, Wu et al. 2011). Hence, the DNA binding protein detected at RE1 sites in induced human pluripotent cells in this study has the characteristics of REST or the REST complex.

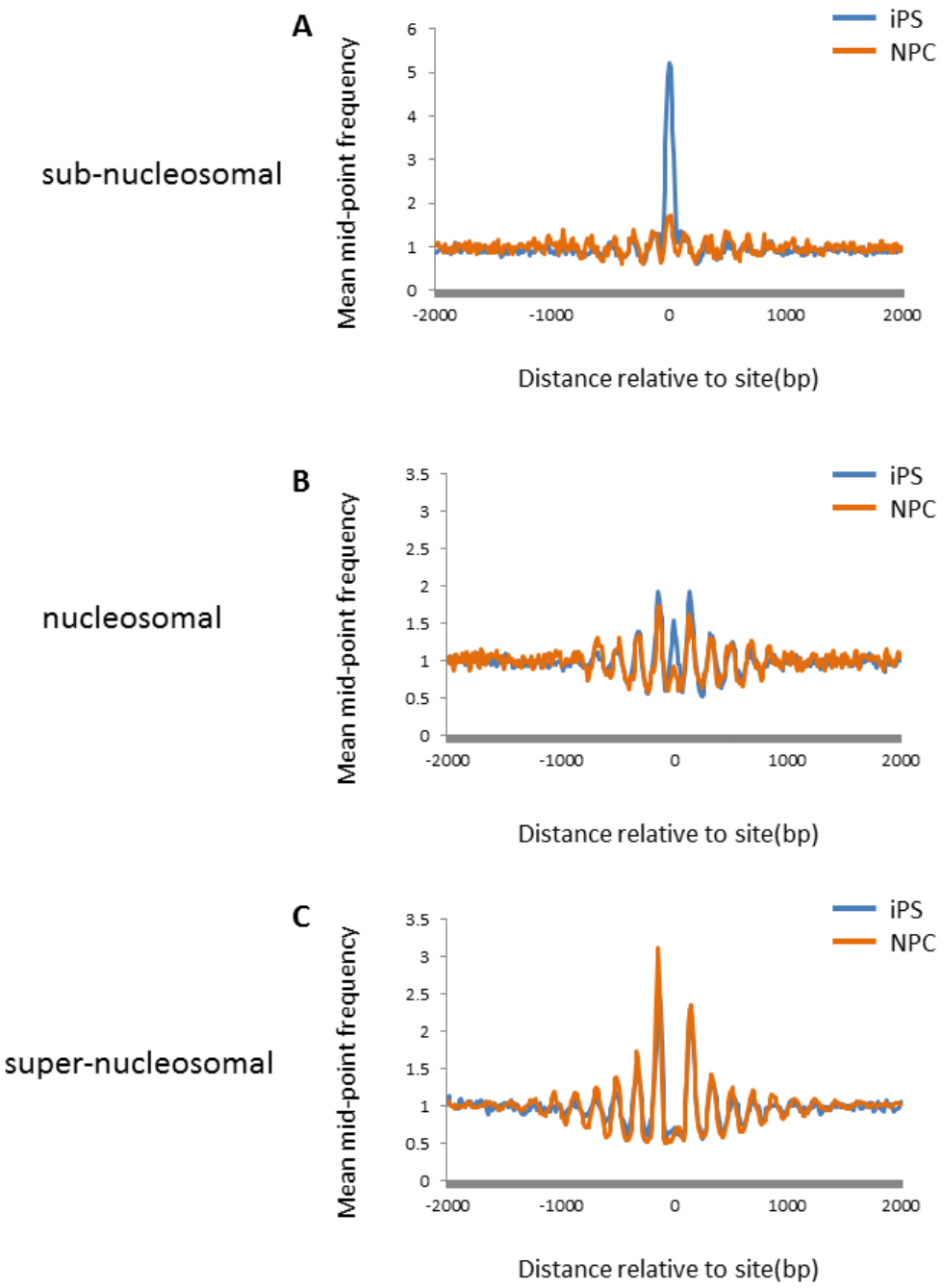


Figure 5.2 Chromatin remodelling occurs at RE1 sites during differentiation from pluripotent to neuronal precursor cells.

Normalised average cumulative frequency distributions of chromatin particle size classes A. 112-137 bp (sub-nucleosomes), B. 138-161 bp (representing nucleosomes) and C. 162-188 bp (super-nucleosomes) show the chromatin structure at and surrounding (+/-2000 bp) RE1 elements in both iPS and NPC cells (n=871).

5.4 Nucleosomes are positioned surrounding RE1 sites in both iPS and NPC cells.

Next, the average cumulative frequency distribution of nucleosomal (138-161 bp) and super-nucleosomal (162-188 bp) chromatin particle size classes at and surrounding RE1 sites was determined. The chromatin particle frequency distributions shown in Fig 5.2 B and C show that there are six distinct peaks on either side of the RE1 site in both cell types. This suggests that an array of six positioned nucleosomes flanks the RE1 site in both iPS and NPC cells.

A peak in the frequency of super-nucleosomal chromatin particles at the RE1 site itself was not detected in either cell type. However, a small peak, representing a high frequency of chromatin particles in the nucleosomal size class (138-161 bp) was found centred on the RE1 site in iPS cells. This peak is smaller than that shown at the RE1 site in the sub-nucleosomal (112-137 bp) chromatin particle size class, hence, the dominant peak at the RE1 site, likely representing the REST complex in pluripotent cells, is in the sub-nucleosomal size class.

It might be expected that REST binds to the RE1 site, recruits its co-repressors which have chromatin re-modelling activity and then nucleosomes become organised surrounding the RE1 site as a result of REST binding. These results are in agreement with this possibility in iPS cells. However, this result shows that nucleosomes remain positioned at RE1 sites during differentiation from iPS to NPC cells when the peak in the sub-nucleosomal (112-137 bp) chromatin particle size class is absent in NPC cells. This suggests that nucleosomes can remain positioned at RE1 sites in the absence of the sub-nucleosomal chromatin particle positioned at RE1 sites in NPC cells.

In order to validate the chromatin structure surrounding RE1 sites shown in this study, the chromatin structure surrounding RE1 sites in both K562 and GM12878 cells was investigated using the nucleosome maps derived by Kundaje *et al* (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012) that were re-processed for comparison purposes. The core nucleosome data (138-161 bp) from iPS and NPC cells in this study is most comparable with the published K562 and GM12878 nucleosome maps since, in these published maps, the predominant fragment length is the size of a mono-nucleosome.

Comparisons of the average cumulative frequency distributions of the (138-161 bp) chromatin particle size class in iPS versus K562 cells and iPS versus GM12978 cells are shown in Fig 5.3 A and C respectively. These results show that a similar number of positioned nucleosomes (approximately 6) flank the RE1 site in all three genomes, but the peak that is centred on the RE1 site in iPS cells, which is likely to be the REST complex, is absent in both K562 and

GM12978 cells. A comparison of the average cumulative frequency distributions of the nucleosomal (138-161 bp) chromatin particle size class in NPC versus K562 cells and NPC versus GM12978 cells is shown in Fig 5.3 B and D. These results show that a similar number (approximately 6) of positioned nucleosomes flank the RE1 site in all three genomes.

Taken together, the comparison of the analysis of patterns of chromatin particle positioning in NPC cells shown here and the results of the analysis of published chromatin maps from two other cell types, these results suggest that positioned nucleosomes flank RE1 sites in differentiated cells, independently of the presence of the sub-nucleosomal chromatin particle positioned at the RE1 site.

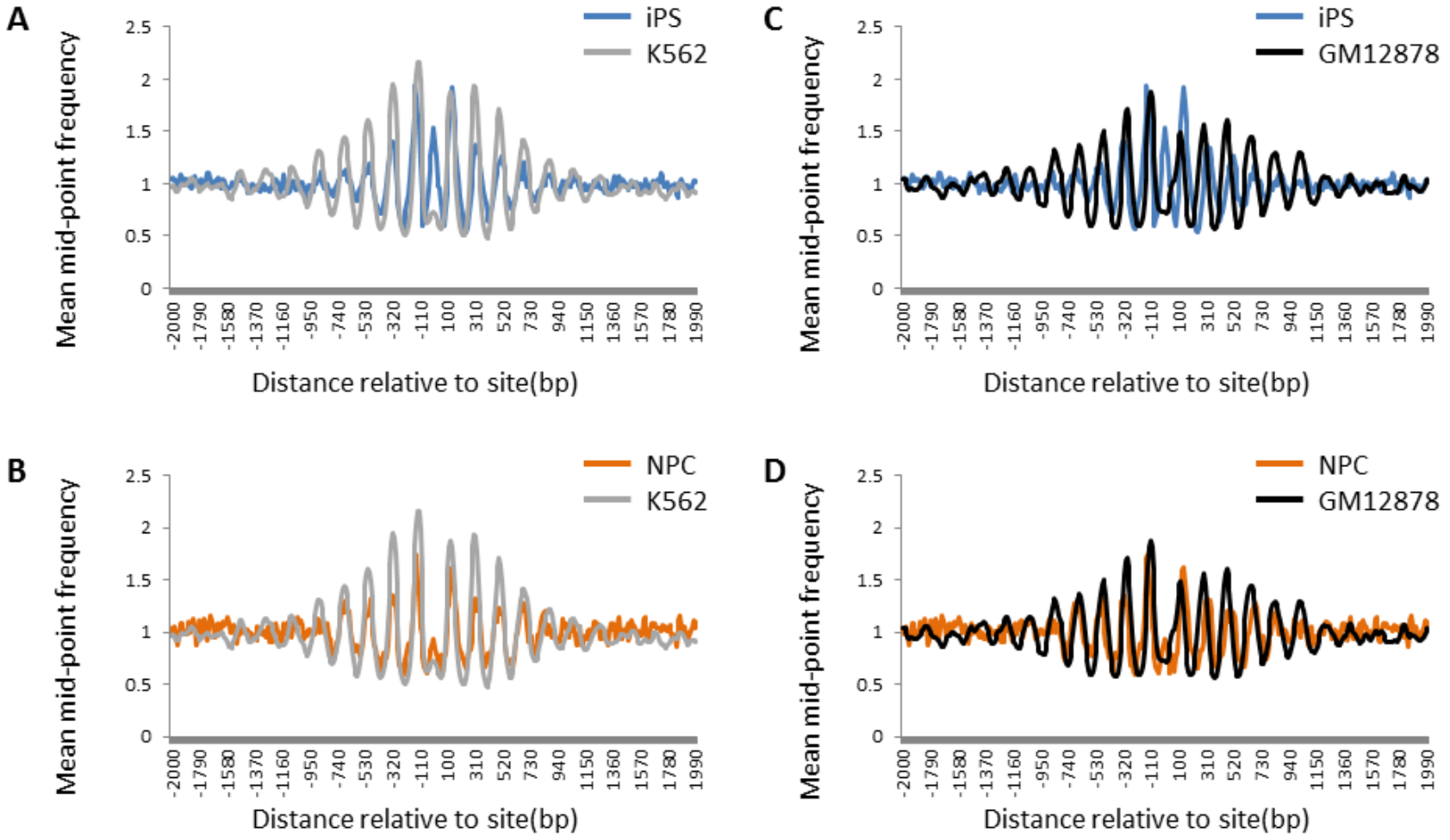


Fig 5.3 See next page for legend.

Figure 5.3 The pattern of positioned nucleosomes flanking RE1 sites is similar in several differentiated cell types.

Comparison of the positions of chromatin particles at and surrounding the RE1 site in the nucleosomal (138-161 bp) chromatin particle size range mapped in iPS and NPC cells by MNase-seq with those mapped in K562 and GM12878 cells. The data from the nucleosome maps generated for the K562 and GM12878 cell lines (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012) was converted from bedgraph to sgr format and the data was re-binned into 10 bp bins, calculating a 3bin moving average. The K562 and GM12878 maps were constructed by sequencing DNA purified from chromatin in the mono-nucleosome (approx. 150 bp) size range after MNase digestion. These data were compared with the iPS and NPC chromatin particle maps generated in the 138-161 bp size range since this is most similar to the size range of the chromatin particles in the K562 and GM12878 maps. An average cumulative frequency distribution of the chromatin particle distribution at and surrounding (+/- 2000 bp) each RE1 site was constructed and a comparison of the average cumulative frequency distribution of the chromatin particle distribution are shown in A. iPS and K562, B. NPC and K562, C. iPS and GM12878, D. NPC and GM12878.

5.5 A subset of RE1 sites are flanked by strongly positioned nucleosomes.

The average cumulative frequency distributions of chromatin particles at and surrounding a set of genomic feature positions represent an average chromatin structure at and surrounding the chosen genomic feature positions. Hence, it is possible that the pattern of chromatin particle positioning observed in Fig 5.2 was derived from a subset of RE1 sites.

In order to investigate the patterns of positioned chromatin particles in the nucleosomal chromatin particle size class in iPS cells further, the chromatin particle data in this size class surrounding RE1 sites within a window of +/-300 bp was clustered into eight groups. Cluster analysis of these data was undertaken as described previously. For each cluster, the pattern of chromatin structure surrounding RE1 sites was plotted as a heat map and an average cumulative frequency distribution. Fig 5.4 shows that approximately 50% (411) of the RE1 sites (clusters C2, C5 and C7) in the iPS genome possess positioned nucleosomal chromatin particles surrounding them. These sites were grouped together to form a subset of RE1 sites in shown in Fig 5.5 as 'Structured'. The average cumulative frequency distribution of positioned sub-nucleosomal particles (Fig 5.5 C) and super-nucleosomal particles (Fig 5.5D), was determined surrounding this subset of RE1 sites in iPS cells. These results show that where there was well-positioned core nucleosomes surrounding RE1 sites in iPS cells, there tended to be, on average, positioned sub-nucleosomal particles at the RE1 site and strongly positioned super-nucleosomal particles surrounding the site. The remaining RE1 sites (n=460) shown in clusters C8, C6, C3, C4 and C1 (Fig 5.4) which have fewer well-positioned nucleosomal particles, were grouped together and are labelled "Less structured" in Fig 5.5. The average cumulative frequency distribution of positioned sub-nucleosomal particles and super-nucleosomal particles was determined surrounding this second subset of RE1 sites, shown in Fig 5.5 F and G respectively. Where there were fewer positioned core nucleosomes surrounding RE1 sites in iPS cells shown by the graph in Fig 5.5 E, this correlates with fewer positioned sub-nucleosomal particles at the RE1 site (Fig 5.5 F) and the presence of less strongly positioned super-nucleosomal particles surrounding the site (Fig 5.5 G). From this analysis, strong sub-nucleosomal chromatin particle positioning occurs centred on a subset (64%) of RE1 sites in iPS cells. The presence of this sub-nucleosomal chromatin particle positioning correlates with the presence of precise nucleosome positioning flanking RE1 sites, both in the nucleosomal and super-nucleosomal chromatin particle size classes. On average, the remaining RE1 sites in iPS cells possess weakly positioned nucleosomes surrounding them (Fig 5.5 E and G). This correlates with weaker positioning of the sub-nucleosomal chromatin particle at the RE1 sites. This suggests that when the REST complex is present at a subset of RE1 sites in pluripotent cells, nucleosomes are organised surrounding the RE1 site.

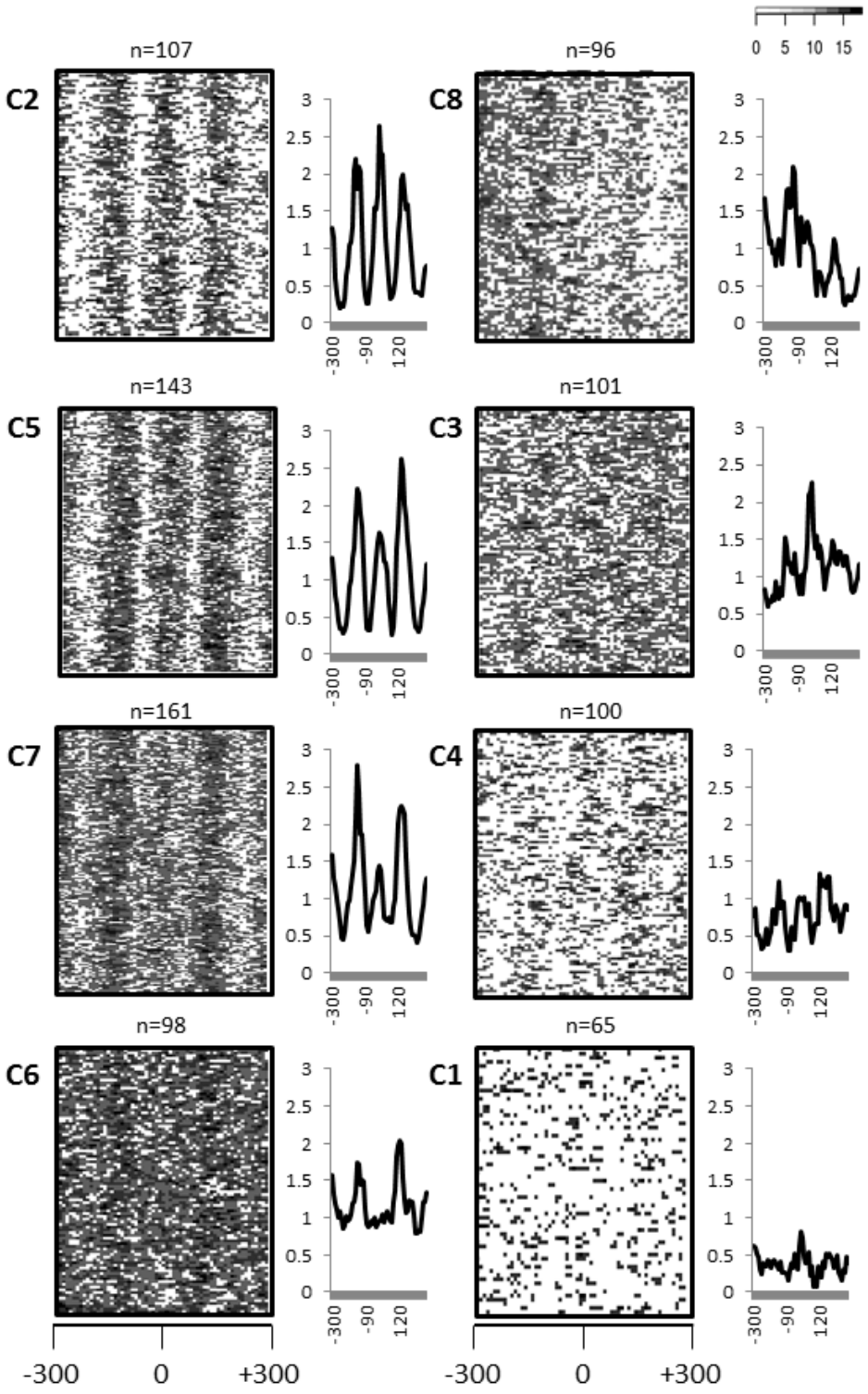


Fig 5.4 See next page for legend.

Figure 5.4 Cluster analysis reveals that chromatin is organised surrounding approximately half of the RE1 sites in iPS cells.

Data representing the positions of chromatin particles in the nucleosomal (162-188 bp) particle size class at and surrounding all 871 RE1 sites (+/- 300 bp) were clustered (k-means; Canberra distance) into eight groups (C1-C8). Each group is shown as a heat map with x-axis =distance relative to the RE1 site (bp) and y-axis =log2 of the locally-normalized dyad frequency values for every bin position. For each group, the mean chromatin particle frequency values centred on and surrounding the RE1 sites were plotted for the nucleosomal size class data and these are shown in the graphs on the right-hand side in each group, x-axis =distance relative to site (bp), y axis = mean mid-point frequency. The groups of RE1 sites that possess organised chromatin at RE1 sites were C2, C5 and C7, totalling 47% of the RE1sites.

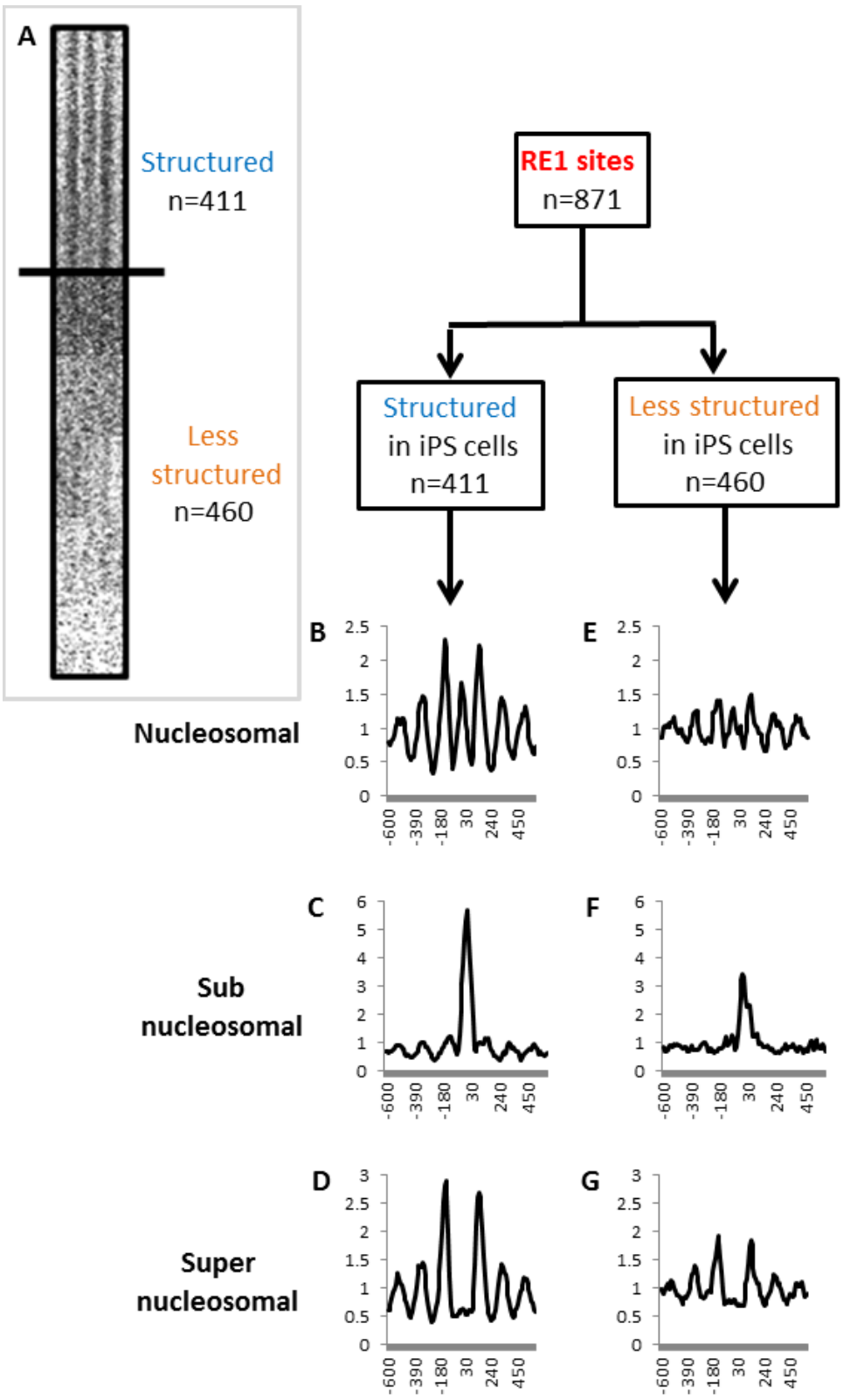


Fig 5.5 See next page for legend.

Figure 5.5 The presence of positioned nucleosomes flanking RE1 sites in iPS cells correlates with sub-nucleosomal chromatin particle binding (REST).

The presence of organised nucleosomes flanking RE1 sites in iPS cells correlates with the presence of a high frequency of chromatin particles in the sub-nucleosomal chromatin particle size class (REST) positioned at the RE1 site. Approximately half (n=411) of the RE1 sites possess flanking positioned nucleosomes in iPS cells, shown in the top half of the combined heat map (A), with x-axis =distance relative to the RE1 site (+/- 300 bp) and y-axis =log2 of the locally-normalized dyad frequency values for every bin position. Average cumulative frequency distributions of the chromatin particle distribution in the nucleosomal (B), sub-nucleosomal (C) and super-nucleosomal particle (D) size-classes at and surrounding (+/- 600 bp) this subset ("Structured") of RE1 sites was plotted with x-axis =distance relative to site (bp), y axis = mean mid-point frequency. Similar average cumulative frequency distributions were plotted for the remaining RE1 sites ("Less structured", n=460), shown in E, F and G. This group of sites had reduced nucleosome positioning surrounding RE1 sites that correlated with a reduction in the frequency of chromatin particles in the sub-nucleosomal chromatin particle size class (REST) positioned at the RE1 site

5.6 Strongly positioned nucleosomes can remain positioned at RE1 sites during neural cell development.

In order to investigate the changes in chromatin structure that occur at RE1 sites during differentiation to NPC cells, the subset of "structured" RE1 sites (Fig 5.6 Class 1 sites) with a high frequency of positioned sub-nucleosomal particles, was investigated in NPC cells. Average cumulative frequency distributions were plotted at and surrounding the class 1 RE1 sites in NPC cells for all three chromatin particle size classes (Fig 5.6 D, E, F). These results suggest that class1 RE1 sites tend to retain positioned nucleosomes at RE1 sites during differentiation to NPC cells, since well-positioned nucleosomes and super-nucleosomes flank the RE1 site (Fig 5.6 D and F), despite the absence of the sub-nucleosomal chromatin particle at the RE1 site in NPC cells (Fig 5.6 E).

These results suggest that the sub-nucleosomal MNase-resistant particle that is positioned at RE1 sites in iPS cells represents REST or the REST complex and that chromatin is organised around RE1 sites when REST or the REST complex is bound in iPS cells. When iPS cells differentiate to NPC cells, the class1 RE1 sites that have positioned nucleosomes in iPS cells retain this chromatin organisation, despite the fact that the positioned sub-nucleosomal particle at the RE1 site is not detected in NPC cells. It has been shown in mouse cells that REST is still present after differentiation to NPC cells but at a much lower level than that in ES cells (Ballas, Grunseich et al. 2005). If there were similar reduction in REST levels in human cells after differentiation to NPC, this might result in a reduction in occupancy at RE1 sites by REST or the REST complex. Nevertheless, the occupancy at RE1 sites in NPC cells has clearly changed without a change in the flanking chromatin structure.

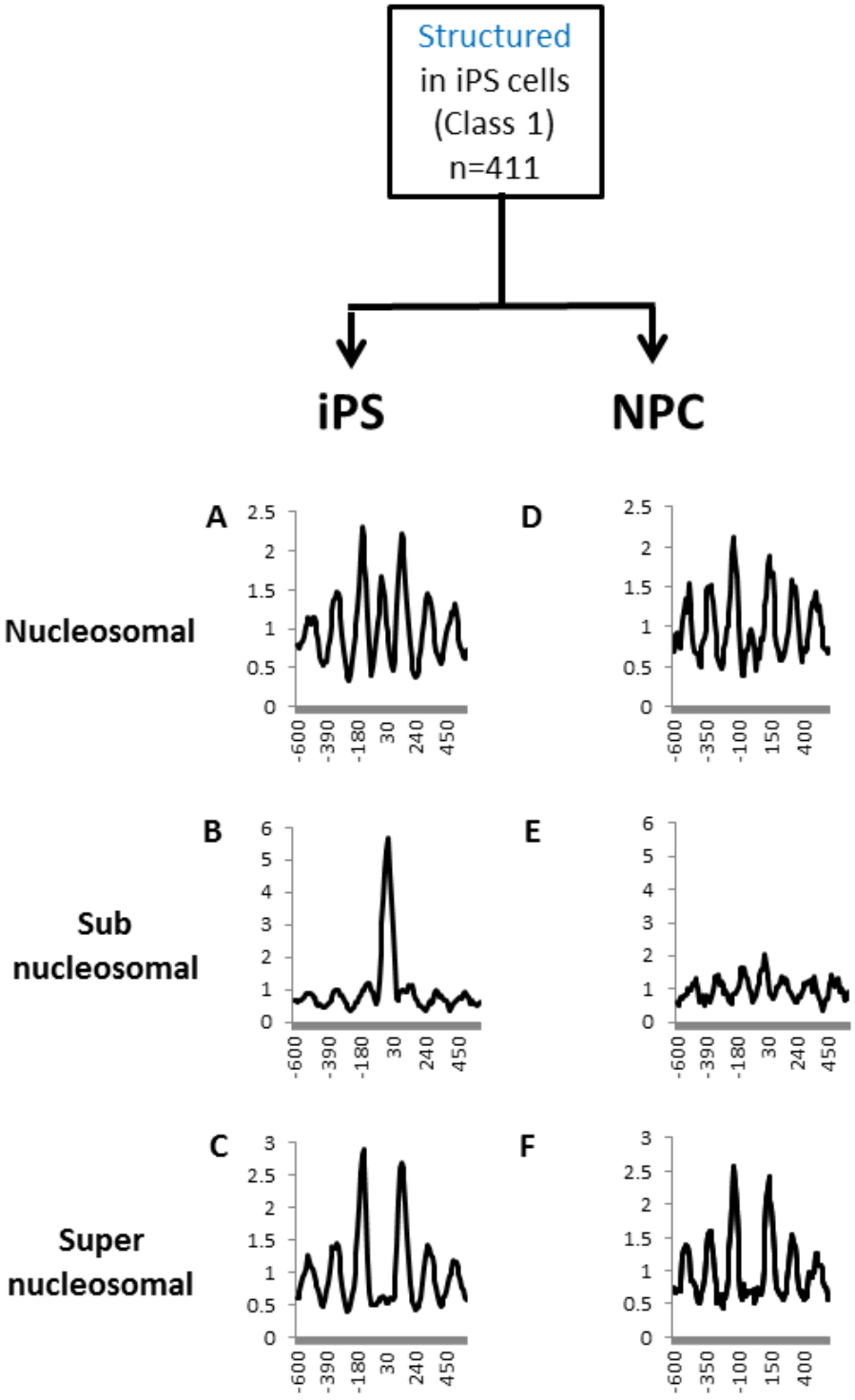


Fig 5.6 See next page for legend.

Figure 5.6 Organised chromatin is maintained surrounding RE1 sites during differentiation from iPS to NPC cells.

Average cumulative frequency distributions were plotted at and surrounding (+/- 600 bp) the "structured, class 1" subset of RE1 sites (n=411) with x-axis =distance relative to the RE1 site (bp), y axis = mean mid-point frequency in iPS and NPC cells for the nucleosomal (A and D), sub-nucleosomal (B and E) and super-nucleosomal particle (C and F) size-classes. The subset of RE1 sites that had a high frequency of chromatin particles in the sub-nucleosomal chromatin particle size class (REST) at the RE1 site in iPS cells (B) maintain strongly positioned nucleosomes flanking these sites in NPC cells (D and F) despite the loss of chromatin particles in the sub-nucleosomal chromatin particle size class (REST) at the RE1 site in NPC cells (E).

The subset of RE1 sites that had less organised flanking chromatin also had a reduced frequency of chromatin particles in the sub-nucleosomal chromatin particle size class at the RE1 site (Fig 5.5 E and F) in iPS cells. This subset of RE1 sites, henceforth referred to Class 2 RE1 sites, was further investigated to determine the chromatin structure at and surrounding these sites in NPC. Fig 5.7 shows average cumulative frequency distributions plotted at and surrounding the Class 2 subset of RE1 sites in NPC cells for the nucleosomal (D), sub-nucleosomal (E) and super-nucleosomal particle (F) size-classes. This group of RE1 sites had reduced nucleosome positioning surrounding RE1 sites in iPS cells. This pattern correlated with a reduction in the frequency of chromatin particles in the sub-nucleosomal chromatin particle size class (REST) positioned at the RE1 site in iPS cells. This less organised chromatin structure is maintained during differentiation to NPC cells, this is shown by comparing Fig 5.7 A with D and C with F. Therefore, at this class of RE1 sites, there is some degree of nucleosome positioning and this chromatin structure does not change during differentiation from iPS to NPC.

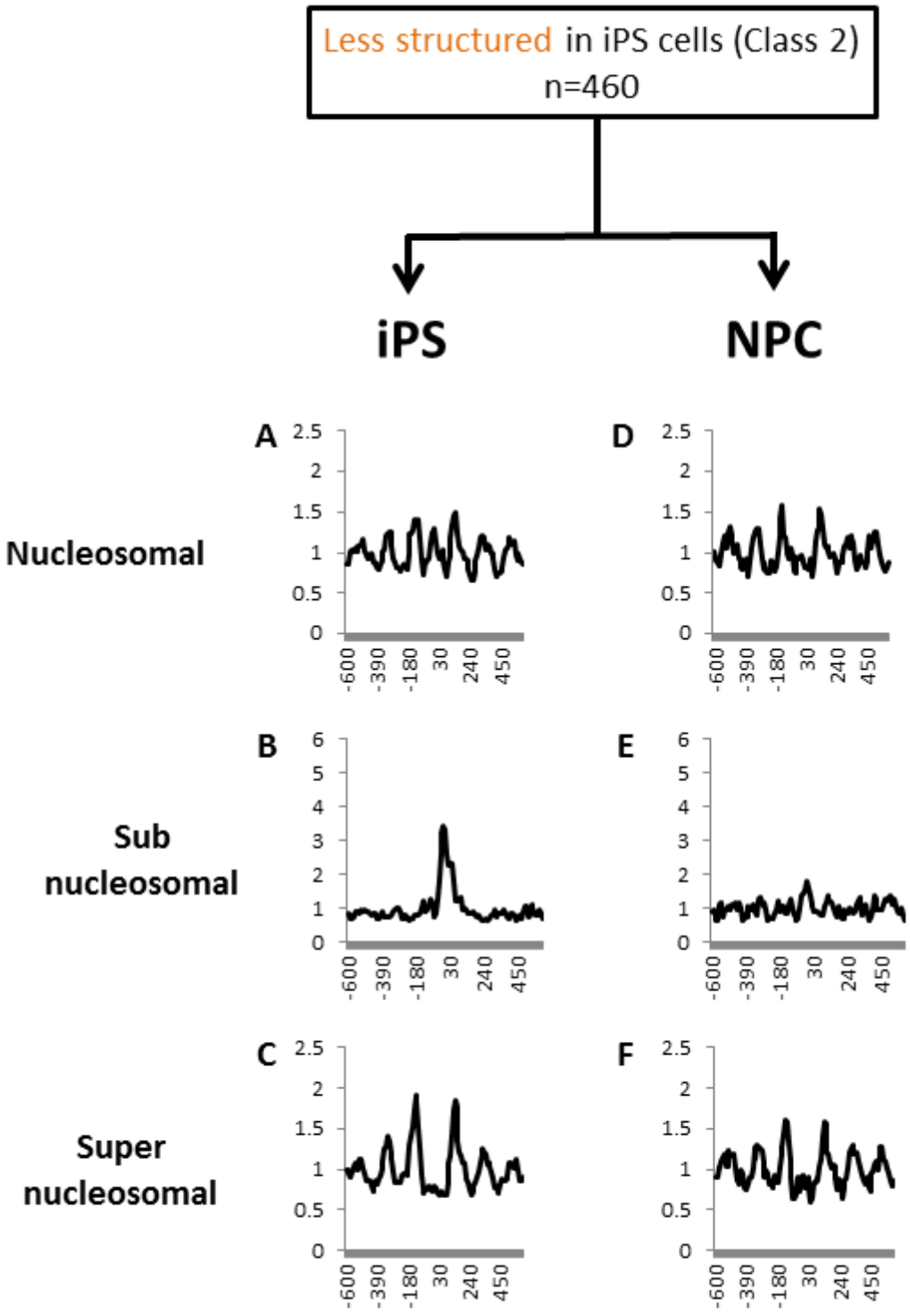


Fig 5.7 See next page for legend.

Figure 5.7 In the absence of REST, dis-organised chromatin is maintained at RE1 sites during differentiation from iPS to NPC cells.

The subset of RE1 sites ("Less structured, class 2") that had a reduced frequency of chromatin particles in the sub-nucleosomal chromatin particle size class (REST) at the RE1 site show little change in nucleosome organisation flanking these sites in NPC cells. Average cumulative frequency distributions were plotted at and surrounding (+/- 600 bp) this subset of RE1 sites (n=460) with x-axis =distance relative to the RE1 site (bp), y axis = mean mid-point frequency, in NPC cells for the nucleosomal (D), sub-nucleosomal (E) and super-nucleosomal particle (F) size-classes. This group of sites had reduced nucleosome positioning surrounding RE1 sites in iPS cells, this correlated with a reduction in the frequency of chromatin particles in the sub-nucleosomal chromatin particle size class (REST) positioned at the RE1 site in iPS cells. This less organised chromatin structure is maintained during differentiation to NPC cells. This is shown by comparing A with D and C with F.

5.7 Determination of the frequency of RE1 sites that possess positioned sub-nucleosomal particles.

The previous analysis in this chapter suggested that there was a correlation in iPS cells between the presence of positioned chromatin particles in the sub-nucleosomal size range, which might represent the REST complex and the presence of organised chromatin flanking the RE1 sites, in the size range of nucleosomes. The cumulative frequency distributions of paired-read midpoint positions at and surrounding a set of genomic feature positions represents an average chromatin structure surrounding the chosen genomic feature positions. Therefore, it is possible that the pattern of chromatin particle positioning observed in the sub-nucleosomal size class in Fig 5.2 was derived from a subset of RE1 sites. Some of the RE1 sites that had been shown to have organised nucleosomal chromatin flanking the RE1 sites might not have had a positioned sub-nucleosomal chromatin particle positioned at the RE1 site itself. Therefore, to derive a precise set of RE1 sites that possessed both a positioned sub-nucleosomal chromatin particle at the RE1 site and positioned nucleosomal chromatin particles flanking the RE1 sites, I decided to analyse the sub-nucleosomal chromatin particle data at RE1 sites in iPS cells further. Chromatin particle data representing the positions of chromatin particles in the sub-nucleosomal particle size class at and surrounding all 871 RE1 sites (+/- 300 bp) was clustered (k-means; Canberra distance) into eight groups (clusters C1-8). Each group is shown as a heat map (Fig 5.8). For each group of RE1 sites, the mean chromatin particle frequency values centred on and surrounding the RE1 sites were plotted for the sub-nucleosomal size class data and these are shown in the graphs on the right-hand side. The groups of RE1 sites that possess positioned sub-nucleosomal chromatin particles at RE1 sites were designated C1, C2, C5, C6 and C7, totalling 64% of the RE1 sites.

The final number and location of RE1 sites that possess both a positioned sub-nucleosomal chromatin particle and positioned nucleosomal chromatin particles flanking it was determined by obtaining the overlap in the locations of RE1 sites from each group. Thus, a final list of 309 RE1 sites was obtained for use in further analysis.

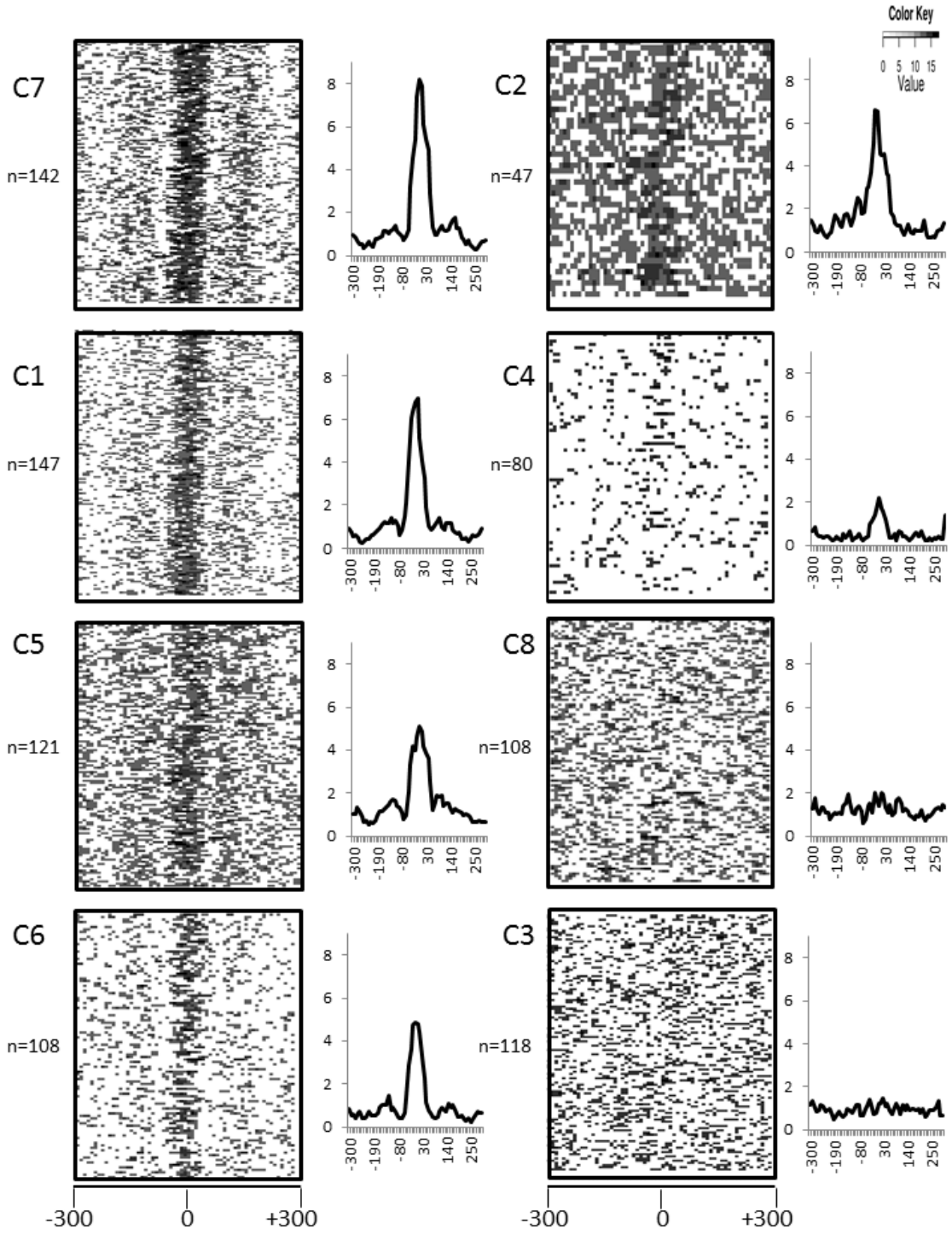


Fig 5.8 See next page for legend.

Figure 5.8 64% of the RE1 sites in iPS cells possess positioned sub-nucleosomal chromatin particles.

Data representing the positions of chromatin particles in the sub-nucleosomal particle size class at and surrounding all 871 RE1 sites (+/- 300 bp) were clustered (k-means; Canberra distance) into eight groups (C1-C8). Each group is shown as a heat map with x-axis =distance relative to the RE1 site (bp) and y-axis =log2 of the locally-normalized dyad frequency values for every bin position. For each group the mean chromatin particle frequency values centred on and surrounding the RE1 sites were plotted for the sub-nucleosomal size class data and these are shown in the graphs on the right-hand side in each group, x-axis =distance relative to site (bp), y axis = mean mid-point frequency. The groups of RE1 sites that possess positioned sub-nucleosomal chromatin particles at RE1 sites were C1 ,C2, C5, C6 and C7, totalling 64% of the RE1 sites.

5.8 Identification of genes targeted by REST during neural cell development.

To identify the genes whose RE1 sites possess positioned nucleosomes during neural cell development, the co-ordinates of the RE1 sites utilised in this study were matched to their respective genes using the Gencode annotation file: gencode.v18.annotation.gtf (Harrow, Frankish et al. 2012). Thus, a list of REST target genes for the RE1 sites used in this study was obtained. The subset (n= 309) of genes with RE1 sites that possessed both positioned sub-nucleosomal chromatin particles and flanking nucleosomal chromatin particles in iPS cells that was obtained section 5.7 was matched with a list of 64 known REST target genes, derived from the literature that had been validated experimentally (Appendix table A.2).This information was used to determine that 504 of the RE1 sites used in this study are within Gencode annotated genes and of these, 33 are known to be REST target genes. 18 of the known REST target genes used in this study have RE1 sites that possess strongly positioned sub-nucleosomal and nucleosomal chromatin particles in iPS cells and strongly positioned nucleosomal chromatin particles in NPC cells. The remaining 15 known REST target genes used in this study have weaker chromatin particle positioning.

Chromatin particle positioning was visualised by plotting histograms for all three chromatin particle size classes (sub-nucleosomal, nucleosomal and super-nucleosomal) in both iPS and NPC cells surrounding the RE1 site (+/-1 kb) in two of the known REST target genes, CACNA1A (Johnson, Gamblin et al. 2006) and BDNF (Schoenherr and Anderson 1995) (Fig 5.9.A and B respectively). CACNA1A encodes a voltage-gated Calcium channel subunit and it is regulated by REST in human cells. This gene has four binding sites for REST, only two of which bind REST with high affinity *in vivo* (Johnson, Gamblin et al. 2006). The RE1 site at chr19: 113,552,886 in intron 3 of the CACNA1A gene is shown in Fig 5.9.A. This is the position at which Johnson *et al* defined a high affinity REST binding site in HeLa cells (Johnson, Gamblin et al. 2006). This provides evidence that this site is a *bona fide* RE1 site. Hence this is an example of a previously characterised RE1 site that retains its nucleosome organisation during neural development.

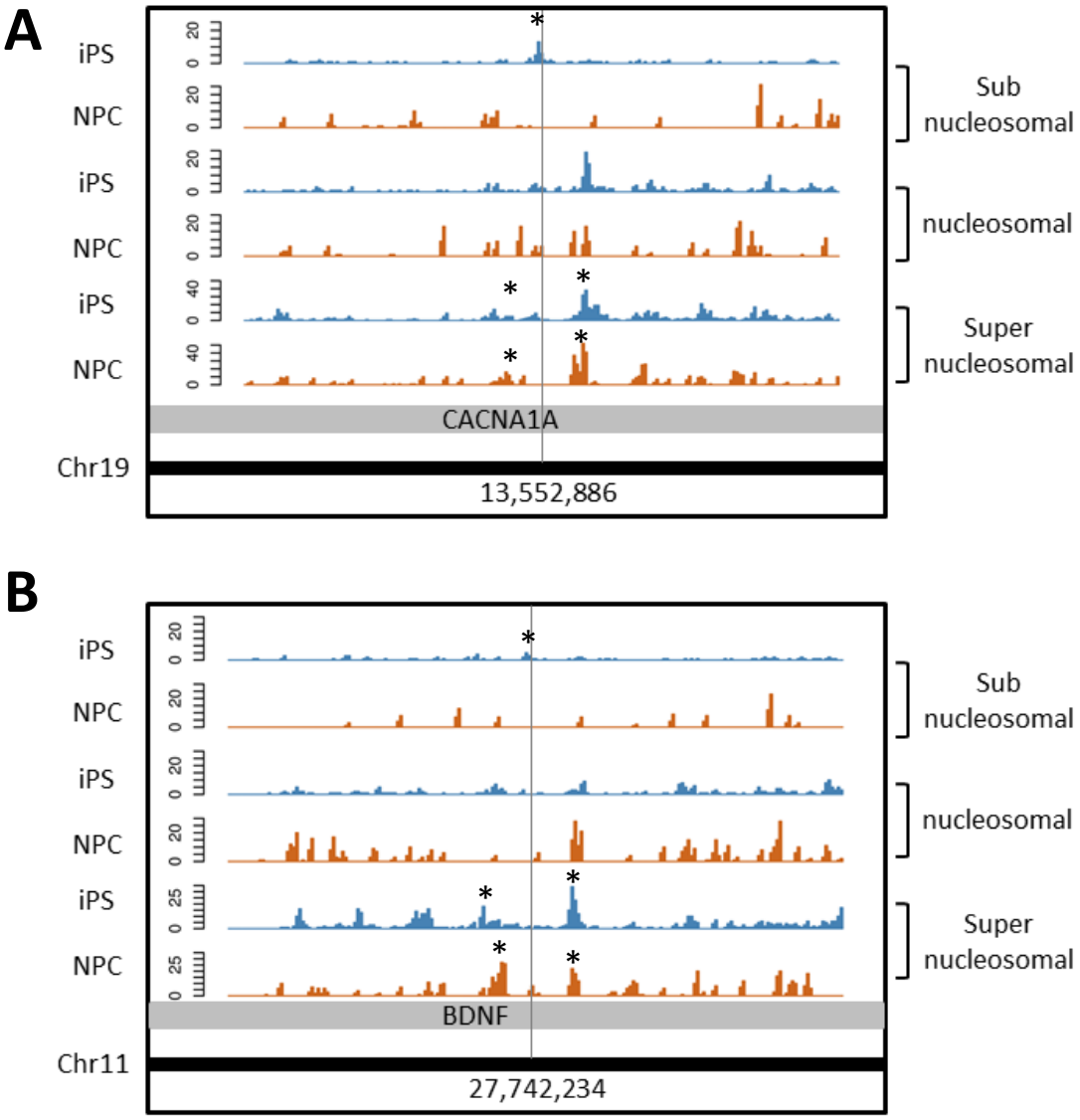


Figure 5.9 Visualisation of chromatin particle positioning at RE1 sites in iPS and NPC cells at REST target genes.

Paired-read midpoint position frequency data in 10 bp bins was extracted from .sgr files at and surrounding selected RE1 sites (+/- 1kb) that were within known REST target genes (See table A.2.) A. CACNA1A and B. BDNF. Histograms of paired-read midpoint position frequency data were plotted in R, x axis =chromosome co-ordinate (hg19), y axis = paired-read midpoint position frequency, for all three chromatin particle size classes (sub-nucleosomal, nucleosomal and super-nucleosomal) in both iPS and NPC cells. RE1 sites are shown by the grey vertical line. Sub-nucleosomal chromatin particles are positioned at RE1 sites in iPS cells (indicated by *), but not in NPC cells. Nucleosomal and super-nucleosomal particles (indicated by *), are positioned flanking RE1 sites in both iPS and NPC cells in both the BDNF and CACNA1A genes. These genes retain positioned nucleosomes flanking their RE1 sites during differentiation to NPC, despite the loss of the positioned sub-nucleosomal particles at the RE1 site (REST) in iPS cells.

The BDNF gene encodes brain-derived neurotrophic factor (BDNF) which is a member of the neurotrophin family of growth factors that are found in the brain (Barde, Edgar et al. 1982). This factor is important in the growth, development and maintenance of nerve cells. BDNF transcription is repressed by REST binding to the RE1 site in intron two of the BDNF gene, and this repression is inhibited by the wild type huntingtin protein (Zuccato, Tartari et al. 2003) which when mutated, causes Huntington's disease (THDCRG 1993). Thus BDNF is an important factor in neuronal gene regulation and the mis-regulation of its expression through the opposing effects of REST and huntingtin can have catastrophic effects.

Two genes that possess RE1 sites and are likely to be REST target genes in humans, but have not been as well studied in terms of their relationship to REST are CHD5 and KCNAB2. These genes have strongly positioned sub-nucleosomal and nucleosomal size chromatin particles at the RE1 sites in iPS cells. This chromatin structure is retained when the cells differentiate to NPC (Fig 5.10). The KCNAB2 gene encodes a subunit of a voltage gated potassium channel and it has been shown that deletion of this gene in mice causes defects in memory and in associative learning (Perkowski and Murphy 2011). Studies in human Huntington's disease cortex samples have shown that KCNAB2 is one of the REST target genes that has significantly higher REST occupancy and downregulated expression in disease samples (Zuccato, Belyaev et al. 2007). CHD5 is a chromatin re-modeller that is important in spermiogenesis (Li, Wu et al. 2014) but it is also preferentially expressed in the nervous system (Thompson, Gotoh et al. 2003) and it is required for terminal neuronal differentiation (Egan, Nyman et al. 2013). It is a strong candidate tumour suppressor gene that is deleted in neuroblastoma (Fujita, Igarashi et al. 2008). It has been shown that CHD5 has a significant role in the regulation of neuronal genes, including polycomb (Egan, Nyman et al. 2013) and it is expressed at high levels in adult mouse brain (Vestin and Mills 2013). Hence, CHD5 could be another 'master regulator' of neural genes though its chromatin remodelling activity. The relationship between CHD5 and REST is not clear, but the data presented here suggest that CHD5 might be regulated by REST and warrants further investigation, but this is beyond the scope of this thesis.

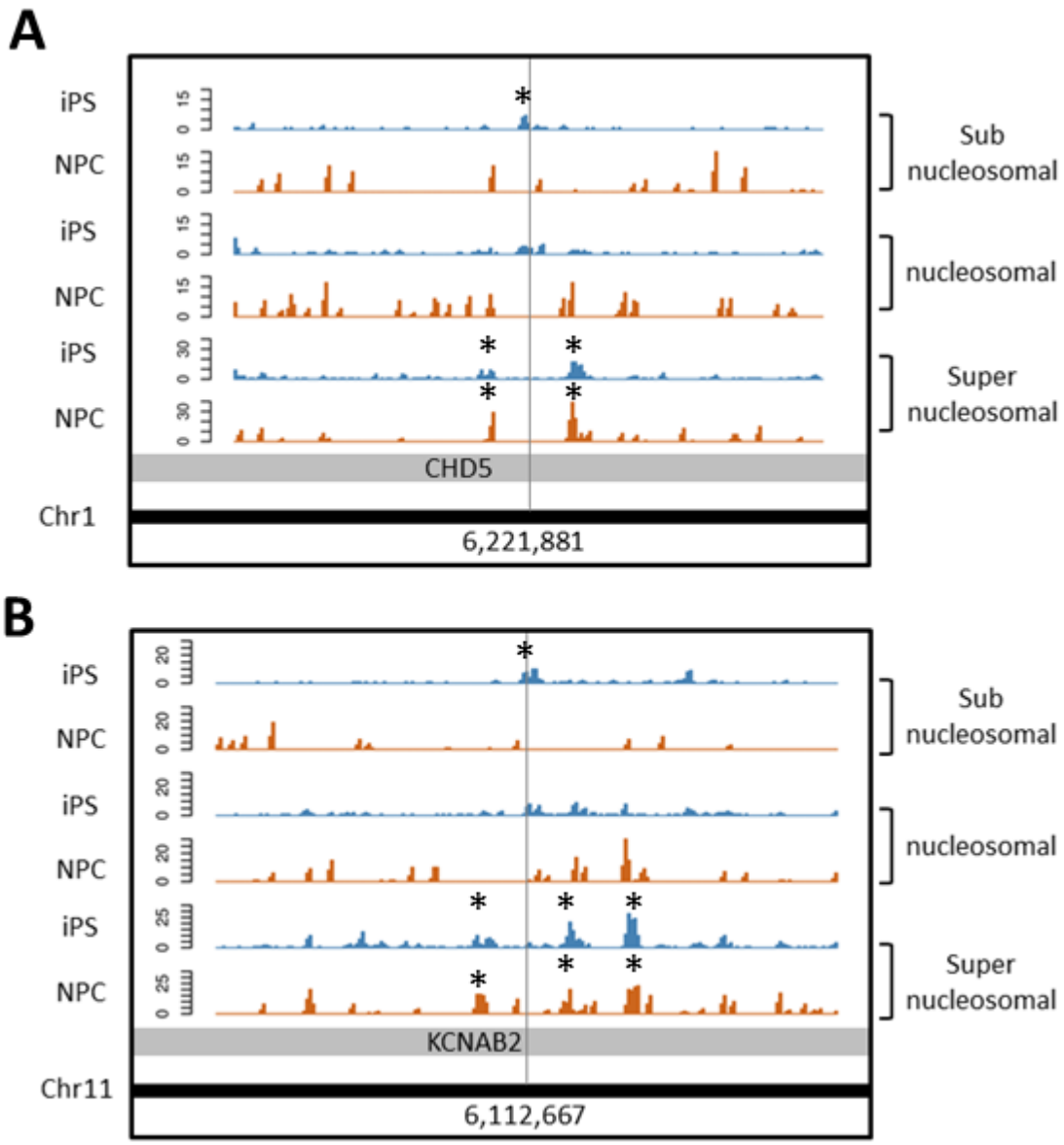


Figure 5.10 Visualisation of chromatin particle positioning at class one RE1 sites in iPS and NPC cells at putative novel REST target genes.

Paired-read midpoint position frequency data in 10 bp bins was extracted from .sgr files at and surrounding selected RE1 sites (+/- 1kb) that were within A. CHD5 and B. KCNAB2. Histograms of paired-read midpoint position frequency data were plotted in R, x axis =chromosome co-ordinate (hg19), y axis = paired-read midpoint position frequency, for all three chromatin particle size classes (sub-nucleosomal, nucleosomal and super-nucleosomal) in both iPS and NPC cells. RE1 sites are shown by the grey vertical line. Sub-nucleosomal chromatin particles are positioned at RE1 sites in iPS cells (indicated by *), but not in NPC cells. Nucleosomal and super-nucleosomal particles (indicated by *), are positioned flanking RE1 sites in both iPS and NPC cells in both the CHD5 and KCNAB2 genes. These genes retain positioned nucleosomes flanking their RE1 sites during differentiation to NPC, despite the loss of the positioned sub-nucleosomal particles at the RE1 site (REST) in iPS cells.

Chromatin particle positioning also was visualised at the RE1 site within the RLTPR gene (Matsuzaka, Okamoto et al. 2004). This gene is a representative example from the class of genes from this study whose RE1 sites do not possess positioned chromatin particles in iPS cells or NPC cells. This gene encodes a lymphoid cell-specific protein that is essential in T cell development (Liang, Cucchetti et al. 2013), thus it would not be expected to be involved in neural cell differentiation.

At some RE1 sites weaker nucleosome positioning was detected that correlates with weaker sub-nucleosomal chromatin particle positioning at the RE1 site (Class2 sites). The RE1 sites utilised in this study were derived from ChIP data from a mixture of cell types. Hence, it is possible that the class 2 RE1 sites characterised by their chromatin structure in this study, are not utilised in iPS cells. Alternatively, the characteristics and dynamics of RE1 sites may vary at different genes and at different developmental stages. In the mouse genome, it has been suggested that there are different levels of REST binding at different stages of development and at different genes (Sun, Greenway et al. 2005). If this were true in the human genome, it might be reflected in the chromatin dynamics at RE1 sites.

Taken together, the results in this chapter suggest that there is a specific set of RE1 sites in iPS cells that possess sub-nucleosomal chromatin particles positioned at the RE1 site. These RE1 sites tend to have positioned nucleosomal chromatin particles flanking them in iPS and NPC cells, suggesting that nucleosomes, once positioned at RE1 sites in pluripotent cells, remain positioned during early neural differentiation.

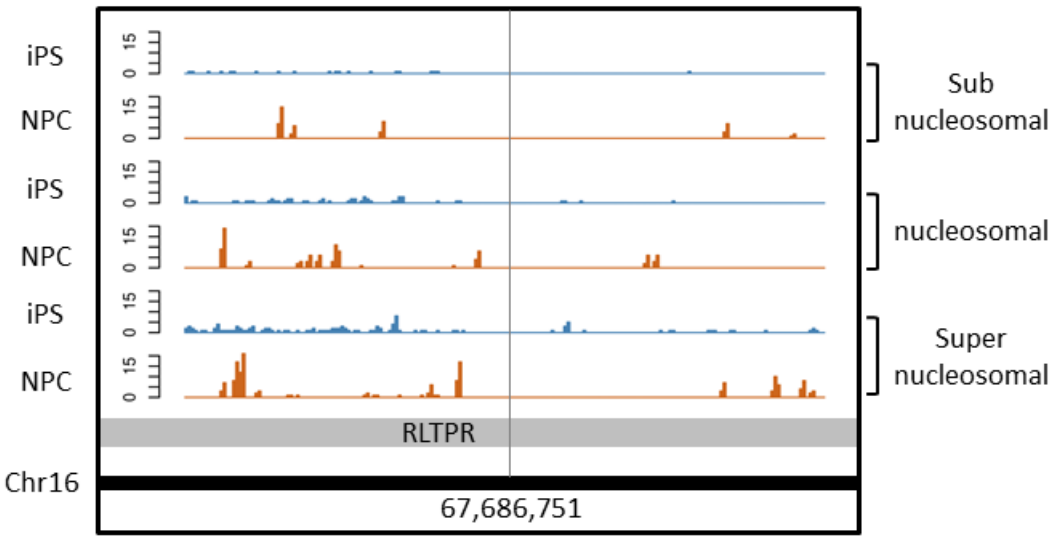


Figure 5.11 Some RE1 sites are devoid of any positioned chromatin particles

Visualisation of chromatin particle positioning at a class two RE1 site in iPS and NPC cells. . Paired-read midpoint position frequency data in 10 bp bins was extracted from .sgr files at and surrounding a class2 RE1 site (+/- 1kb) in the RLTPR gene. Histograms of paired-read midpoint position frequency data were plotted in R, x axis =chromosome co-ordinate (hg19), y axis = paired-read midpoint position frequency, for all three chromatin particle size classes (sub-nucleosomal, nucleosomal and super-nucleosomal) in both iPS and NPC cells. RE1 sites are shown by the grey vertical line. This illustrates that not all of the putative RE1 sites in this study possess positioned chromatin particles. This absence of chromatin organisation suggests that the RE1 site in the RLTPR gene may not be utilised as a transcription factor binding site in iPS and NPC cells.

5.9 Discussion

The work presented in this chapter has shown that it is possible to detect positioned nucleosomal chromatin particles flanking the binding motif of a master regulator of neural genes, in both iPS cells and NPC cells. This is in agreement with recent work in the human genome that has shown that regulatory sites tend to have high nucleosome occupancy in human cells (Tillo, Kaplan et al. 2010) and that well positioned nucleosomes are found flanking repressor sites (Nie, Cheng et al. 2014). Sub-nucleosomal chromatin particles were positioned at the RE1 site iPS cells, but they were not detected at the RE1 site in NPC cells. It has been shown in mouse ES cells that the levels of REST protein are high (Ballas, Grunseich et al. 2005). If this were true in human iPS cells then the data presented here suggests that the sub-nucleosomal chromatin particles positioned at the RE1 site in iPS cells detected here, using MNase-seq, might be REST or the REST complex. This could be tested using ChIP-seq.

In this study, a comparison of the nucleosomal chromatin particle positioning at RE1 sites in both iPS and NPC cells investigated whether there were changes in nucleosomal chromatin structure during neural cell differentiation. It was shown that nucleosomal chromatin particles remain positioned flanking RE1 sites during differentiation from iPS to NPC cells in approximately 50% of the RE1 sites investigated. The patterns of strongly positioned nucleosomes at RE1 sites in pluripotent cells correlate with the presence of sub-nucleosomal chromatin particles at the RE1 site. However, sub-nucleosomal chromatin particles were not detected at RE1 sites in NPC cells. In mouse, REST is still present after differentiation to NPC cells but at a much lower level than in ES cells (Ballas, Grunseich et al. 2005). It is possible that this is true in human cells, but that this low level of REST is not detected using the MNase-seq technology used in this study. However, the results presented here suggest that in human cells, well organised chromatin structure, once established, tends to be maintained at a subset of RE1 sites during the early stages of neuronal differentiation and that this is independent of the presence of the positioned sub-nucleosomal chromatin particles at the RE1 site. The findings in this chapter are summarised in Fig 5.12.

The results presented here show that the chromatin structure at RE1 sites in iPS cells is maintained in NPC cells, hence further epigenetic changes affecting REST binding or REST target gene regulation would be most likely through post-translational modifications of chromatin or DNA rather than through changes in the position of nucleosomes at this stage of development. The positioned nucleosomes might mark these RE1 sites and allow REST to rebind to some or all of the same binding motifs in NPC cells, or in later development to regulate a specific set of target genes. There is evidence that there are different types of RE1

sites, having different binding characteristics that can be utilised in different ways in development. Bruce *et al* 2009 (Bruce, Lopez-Contreras et al. 2009) showed that RE1 sites in human cells have a binding affinity hierarchy and that the degree of binding is sequence specific. Sites at which there is strong REST binding, with canonical binding site sequences, tend to be utilised in all cell types, whereas and sites at which there is weak REST binding, with divergent binding site sequences, are utilised in a cell type specific manner.

There is evidence that REST levels are high in cell types other than pluripotent cells. Although REST levels are low in NPC cells and cortical neurons, it is expressed at high levels in the adult hippocampal granule and pyramidal neurons (Palm, Belluardo et al. 1998). In addition, REST expression increases in rat brain after kainate induced seizures (Palm, Belluardo et al. 1998) and after global ischemic attacks. Calderone *et al* showed that increased REST expression after global ischemic attacks leads to the suppression of the GluR2 promoter, affecting AMPA receptor function and resulting in the death of hippocampal CA1 neurons (Calderone, Jover et al. 2003). Hence, although REST levels are low in NPC cells, REST levels can increase in later development or in response to cellular stress. Recent work by McClelland *et al* (McClelland, Dube et al. 2011; McClelland, Flynn et al. 2011) has shown that after seizures, REST levels increase, but this increase does not affect genes that bind REST with high or low affinity, rather, the increase REST levels seemed to affect a subset of genes with 'mid-range' affinity binding RE1 sites, suggesting that these sites are poised to respond to changes in REST levels. This subset of genes is critical for neuronal function.

Patterns of nucleosome positioning may define different types of REST sites, the degree of REST binding and whether and how REST sites are used at different times in development or in response to cellular stress.

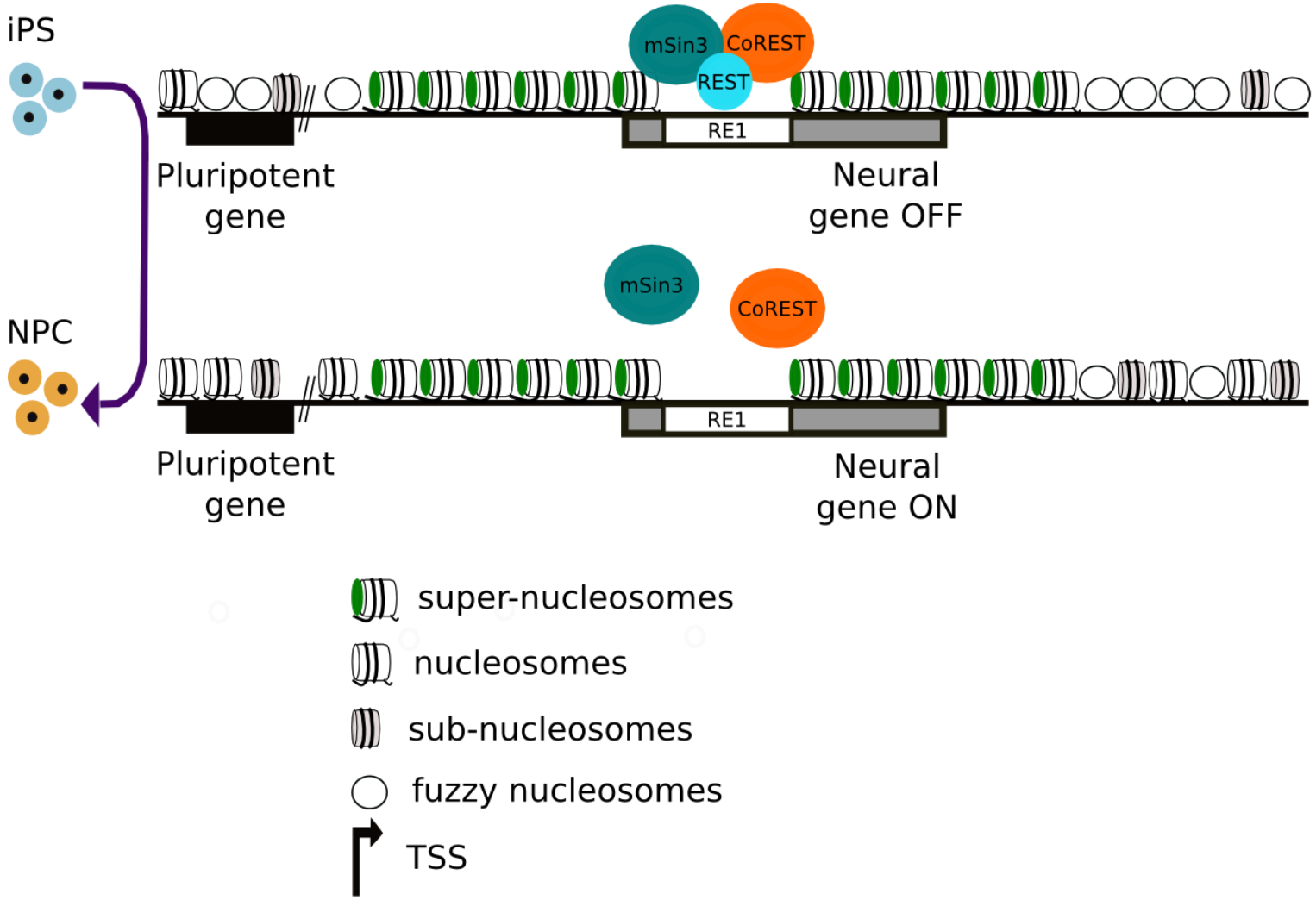


Fig 5.12 See next page for legend.

Figure 5.12 Summary of Chromatin structure at RE1 sites during early neuronal differentiation

A. In pluripotent cells, REST binds to the RE1 site and recruits co-repressor complexes. Nucleosomes are positioned flanking the RE1 site. **B.** In neural progenitor cells, REST is not bound to the RE1 site but nucleosomes remain positioned flanking the RE1 site.

Chapter 6: Chromatin at CTCF binding motifs during neural cell differentiation.

6.1 Introduction

CTCF has a dominant role in chromatin architecture but the relationship between CTCF binding, chromatin architecture and genome regulation is not clear. At the local level of chromatin structure, it is known that CTCF binding motifs have positioned nucleosomes flanking them (Fu, Sinha et al. 2008; Teif, Vainshtein et al. 2012). The aim of this chapter was to investigate whether changes occur in nucleosome positioning at CTCF binding motifs during early neural cell differentiation in human cells. This would generate information about the role of CTCF in early neural cell differentiation and provide insight into the mechanisms involved in changes in chromatin architecture that lead to changes in the regulation of cell type specific genes. The roles of CTCF were discussed in Chapter1, section 1.15.1 and they are summarised in Fig 6.1.

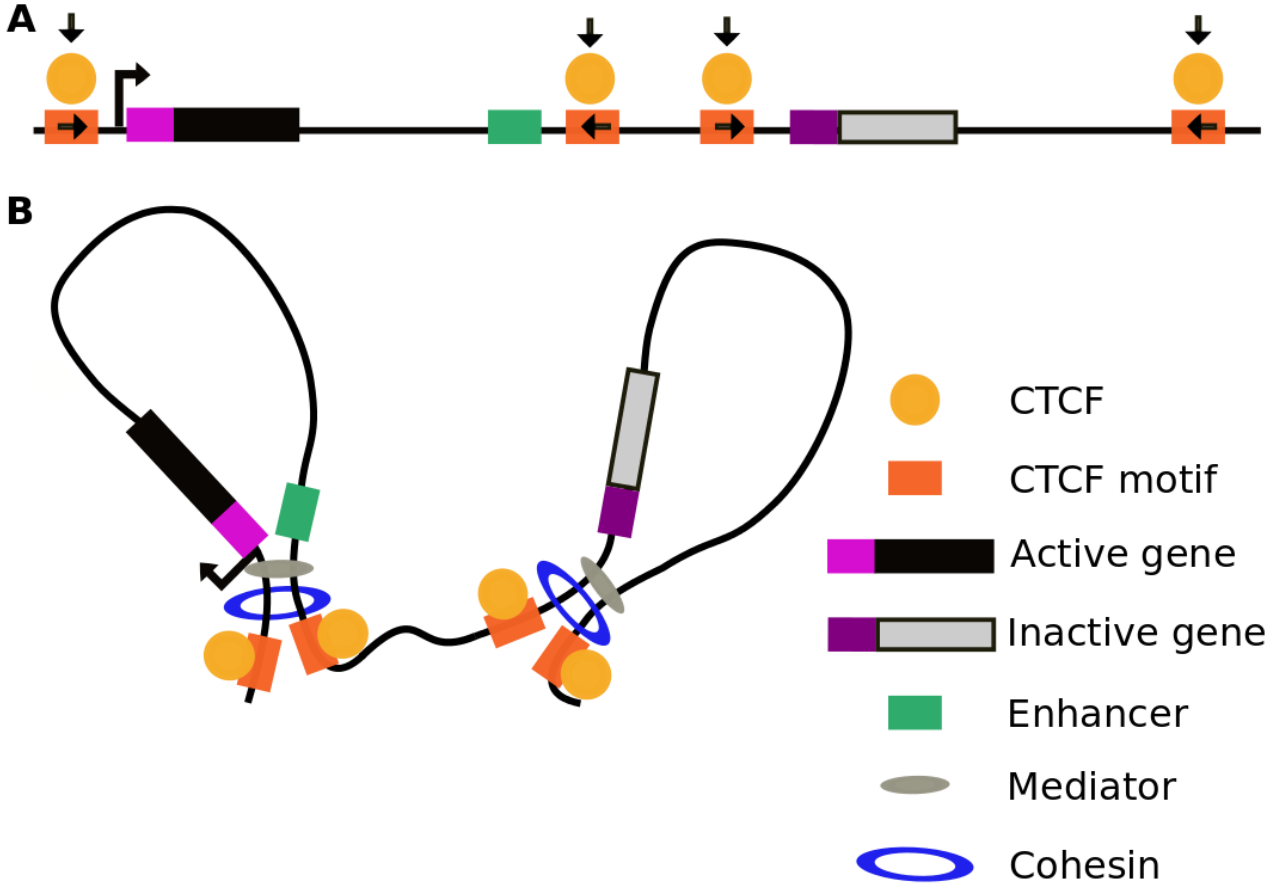


Fig 6.1 See next page for legend

Figure 6.1 The roles of the insulator protein CTCF

CTCF can act as a barrier between areas of active and inactive chromatin and it is involved in the formation of chromatin loops. A. Cartoon of a linear section of an active chromatin domain in the human genome flanked at its edges by CTCF sites which act as barriers to flanking heterochromatin. The cartoon depicts an actively transcribed gene, an inactive gene and an enhancer with interspersed CTCF binding motifs. CTCF binding motifs have directionality (shown by arrows in the motif), affecting the direction in which chromatin loops form. B. After CTCF has bound to its motif; cohesin and the co-activator Mediator are recruited, leading to the formation of chromatin loops or topologically active domains (TADs). Looping may bring together a gene promoter and an enhancer, resulting in an active gene, shown in the left-hand loop. Alternatively looping might block the interaction of a gene promoter with an enhancer, preventing gene activation, shown in the right hand loop.

6.2 Derivation of CTCF binding motifs.

In this this study, a human core CTCF consensus binding motif was mapped to ChIP-seq data to derive a set of CTCF binding motifs to be, as near as possible, specific to human pluripotent cells. The 12 bp human core consensus CTCF binding motif, [C][C][A|G][C|G][C|T][A][G][G|A][T|G][G][G][C|T], encompassing modules 2 and 3 of the total 52 bp motif was used (Ong and Corces 2014). This motif was derived in HeLa cells using ChIP-exo (chromatin immunoprecipitation combined with lambda exonuclease digestion). This 12 bp motif was shown to be utilised by half of the locations bound by CTCF in HeLa cells, with intermediate levels of occupancy (Rhee and Pugh 2011) and it is similar to the motif derived by Nakahashi *et al* (Nakahashi, Kwon et al. 2013). As publicly available CTCF ChIP data from iPS cells was not available, it was decided to use CTCF ChIP data derived from human embryonic stem cells (H1-hESC) from the ENCODE project (Gerstein, Kundaje et al. 2012). The extent to which the characteristics of iPS cells are the same as those of ESC cells has been a topic of debate for some time (Chin, Mason et al. 2009; Hu, Weick et al. 2010). However, recent studies have shown, in several cell lines tested, that iPS cells and ESC have similar patterns of gene expression and chromatin marks (Guenther, Frampton et al. 2010; Mallon, Chenoweth et al. 2013). Specifically, Mallon *et al* have studied gene expression in the male H1-hESC line and in both differentiated NPC cells and undifferentiated iPS cells derived from them. This work showed that the undifferentiated isogenic cell populations did not show significant differences in gene expression (Mallon, Hamilton et al. 2014).

In this study, the positions of the centre of the 12 bp core CTCF consensus binding motif described above was extracted from the hg19 human genome autosomal fasta files. The positions of these sites were mapped to the H1-hESC cell CTCF ChiP-seq data from the Broad institute (for methods see Chapter 2 section 2.5.2).This provided a set of precise co-ordinates at the centre of each consensus CTCF binding sequence at which it has been shown by ChiP-seq that CTCF binds in H1-hESC cells The derived set of 9,516 CTCF binding elements used in this thesis herein are referred to as CTCF sites. To study genome-wide, cell-type specific changes in patterns in chromatin structure surrounding CTCF sites during neural cell development, the average cumulative frequency distributions of all three chromatin particle size classes (sub-nucleosomal, nucleosomal and super-nucleosomal) used in this study were plotted at and surrounding the derived set of CTCF sites.

6.3 Detection of sub-nucleosomal MNase-protected chromatin particles at CTCF sites during differentiation.

The average cumulative frequency distribution of the sub-nucleosomal (112-137 bp) chromatin particle size class at and surrounding CTCF sites shows that there is a large peak in the chromatin particle mid-point frequency centred on the CTCF site in iPS cells (Fig 6.2 A). This suggests that, on average, protection from MNase digestion by sub-nucleosomal chromatin particles at this derived set of CTCF sites in iPS cells can be detected.

In NPC cells the peak in the sub-nucleosomal chromatin particle positioned at the CTCF binding site is reduced to approximately 30% of that found in iPS cells. This result could be explained by two scenarios: i) on average, protection from MNase digestion by sub-nucleosomal chromatin particles is reduced after differentiation from iPS to NPC cells. This suggests that CTCF binding is reduced at this set of CTCF sites in NPC cells or ii) CTCF occupancy is reduced across the CTCF sites in the genomes of the NPC cell population, i.e. CTCF does not bind at the same place in every cell in the population (Fig 6.3).

6.4 Nucleosomal and super-nucleosomal chromatin particles are positioned surrounding CTCF sites in both iPS and NPC cells.

Next, the average cumulative frequency distribution of the nucleosomal (138-168 bp) and super- nucleosomal (162-188 bp) chromatin particle size classes at and surrounding CTCF sites was determined. The peaks in the average cumulative frequency distribution suggest that in both chromatin particle classes, there is an array of seven positioned particles on each side of the CTCF site in both iPS and NPC cells (Fig 6.2B and C). In contrast, there are few positioned chromatin particles in these size classes at the CTCF site itself. This result suggests that the nucleosomal chromatin structure flanking CTCF sites is, on average, retained during differentiation from pluripotent cells to NPC cells. However, there is a peak representing a high frequency of positioned chromatin particles in the nucleosomal size class (138-161 bp) centred on the CTCF site in iPS cells. This peak is smaller than that shown at the CTCF site in the sub-nucleosomal (112-137 bp) chromatin particle size class. Hence, the dominant peak at the CTCF site, most likely representing CTCF in a complex with other proteins, is in the sub-nucleosomal size class.

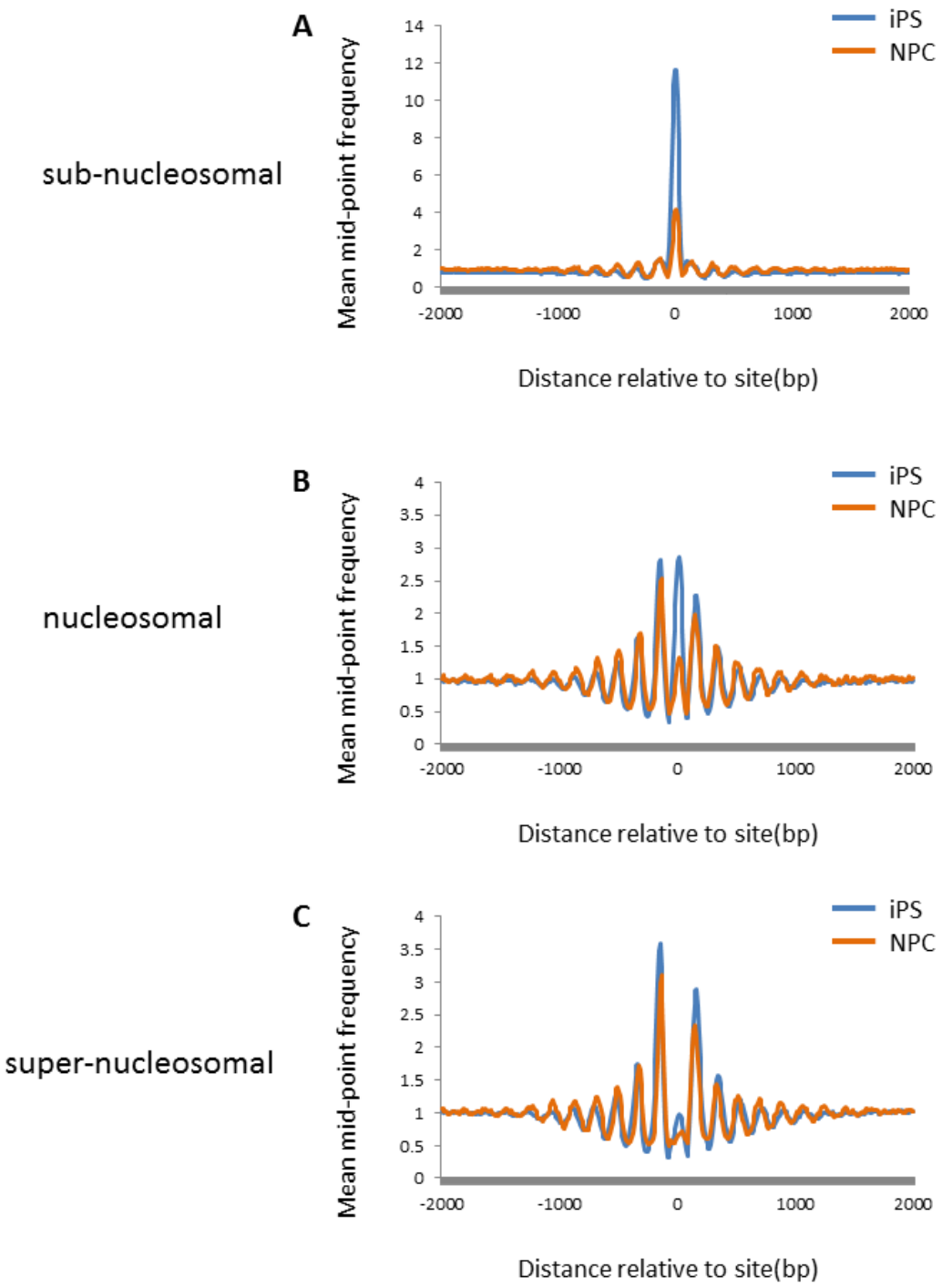


Figure 6.2 Chromatin remodelling occurs at CTCF sites during differentiation from pluripotent to neuronal precursor cells.

Normalised average cumulative frequency distributions of chromatin particle size classes A. 112-137 bp (sub-nucleosomes), B. 138-161 bp (representing nucleosomes) and C. 162-188 bp (super-nucleosomes) show the chromatin structure at and surrounding (+/-2000 bp) CTCF elements (n=9516) in both iPS and NPC cells.

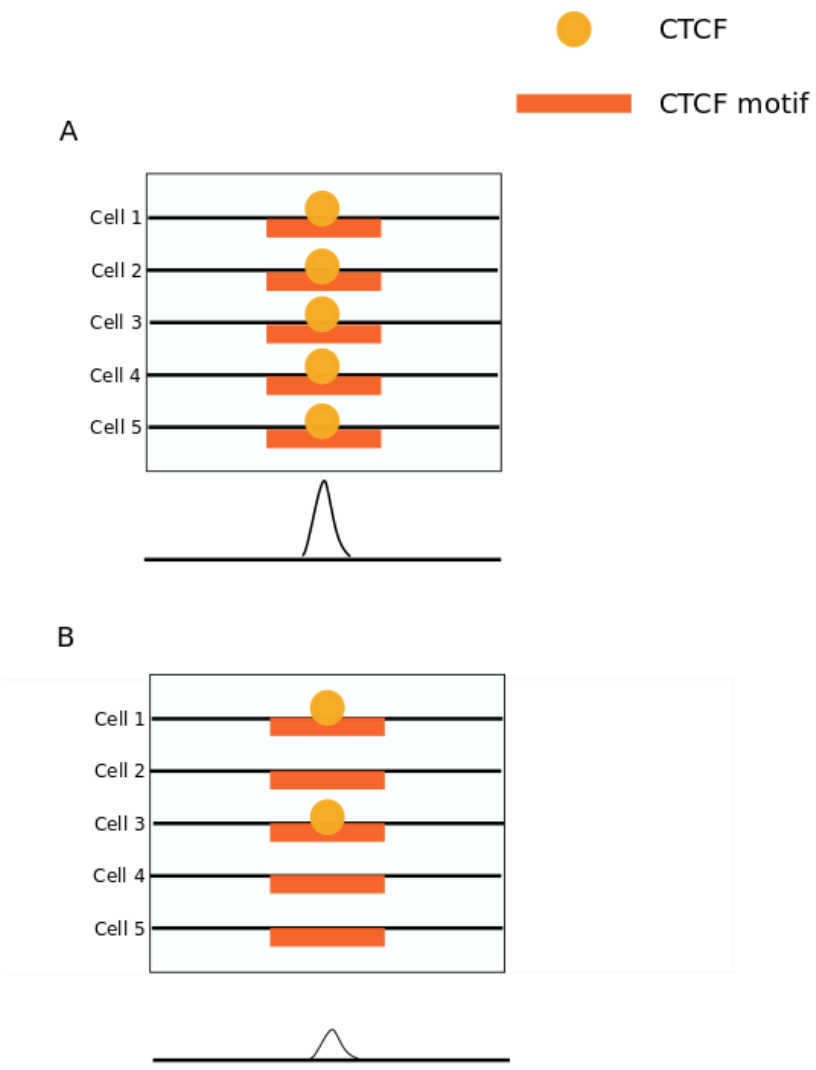


Figure 6.3 Model of CTCF occupancy.

A. In a population of cells, transcription factors that bind to their motif in the same place in the genome from one cell to the next in all cells have high occupancy. B. Transcription factors that do not bind to their binding motif in the same place in every cell have low occupancy.

In order to validate the chromatin structure surrounding CTCF sites shown in this study, the chromatin structure surrounding CTCF sites in both K562 and GM12878 cells was investigated using the nucleosome maps derived by Kundaje *et al* (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012) that were re-processed for comparison purposes as described in Chapter 3. Comparisons of the average cumulative frequency distributions in the nucleosomal (138-161 bp) chromatin particle size class in iPS versus K562 cells and iPS versus GM12878 cells are shown in Fig 6.4 A and C respectively. These results show that a similar number of positioned nucleosomes flank both sides of the CTCF site in all three genomes, but the peak that is centred on the CTCF site in iPS cells, is completely absent in K562 and GM12978 cells. This suggests that the peak that is centred on the CTCF site in iPS cells is specific to pluripotent cells. Comparisons of the average cumulative frequency distributions in the (138-161 bp) chromatin particle size class in NPC versus K562 cells and NPC versus GM12978 cells are shown in Fig 6.4 B and D respectively. Again, these results show that a similar number of positioned nucleosomes flank the CTCF site in all three genomes.

Taken together, these results confirm the finding from NPC cells shown here, that positioned nucleosomal chromatin particles flank CTCF sites in differentiated cells. In addition, these results suggest that positioned nucleosomal particles flank the CTCF binding motif even when sub-nucleosomal chromatin particle positioning at the CTCF site is reduced, as in NPC cells.

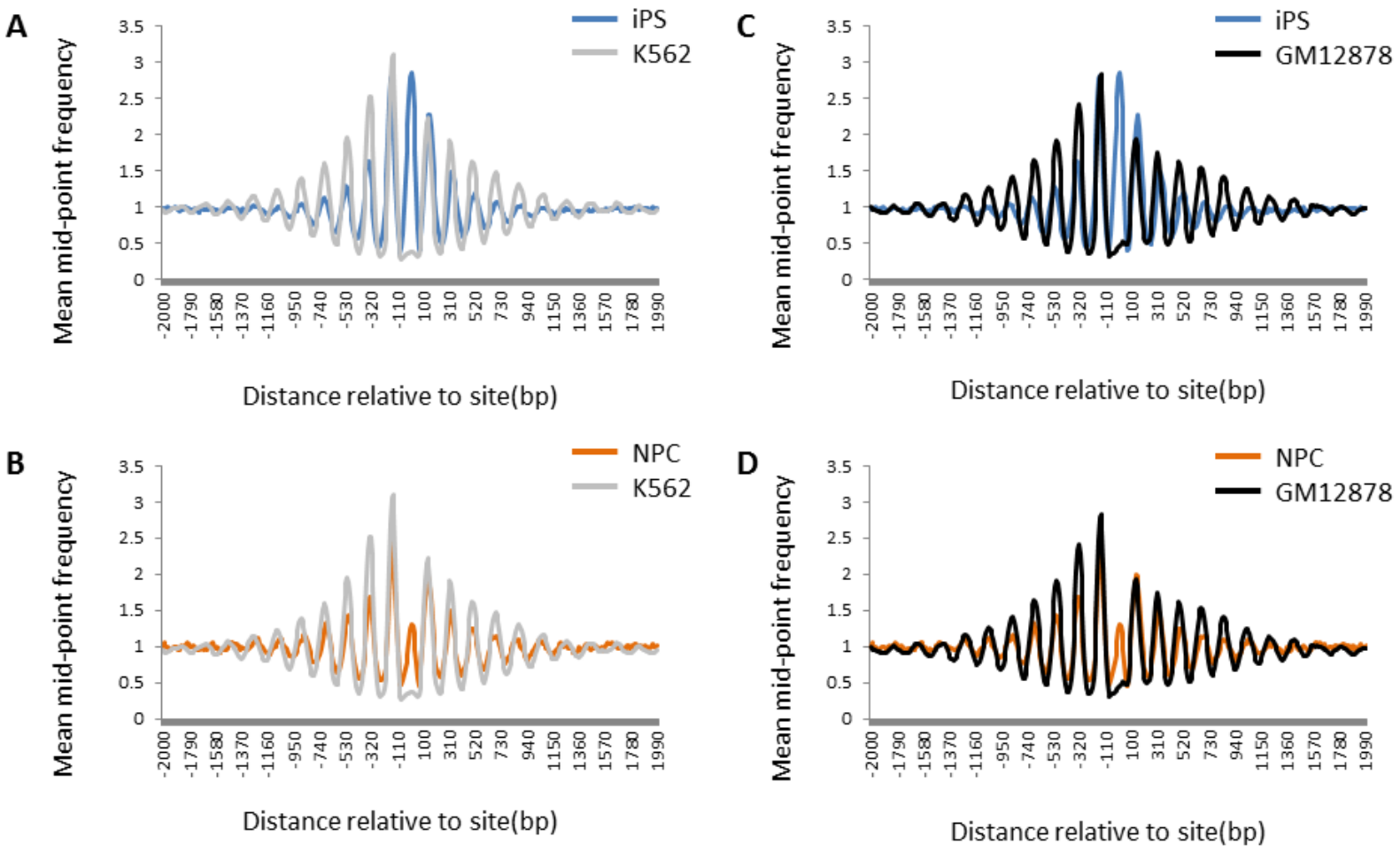


Fig 6.4 See next page for legend.

Figure 6.4 The pattern of positioned nucleosomes flanking CTCF sites is similar in differentiated cell types

Comparison of the positions of chromatin particles at and surrounding the CTCF site in the nucleosomal (138-161 bp) chromatin particle size range mapped in iPS and NPC cells by MNase-seq with those mapped in K562 and GM12878 cells. The data from the nucleosome maps generated for the K562 and GM12878 cell lines (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012) was converted from bedgraph to sgr format and the data was re-binned into 10 bp bins, calculating a 3bin moving average. The K562 and GM12878 maps were constructed by sequencing DNA purified from chromatin in the mono-nucleosome (approx. 150 bp) size range after MNase digestion. These data were compared with the iPS and NPC chromatin particle maps generated in the 138-161 bp size range since this is most similar to the size range of the chromatin particles in the K562 and GM12878 maps. A comparison of the average cumulative frequency distributions of the chromatin particles in different cell types at and surrounding (+/- 2000 bp) CTCF sites are shown A. iPS and K562, B. NPC and K562, C. iPS and GM12878, D. NPC and GM12878.

6.5 CTCF sites possess positioned sub-nucleosomal particles in iPS cells.

The cumulative frequency distributions of chromatin particles at and surrounding a set of genomic feature positions represented an average chromatin structure at and surrounding the chosen genomic feature positions. Hence, it was possible that the pattern of chromatin particle positioning observed in Fig 6.2 was derived from a subset of CTCF sites. So it was decided to investigate the patterns of positioned chromatin particles in the sub-nucleosomal chromatin particle size class in iPS cells further. This would determine whether sub-nucleosomal chromatin particles were positioned at all of the CTCF sites, or whether the pattern observed in Fig 6.2A was derived from a subset of CTCF sites. The chromatin particle data in the sub-nucleosomal size class surrounding CTCF sites within a window of +/-300 bp was clustered into six groups. For each cluster, the pattern of chromatin structure surrounding CTCF sites was plotted as a heat map and an average cumulative frequency distribution.

Fig 6.5 shows that sub-nucleosomal chromatin particles are positioned at all of the CTCF sites in iPS cells used in this study (n=9516), but the level of signal, indicating the level of protection from MNase digestion at CTCF sites by sub nucleosomal particles is variable. For example, cluster C5 has a signal intensity that is more than two fold greater than that in cluster C3. This suggests that all of the CTCF sites defined in this study are bound by CTCF and its co-binding proteins in iPS cells.

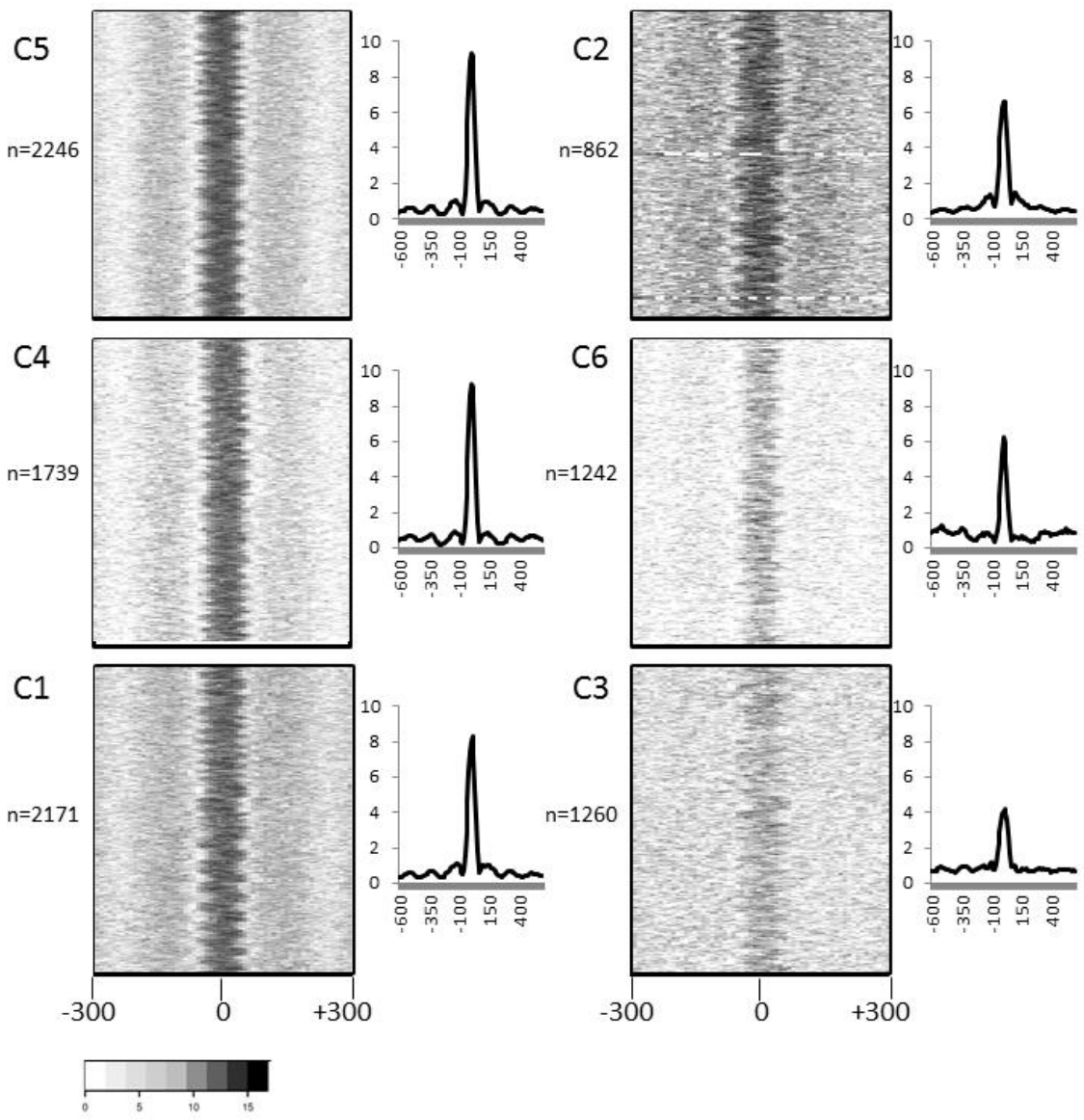


Figure 6.5 All of the CTCF sites possess positioned sub-nucleosomal particles in iPS cells

Data representing the positions of sub-nucleosomal chromatin particles at and surrounding all 9516 CTCF sites (+/- 300 bp) were clustered (k-means; Canberra distance) into six groups (clusters C1-6). Each group is shown as a heat map with x-axis =distance relative to the CTCF site (bp) and y-axis =log2 of the locally-normalized dyad frequency values for every bin position. For each group, the mean chromatin particle frequency values centred on and surrounding the CTCF sites was plotted for the sub-nucleosomal size class data and these are shown in the graphs on the right-hand side in each group: x axis = distance relative to CTCF site (bp), y axis = mean mid-point frequency. All of the groups of CTCF sites possess positioned sub-nucleosomal chromatin particles at the CTCF site.

6.6 Sub-nucleosomal particles are positioned at all CTCF sites after differentiation to NPC cells.

Next, the positioning of sub-nucleosomal particles at CTCF sites after differentiation of iPS to NPC cells was investigated. The clusters of CTCF sites generated in the analysis of sub-nucleosomal particle positioning at CTCF sites in iPS cells was assembled into a master list of sites in order of the signal intensity obtained (i.e. cluster C5>C4>C1>C2>C6>C3). This ordered list of all CTCF sites was used to extract the corresponding data, in the same order, representing the positions of sub-nucleosomal chromatin particles at and surrounding all 9516 CTCF sites (+/- 300 bp) in NPC cells. A heat map was constructed from these data (Fig 6.6). This analysis suggests that sub-nucleosomal particles are positioned in NPC cells at all of the CTCF sites tested in this study. This result is similar to the result found in iPS cells, however in NPC cells there is generally a much lower signal intensity at CTCF sites in NPC cells (compare Fig 6.6 A and B). This suggests weaker positioning of sub-nucleosomal particles at all of the CTCF sites in NPC cells. To investigate this further and to establish whether there might be a sub population of CTCF sites in NPC cells at which positioning of sub-nucleosomal particles is completely absent, the chromatin particle data in this size class surrounding CTCF sites within a window of +/-300 bp was clustered into six groups. For each cluster, the pattern of chromatin structure surrounding CTCF sites was plotted as a heat map and an average cumulative frequency distribution (Fig 6.7). This result suggests that, although it is possible to cluster these data into groups with slightly different patterns of chromatin particle positioning, sub-nucleosomal particles are weakly positioned at all CTCF sites in NPC cells.

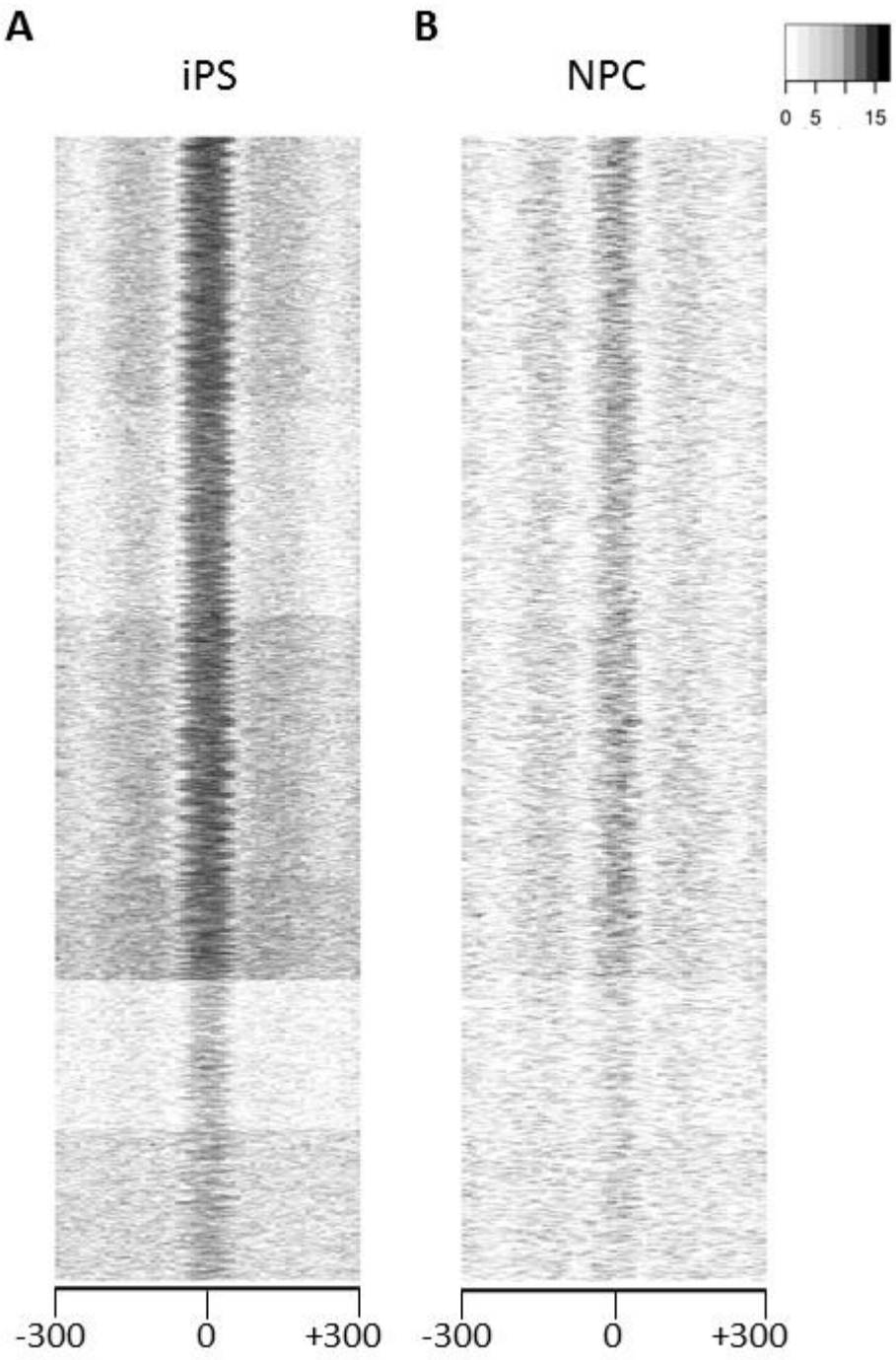


Figure 6.6 Sub-nucleosomal chromatin particles are positioned at all of the CTCF sites in NPC cells

The clusters of CTCF sites generated in the analysis of sub-nucleosomal particle positioning in iPS cells (shown in Fig 6.5) were assembled into an ordered master list of sites , according to the absolute particle position mid-point frequency (from high to low, i.e. clusters C5>C4>C1>C2>C6>C3). The corresponding data representing the positions of sub-nucleosomal chromatin particles at and surrounding all 9516 CTCF sites (+/- 300 bp) in iPS cells was extracted in the same order as the master site list. These data are represented as a heat map (A) with x-axis =distance relative to the CTCF site (bp) and y-axis =log2 of the locally-normalized dyad frequency values for every bin position. The corresponding data representing the positions of sub-nucleosomal chromatin particles at and surrounding all 9516 CTCF sites (+/- 300 bp) in NPC was extracted in the same way and it is represented as a heat map (B).

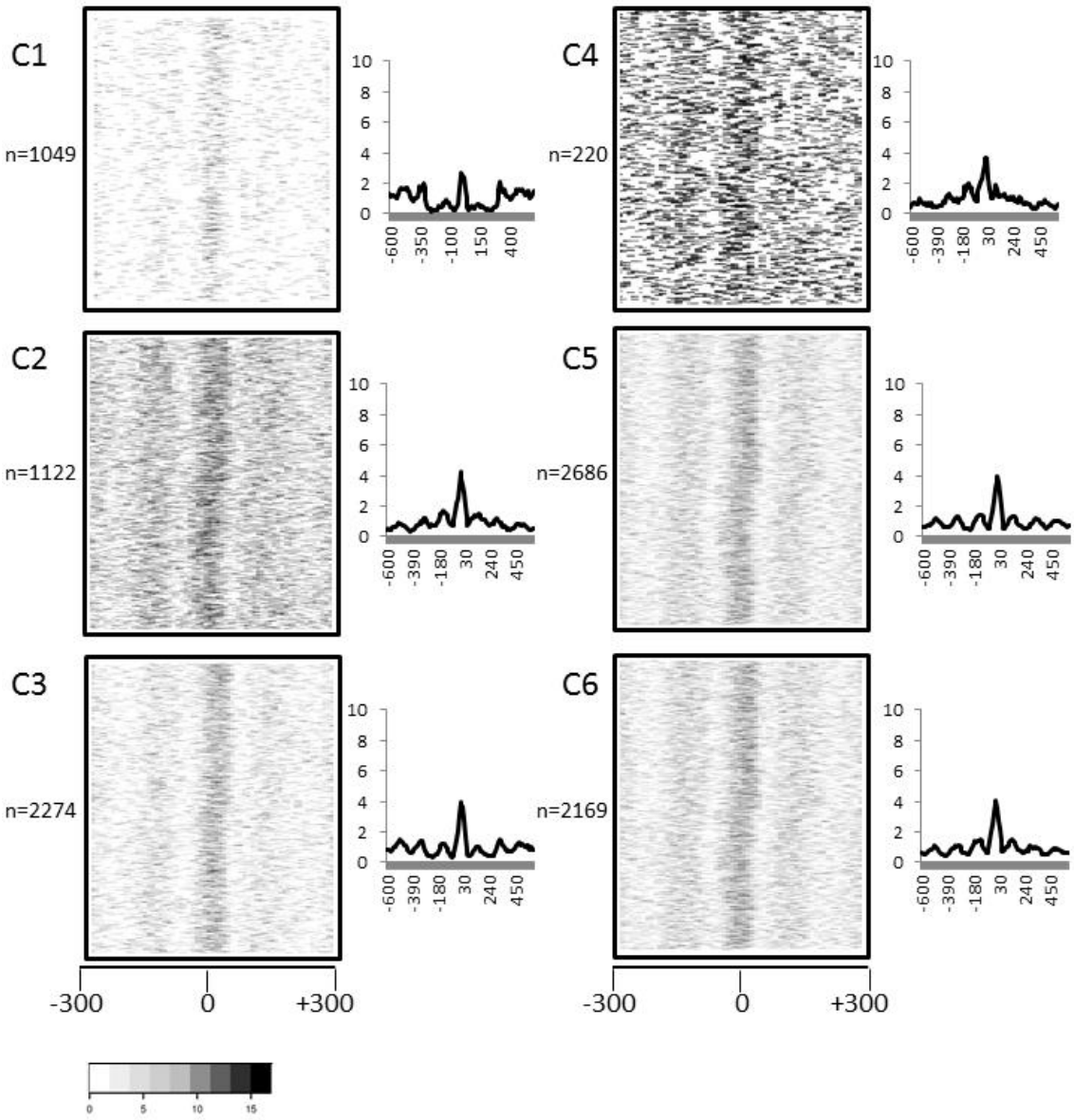


Figure 6.7 All of the CTCF sites possess positioned sub-nucleosomal particles in NPC cells

Data representing the frequency of mid-point positions of sub-nucleosomal chromatin particles at and surrounding all 9,516 CTCF sites (+/- 300 bp) were clustered (k-means; Canberra distance) into six groups (clusters C1-6). Each group is shown as a heat map with x-axis =distance relative to the CTCF site (bp) and y-axis =log2 of the locally-normalized dyad frequency values for every bin position. For each group, the mean chromatin particle frequency values centred on and surrounding the CTCF sites were plotted for the sub-nucleosomal size class data. These are shown in the graphs on the right-hand side in each group: x axis = distance relative to CTCF site (bp), y axis = mean mid-point frequency. All of the groups of CTCF sites possess positioned sub-nucleosomal chromatin particles at the CTCF site.

6.7 Nucleosomal chromatin particles flank CTCF sites during differentiation from iPS to NPC.

The positioning of nucleosomal particles surrounding CTCF sites in both iPS and NPC cells was investigated using the ordered list of CTCF sites generated according to the signal intensity obtained in the iPS sub-nucleosomal particle size class. The ordered list of CTCF sites was used to extract the corresponding chromatin particle midpoint position frequency data for the nucleosomal (168-161 bp) chromatin particle size class for both iPS and NPC. These data were plotted as heat maps, shown in Fig 6.8. This result shows that positioned nucleosomal chromatin particles can be detected flanking all CTCF sites in both iPS and NPC cells. Interestingly, the signal strength shown in the heat map for positioned nucleosomal chromatin particles in NPC cells is weaker than that in iPS cells. This suggests that nucleosomal chromatin particle positioning at CTCF sites is weaker in NPC cells than that in iPS cells.

6.8 Super-nucleosomal chromatin particles flank CTCF sites during differentiation from iPS to NPC.

The positioning of super-nucleosomal particles surrounding CTCF sites in both iPS and NPC cells was investigated using the ordered list of CTCF sites generated according to the signal intensity obtained in the iPS sub-nucleosomal particle size class. The ordered list of CTCF sites was used to extract the corresponding chromatin particle midpoint position frequency data for the super-nucleosomal chromatin particle size class in both iPS and NPC cells. These data were plotted as heat maps, shown in Fig 6.9. This result shows that positioned super-nucleosomal chromatin particles can be detected flanking all CTCF sites in both iPS and NPC cells. The signal strength shown in the heat map for positioned super-nucleosomal chromatin particles in NPC cells appears to be slightly weaker than that in iPS cells. This suggests that super-nucleosomal chromatin particle positioning at CTCF sites is weaker in NPC cells than that in iPS cells.

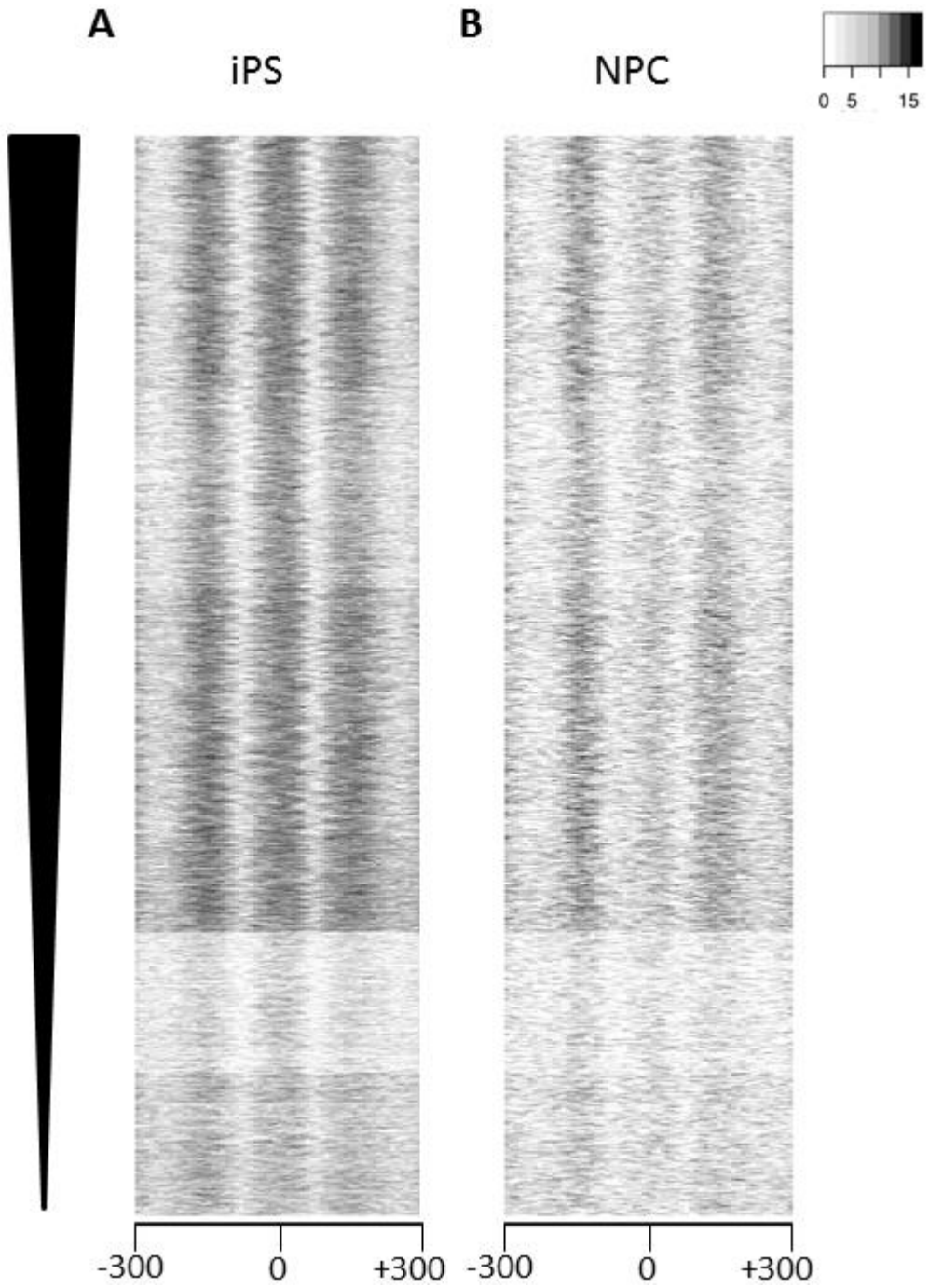


Figure 6.8 Nucleosomal chromatin particles remain positioned flanking CTCF sites during differentiation from iPS to NPC.

The clusters of CTCF sites generated in the analysis of sub-nucleosomal particle positioning in iPS cells (shown in Fig 6.5) was assembled into an ordered master list of sites, according to the absolute particle position mid-point frequency (from high to low, i.e. clusters C5>C4>C1>C2>C6>C3 represented by the black triangle). The corresponding chromatin particle midpoint position frequency data for the nucleosomal chromatin particles at and surrounding all 9,516 CTCF sites (+/- 300 bp) in iPS cells was extracted in the same order as the master site list. These data are represented as a heat map (A) with x-axis = distance relative to the CTCF site (bp) and y-axis =log2 of the locally-normalized dyad frequency values for every bin position. The corresponding data representing the positions of nucleosomal chromatin particles at and surrounding all 9516 CTCF sites (+/- 300 bp) in NPC was extracted in the same way and represented as a heat map (B).

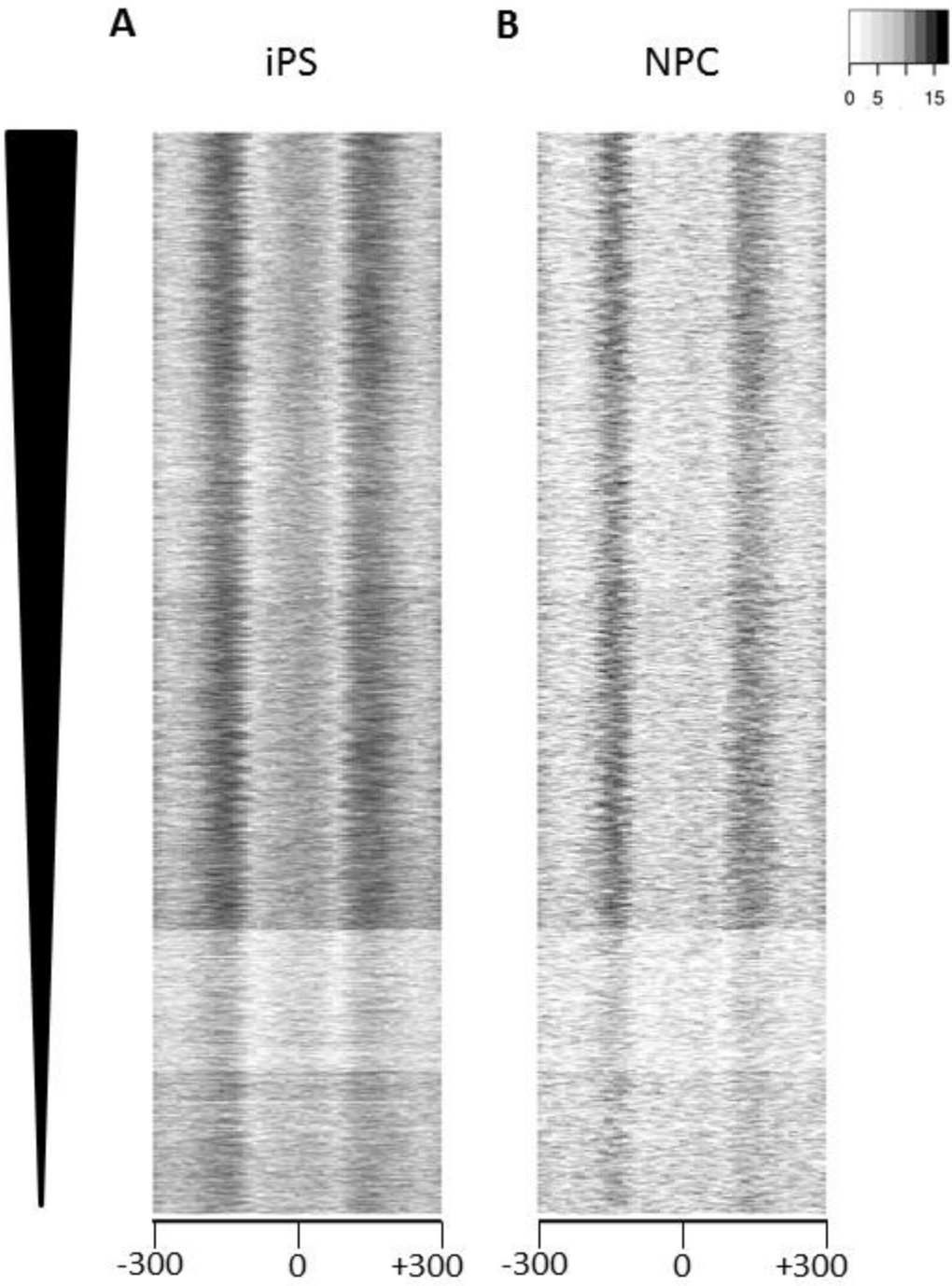


Figure 6.9 Super-nucleosomal chromatin particles remain positioned flanking CTCF sites during differentiation from iPS to NPC.

The clusters of CTCF sites generated in the analysis of sub-nucleosomal particle positioning in iPS cells (shown in Fig 6.5) was assembled into an ordered master list of sites , according to the absolute particle position mid-point frequency (from high to low, i.e. clusters C5>C4>C1>C2>C6>C3 represented by the black triangle). The corresponding chromatin particle midpoint position frequency data for the super-nucleosomal chromatin particle size class at and surrounding all 9,516 CTCF sites (+/- 300 bp) in iPS cells was extracted in the same order as the master site list. These data are represented as a heat map (A) with x-axis =distance relative to the CTCF site (bp) and y-axis =log2 of the locally-normalized dyad frequency values for every bin position. The corresponding chromatin particle midpoint position frequency data for the super-nucleosomal chromatin particle size class at and surrounding all 9,516 CTCF sites (+/- 300 bp) in NPC was extracted in the same way and represented as a heat map (B).

6.9 Super-nucleosomal chromatin particles are re-positioned flanking CTCF sites during differentiation from iPS to NPC.

The average cumulative frequency distributions of the chromatin particle size classes shown in figure 6.2 were examined further. Fig 6.10A shows an expanded version of the average cumulative frequency distributions in the super-nucleosomal chromatin particle size class surrounding CTCF sites in both iPS and NPC cells derived from figure 6.2C. Closer examination of the chromatin particle positioning data in the super-nucleosomal size classes revealed that the positioning of the nucleosomes flanking CTCF sites differs in iPS and NPC cells. Fig 6.10A shows that the super-nucleosomal chromatin particles both upstream and downstream of the CTCF site in NPC were positioned nearer to the CTCF site than the counterpart super-nucleosomal chromatin particle in iPS cells. The extent of the difference in positioning between iPS and NPC cells in this chromatin particle size class was investigated further. The positions of the peak maxima from the average cumulative frequency distributions shown in Fig 6.10A in both iPS and NPC cells were plotted as a bar chart. The bar chart shown in Fig 6.10B shows the positions of the peak maxima on the x axis relative to the CTCF site (x=0) and y was given an arbitrary constant value. Fig 6.10B illustrates that super-nucleosomal chromatin particles flanking CTCF sites are repositioned during differentiation from iPS to NPC. The first three super-nucleosomal chromatin particles both upstream and downstream of the CTCF site were positioned 10 bp nearer to the CTCF site than their counterpart iPS nucleosomes. The fourth flanking super-nucleosomal chromatin particle both upstream and downstream of the CTCF site was repositioned by approximately 20 bp in NPC cells relative to the position of the same super-nucleosomal chromatin particle in iPS cells.

These results suggest that re-positioning of nucleosomes flanking CTCF sites occurs during the early stages of neural cell differentiation. This re-positioning of nucleosomes flanking CTCF sites correlated with a reduction in the positioning of sub-nucleosomal particles at the CTCF site in NPC cells.

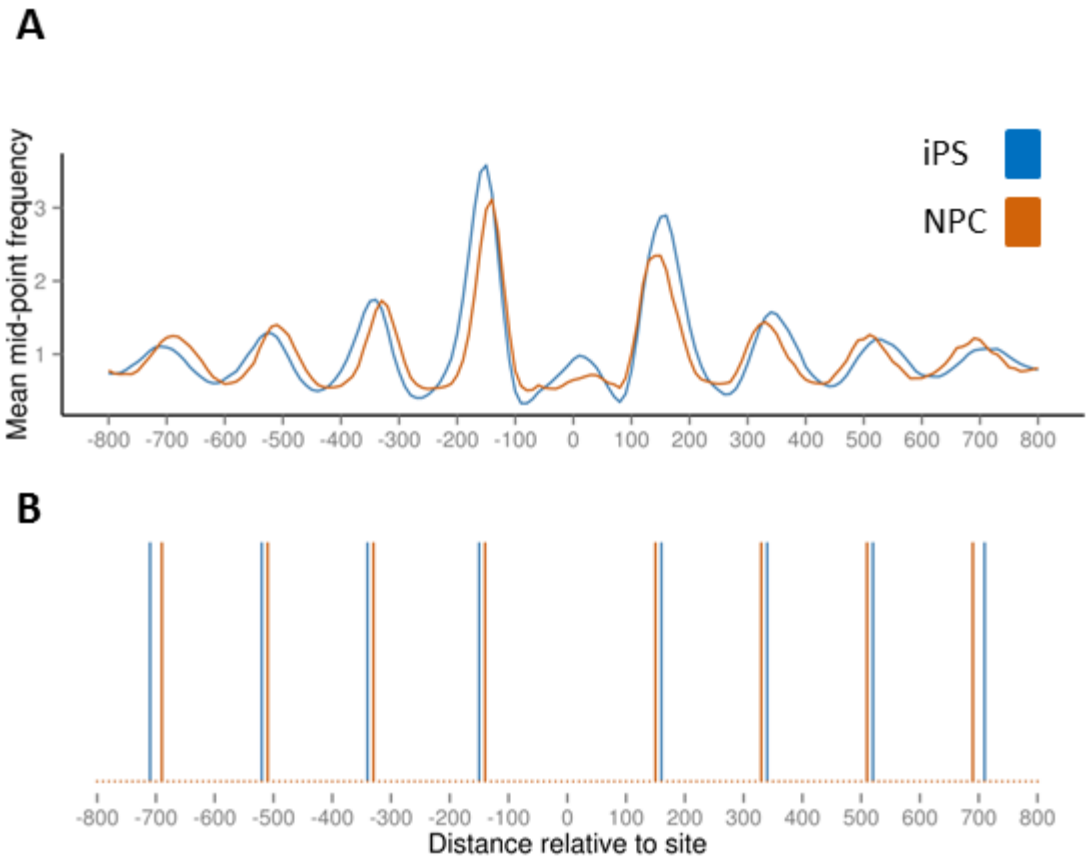


Figure 6.10 Super-nucleosomal chromatin particles flanking CTCF sites are repositioned, or remodelled, during differentiation from iPS to NPC

A. Normalised average cumulative frequency distributions of the super-nucleosomal (162-188 bp) chromatin particle size class show the chromatin structure at and surrounding (+/-2000 bp) CTCF sites in both iPS and NPC cells (n=9516). B. The positions of the peak maxima (x axis) relative to the CTCF site (x=0) were derived from each of the average cumulative frequency distributions in A. The positions of the peak maxima values were plotted for both iPS and NPC cells as a bar chart, with y as an arbitrary constant value. The first three nucleosomes both upstream and downstream of the CTCF site were positioned 10 bp nearer to the CTCF site than their counterpart iPS nucleosomes. The fourth flanking both upstream and downstream of the CTCF site was repositioned by approximately 20 bp in NPC cells relative to the position of the same nucleosome in iPS cells.

6.10 Discussion.

The work presented in this chapter has shown that nucleosomal chromatin particles are positioned flanking the CTCF binding motif in both human iPS and NPC cells. This is in agreement with previous studies of nucleosome positioning at CTCF sites in both the human lymphocytes (Fu, Sinha et al. 2008; Valouev, Johnson et al. 2011) and the mouse genomes (Teif, Vainshtein et al. 2012). In addition, a dominant peak in the sub-nucleosomal chromatin particle mid-point frequency was shown to occur at all of the CTCF sites in iPS cells. The peak in the chromatin particle mid-point frequency representing sub-nucleosomal chromatin particles positioned at the CTCF site also was present in NPC cells, but at only approximately 30% of the level seen in iPS cells. This suggested that there may be sub-groups of CTCF sites in NPC cells, some having strongly positioned sub-nucleosomal chromatin particles and others where sub-nucleosomal chromatin particle positioning does not occur. Cluster analysis showed that this was not the case; rather, all of the CTCF sites in NPC cells appeared to possess sub-nucleosomal chromatin particle positioning, albeit weak. This suggests that the change that occurs during differentiation from iPS to NPC cells is not simply that sub-nucleosomal particles are 'on' or 'off' a CTCF site. This suggests that CTCF does not bind to its binding motif in the same place in every cell in the NPC population; therefore CTCF sites have low occupancy in NPC cells (Fig 6.3).

The presence of the peak in the average cumulative frequency distribution of the sub-nucleosomal chromatin particles positioned at CTCF sites could be explained by direct protection of DNA from MNase cleavage by CTCF and other binding partners at CTCF sites in pluripotent cells. It has been shown in CD4+ cells, by DNase foot-printing analysis, that CTCF protects 50-72 bp of DNA (Fu, Sinha et al. 2008). It is known that CTCF binds a number of other proteins, for example, cohesin (Rubio, Reiss et al. 2008; Lee and Iyer 2012), TAF3 (Liu, Scannell et al. 2011), SIN3A (Lutz, Burke et al. 2000) and several other proteins such as YY1 and RNA polymerase II (Zlatanova and Caiafa 2009). Also, CTCF footprints can be cell type specific (Boyle, Song et al. 2011). As the CTCF sites used in this study were derived using ChIP data derived from H1-hESCs, it would be expected that these sites would be bound by CTCF and its binding partners in hiPS cells. However, there are other possibilities that cannot be excluded. For example, it is possible that the positioned sub-nucleosomal chromatin particle at CTCF sites represents a non-standard nucleosome. At its simplest this could be an 'unwrapped' nucleosome that is more susceptible to MNase cleavage (Li, Levitus et al. 2005; Henikoff, Belsky et al. 2011). Alternatively, sub-nucleosomal chromatin particles could be constructed from two nucleosomes that lose the H2A and H2B histones, then the remaining two histone $(H_3/H_4)_2$ dimers slide together to form an octamer. Weak protection from MNase at the ends

of the octamer would mean that 10 bp of DNA is trimmed from each end of the particle, leaving 125 bp of DNA (Read and Crane-Robinson 1985). More recently, precise nucleosome mapping in *S. cerevisiae* has suggested that half nucleosomes comprising one copy of each histone exist in dynamic regions of chromatin (Rhee, Bataille et al. 2014). The existence of any of these proposed half-nucleosome structures is possible at CTCF sites. However, it has been shown that nucleosomes flanking CTCF sites tend to be enriched for H2AZ (Fu, Sinha et al. 2008) and that human H2A.Z nucleosomes protect approximately 120 bp of DNA (Tolstorukov, Kharchenko et al. 2009). Hence, it is possible that the positioned sub-nucleosomal chromatin particle at the CTCF is an H2A.Z -containing nucleosome.

Arrays of approximately seven strongly positioned nucleosomes were detected flanking each side of all CTCF sites in both iPS and NPC, despite the weaker positioning of the sub-nucleosomal chromatin particle in NPC cells. However, it was shown here that nucleosomes are re-positioned flanking CTCF sites in human NPC cells. Nucleosomes both upstream and downstream of the CTCF site in NPC were positioned nearer to the CTCF site than the counterpart nucleosome in iPS cells. As a single turn of DNA is approximately 10.5 bp (Wang 1979), a 10 bp change in the position of the nucleosomes flanking CTCF sites in NPC cells would not affect the rotational position. It has been shown *in vitro* that ATP-dependent remodellers can move nucleosomes in 10 bp steps (Schwanbeck, Xiao et al. 2004). Therefore, this change in positioning of the nucleosomes flanking CTCF sites in NPC cells, relative to that seen in iPS cells, could be brought about by the action of the ATP-dependent remodellers that slide nucleosomes, such as the ISWI remodellers (Deindl, Hwang et al. 2013) or by CHD1 (Stockdale, Flaus et al. 2006), which is a member of the CHD remodellers. It has been shown in *S .cerevisiae* that the chromatin remodeller Isw2 moves nucleosomes into nucleosome free regions which might act to occlude transcription factor binding sites and repress transcription (Whitehouse and Tsukiyama 2006). One candidate remodeller in this case in human cells is the NURF complex from the ISWI group of remodellers as it has been shown to be enriched in the brain and is involved in bringing about neurite outgrowth (Barak, Lazzaro et al. 2003). Alternatively, positioning of the sub-nucleosomal particle at the CTCF site in iPS cells might act as a barrier, resulting in statistical positioning of the flanking nucleosomes (Mavrich, Ioshikhes et al. 2008). Alternatively, the re-positioning of nucleosomes could be due to a change in the size of the nucleosomes. This could occur through the binding of a small protein to one side of each nucleosome. Therefore the amount of DNA protected from MNase digestion would increase, resulting in a small change in the mid-point position of MNase-protection for these nucleosomes. However, it seems most likely that this re-positioning is brought about by the action of ATP-dependent chromatin re-modelling enzymes.

The work in this chapter is summarised in Fig 6.11. This work has shown that all of the CTCF sites in iPS cells possess positioned sub-nucleosomal particles and flanking nucleosomal particles. Broadly, this pattern of chromatin organisation is retained at all CTCF sites in NPC cells, but significantly, the nucleosomes flanking the CTCF site are re-positioned (Fig 6.10).This suggests that nucleosome positioning at CTCF sites is important during early neural cell differentiation, perhaps affecting higher order chromatin architecture to regulate neural specific genes.

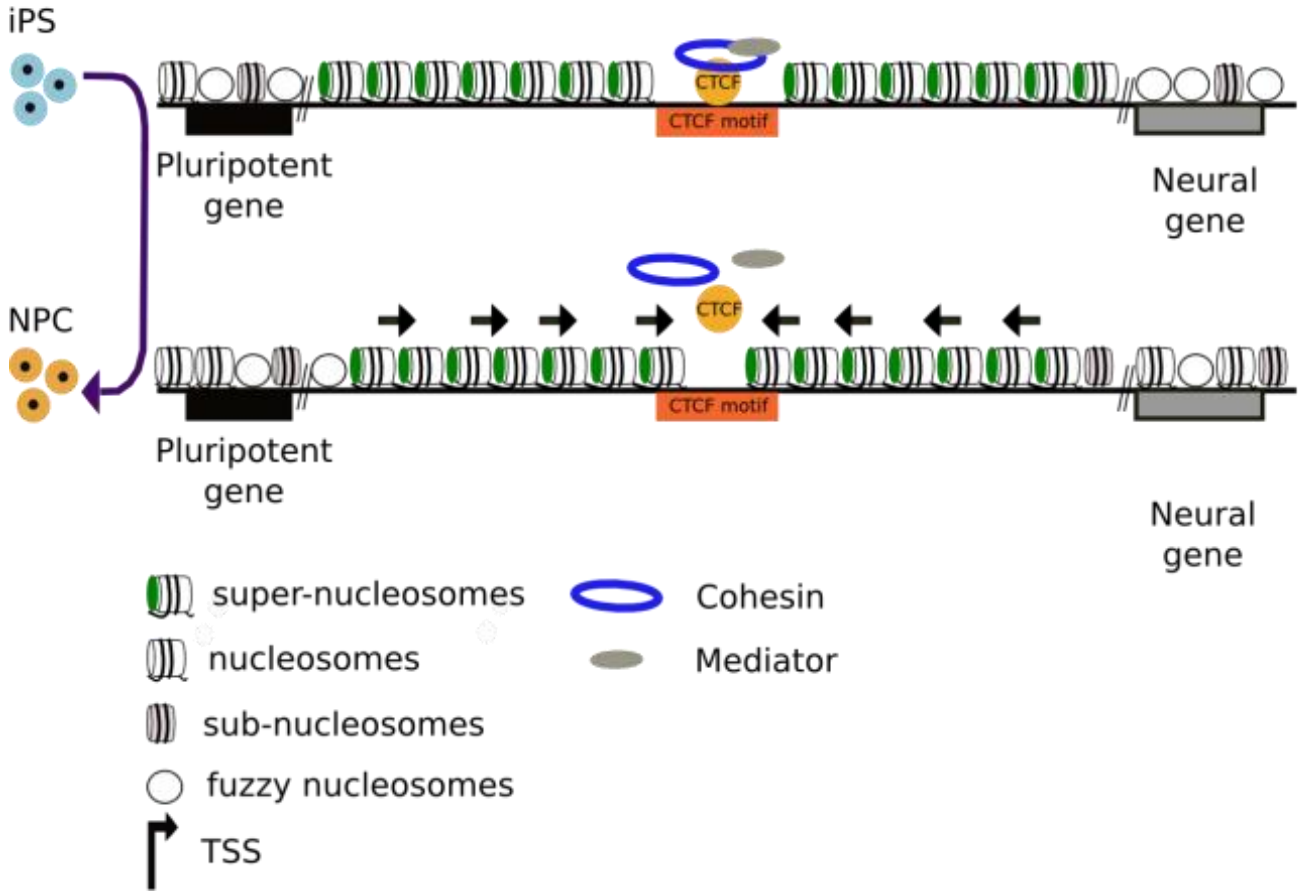


Fig 6.11 See next page for legend.

Figure 6.11 Model of nucleosome positioning surrounding CTCF sites during neural cell differentiation

CTCF binds to its motif in iPS cells and recruits cohesin and mediator. Super-nucleosomal chromatin particles are positioned flanking the CTCF binding motif. After differentiation to NPC cells, CTCF and its co-binding proteins are lost from the CTCF motif and the flanking nucleosomes are re-positioned towards the CTCF motif, suggesting that chromatin remodelling occurs at CTCF sites in the early stages of neural cell differentiation.

Chapter 7: General Discussion

7.1 Project aims

The aim of this project was to study changes in genome-wide chromatin particle positioning in the human genome during the earliest stages of neural cell development to answer the following questions:

1. Are nucleosomes randomly distributed or strategically placed in the human genome?

2. Does chromatin structure change during neural development?

3. Does chromatin structure change at regulatory sites during neural development?

4. Can I identify genes that are targeted for regulation during neural development?

ATP-dependent chromatin remodelling complexes regulate nucleosome positioning and occupancy in chromatin. It has been shown that mutations in genes that encode chromatin remodellers are highly associated with genetic risk (De Rubeis, He et al. 2014). Prevalent among these are genes whose mis-regulation occurs in neurodevelopmental disorders such as Rett's, Kleefstra's, and CHARGE syndromes (Amir, Van den Veyver et al. 1999; Vissers, van Ravenswaaij et al. 2004; Kleefstra, van Zelst-Stams et al. 2009). Hence, mis-regulation of chromatin structure is implicated in a range of childhood neuro-developmental syndromes and mental health disorders. Chromatin dynamics appears to have a pivotal role in gene regulation and its mis-regulation is implicated neuro-psychiatric disorders. This raises the questions of how and why does the control of nucleosome positioning increase the risk of neuropsychiatric disorders. By studying the dynamic changes in chromatin structure that occur during human neuro-development, the work in this thesis has revealed fundamental information about nucleosome positioning that is required to begin to answer this question.

7.2 Bioinformatics Methods.

In order to determine the changes in patterns of chromatin structure that occur during differentiation from induced pluripotent cells to neural progenitor cells, high resolution genome-wide chromatin maps from both cell types were constructed and compared. As the human genome is large (approximately $3x10^9$ base pairs) this study involved developing a work flow for processing 'big data' using high performance computing. To analyse these maps, I used both previously published methods, such as 'SiteWriter' (Kent, Adams et al. 2011) and methods that I developed during the course of this work. The 'SiteWriter' tool generates average genome-wide pattern of chromatin particle positioning at specific genomic sites normalised to the mean particle position per bin value in the window surrounding the site.

To investigate the patterns of positioned nucleosomes at regulatory sites in more detail, I used two other methods of analysis. Firstly, I optimised the methodology for cluster analysis in R to apply to chromatin particle positioning at regulatory sites in the human genome. Secondly, in order to locate individual positioned chromatin particles genome-wide and to align them to regulatory regions of the human genome, I developed a computational method for locating the positions of Gaussian peaks derived from the cumulative read count in each of three consecutive 10 bp bins genome-wide. The resulting tool, PeakFinder, was a simple heuristic peak marker. The parameters used to define this peak were set stringently to detect high numbers of reads in the central bin of the peak, so as to detect well positioned chromatin particles. Hence, this tool detects only very strongly positioned chromatin particles. This means that the estimates of the total numbers of positioned particles might be low, since chromatin particles that are positioned, albeit weakly would not have been detected using this tool. However, the total of numbers of positioned chromatin particles obtained was in a similar range to those found by others using different cell types and different methods of peak detection (Gaffney, McVicker et al. 2012). The major way in which this tool was used in this thesis was to locate well-positioned chromatin particles in the maps from undifferentiated iPS cells and the same cells developed to NPC. These three complementary methods provided information about both the global patterns of nucleosome positioning across the genomes of both iPS and NPC cells and local patterns at specific regulatory regions of the genome.

7.3 Key findings

7.3.1 Global patterns of chromatin particle positioning in the human genome

The genome-wide analysis of the chromatin particle positioning maps in both iPS and NPC cells in this thesis showed that it was possible to detect individual positioned chromatin particles. These were divided into in three size classes: sub-nucleosomal, nucleosomal and super-nucleosomal. The work presented in this thesis shows that broadly, well positioned chromatin particles in the human genome are not positioned at high density across the genome; they are usually distributed singly or in groups on two or three. This result is in contrast to the nucleosome organisation in the *S. cerevisiae* genome in which 85% of the genome possesses positioned nucleosomes and they tend to be organised in arrays (Mavrich, Ioshikhes et al. 2008; Kent, Adams et al. 2011). This contrast in chromatin organisation between the human and yeast genomes may not be as surprising as it seems, as the structure and complexity of the human genome is very different from that in *S. cerevisiae*, in overall size, gene density and gene organisation. A human cell has 20 Mb of DNA per μm^3 whereas yeast has 4Mb of DNA per μm^3. In addition, the *S. cerevisiae* genome is approximately 12 MB in size and possesses

approximately 6,000 genes (Goffeau, Barrell et al. 1996), hence is gene-dense. The human genome is 240 times larger (3GB of DNA) with approximately 20,000 protein coding genes (Lander, Linton et al. 2001). It has been shown that the *S. cerevisiae* genome has 67,517 positioned nucleosomes (Brogaard, Xi et al. 2012). Hence in *S. cerevisiae*, on average, for every kilo-base of DNA there are 5 positioned nucleosomes, leaving only 200 bp of nucleosome free DNA. Using the peak finding tool that I developed to detect well positioned nucleosomes (PeakFinder), the work presented here makes a conservative estimate that human NPC cells possess approximately $1x10^6$ well-positioned nucleosomes. This means that, on average, there is only one nucleosome for every 3kb of DNA in NPC cells. This low density of well positioned nucleosomes in the human genome has been shown in other human cell types, suggesting that this may occur generally across human cell types (Valouev, Johnson et al. 2011; Gaffney, McVicker et al. 2012) . The human genome possesses only fifteen fold more positioned chromatin particles than the *S. cerevisiae* genome. This difference in number of positioned chromatin particles is not in proportion to the difference in size of these two genomes, illustrating that the density of strongly positioned nucleosomes across the genome is much lower in human cells than in *S. cerevisiae*.

The presence of a nucleosome free region (NFR) in expressed genes determined from micro-array data in human cells suggests that there might be a similar local regulatory function of nucleosome positioning at gene promoters in the human genome to that in *S. cerevisiae* (Ozsolak, Song et al. 2007; Tirosh and Barkai 2008). However, global chromatin structure may differ in humans from that in *S. cerevisiae* because of differences in the overall organisation of the regulatory components of the genome. This suggests that that the mechanisms of genome regulation that occur as a result of nucleosome positioning might differ between the human and the *S. cerevisiae* genomes; individual nucleosomes appear to be strategically placed rather than being randomly distributed in the human genome. This work provides the answer to question 1.

7.3.2 Patterns of nucleosome positioning at regulatory regions of the human genome

Having examined the general patterns of well -positioned chromatin particles across the human genome, I examined the patterns of chromatin particle positioning at specific regulatory sites: selected transcription factor binding sites, TSS, CTCF and RE1 sites. This work showed that it is possible to detect positioned nucleosomes at TSS, CTCF and RE1 binding sites across the genome but not at the other selected transcription factor binding sites tested in the work of this thesis.

Examination of the global patterns of chromatin particle positioning at TSS of protein coding genes in both undifferentiated and differentiated cells did not show a canonical pattern of chromatin particle positioning. This is in contrast to the canonical patterns of nucleosome positioning found at TSS in the *S. cerevisiae* genome, which have a nucleosome free region flanked by arrays of well-positioned nucleosomes with weaker positioning upstream than downstream of the TSS (Mavrich, Ioshikhes et al. 2008; Jiang and Pugh 2009). This could be explained if the patterns of chromatin particle positioning are variable at TSS across the human genome. In this case, any patterns of chromatin particle positioning would be masked by plotting the average chromatin particle frequency distribution. In addition, the size and complexity of the human genome should be taken into account. There are approximately 198,000 annotated transcripts in the human genome, of which approximately 80,000 are protein coding transcripts. As there are only approximately 20,000 protein coding genes in total in the human genome this means that each one generates multiple transcripts (GENCODE Release (version 22) www.gencodegenes.org). In addition, it has been shown using cluster analysis, that there are different patterns of chromatin structure at TSS in human cells (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012). Taken together, this means that the number and complexity of the TSS in the human genome makes global analysis of patterns of chromatin particle positioning at TSS challenging. Hence, in this study it was decided to use the peak finding method developed in this thesis. This was used firstly to locate strongly positioned chromatin particles at TSS in both iPS and NPC cells and secondly to determine how many and which of these positioned chromatin particles were exclusive or in common between the two cell types. This led to the finding that it was possible to detect loss and gain of individual positioned nucleosomes at TSS during differentiation from iPS to NPC cells. These changes were detected at the TSS of genes involved in both the maintenance of pluripotency (e.g. Nanog) and at the TSS of neural -specific genes (e.g. GRIA1 which encodes an AMPA receptor subunit). At the TSS of selected genes involved in the maintenance of pluripotency and at neural-specific genes it was shown that the loss and gain of nucleosomes occurring during differentiation from iPS to NPC cells, correlated with changes in gene expression. In addition, it was shown that extent of the increase in number of positioned nucleosomal size chromatin particles at TSS that occurred during differentiation from iPS to NPC cells broadly reflected the global increase in the number of positioned nucleosomal chromatin particles at TSS during differentiation from iPS to NPC.

I have shown from global analysis of positioned chromatin particles, that the additional nucleosomal chromatin particles in NPC cells are most likely positioned *de novo*. This suggests that independent *de novo* positioning of chromatin particles may occur at the TSS of neural

specific genes during neural cell differentiation, rather than through the inter-conversion of existing particles of another size class. However, it cannot be excluded that the pattern of chromatin particle positioning at TSS may differ from the global pattern described here. It has been shown in *S. cerevisiae* that nucleosomes at gene promoters can be partially unwrapped (Floer, Wang et al. 2010), that they can be more accessible to MNase cleavage (Weiner, Hughes et al. 2010) and that smaller, approximately 100 bp chromatin particles are located upstream of the TSS of highly expressed protein coding genes (Kent, Adams et al. 2011). Hence it is possible that larger nucleosomal chromatin particles might convert to sub-nucleosomal chromatin particles at TSS in human cells and this might account for the small overlap in the positions of chromatin particles shown in this study.

Patterns of positioned nucleosomes have been shown to occur flanking CTCF sites and RE1 sites (Fu, Sinha et al. 2008; Valouev, Johnson et al. 2011; Teif, Vainshtein et al. 2012; Nie, Cheng et al. 2014), but this the first time that the changes in chromatin particle positioning at these sites have been investigated in detail in human cells during early neural development. It was shown here that MNase-seq can be used to detect sub-nucleosomal chromatin particles at both RE1 and CTCF sites. These sub-nucleosomal chromatin particles most likely represent transcription factor complexes at their binding sites, but could represent smaller/remodelled nucleosomes. Arrays of positioned nucleosomes were detected flanking both RE1 sites and CTCF sites in iPS cells. In both cases this correlated with the presence of a sub-nucleosomal particle at these sites. After differentiation to NPC cells changes in chromatin particle positioning were detected both at RE1 and CTCF sites. However, the types of changes in nucleosome positioning occurring during neural cell differentiation were specific to each transcription factor. CTCF appears to remain bound to its site in NPC cells and nucleosomes remain positioned flanking it. This suggested that nucleosome positioning at CTCF sites in both cell types might be dependent upon CTCF binding. Initially it appeared that there was little change in CTCF binding or nucleosome positioning at the set of CTCF sites utilised here during differentiation from iPS to NPC. A possible explanation for this would have been that the set of CTCF sites used in this study, derived from H1-ESC cells, was specific to pluripotent cells and therefore a completely different set of sites might be utilised in NPC cells to regulate neural specific genes. It has been shown in mouse cells that CTCF site utilisation might be cell type specific during early neural differentiation (Teif, Vainshtein et al. 2012), so it would be interesting to investigate this possibility in human cells. However, closer examination of the chromatin particle positioning at CTCF sites in NPC cells revealed that the levels of CTCF binding might be lower in NPC cells. Interestingly, loss of occupancy at CTCF sites also has been shown during differentiation of mouse ESC cells to embryonic fibroblasts (MEFs) (Teif,

Beshnova et al. 2014). In addition to a loss of CTCF occupancy, it was shown here that nucleosomes are re-positioned towards the CTCF site in NPC cells. This suggests that chromatin remodelling occurs during neural cell development at the H1-ESC cell derived set of CTCF sites analysed here, but does not rule out the possibility that other CTCF sites might be used to activate neural-specific genes. CTCF might bind to specific CTCF sites in iPS cells and recruit remodelling complexes to position nucleosomes flanking them. During differentiation, loss of CTCF at these sites might result in the removal of a 'barrier' and allow the flanking nucleosomes to redistribute , encroaching on the CTCF site, perhaps blocking the activity of transcription factors that activate genes involved in the maintenance of pluripotency. This nucleosome re-positioning might be brought about remodellers that slide nucleosomes, for example, the NURF complex. Kagey *et al* have shown that the formation of DNA loops between enhancers and promoters in mouse ESC's brings about cell type specific gene expression (Kagey, Newman et al. 2010). In addition, Zhang *et al* have shown that in order to reprogram human somatic cells to iPSC, intra-chromosomal looping is required to re-activate pluripotency genes (Zhang, Jiao et al. 2013). These studies emphasise the importance of higher order chromatin structure in the regulation of pluripotency and thus in development.

In contrast to CTCF binding, occupancy at RE1 sites, perhaps by REST or the REST complex, was not detected during differentiation to NPC cells in this study. In Chapter 5, a large peak in the frequency of sub-nucleosomal chromatin particles (112-137 bp) was detected centred on RE1 sites in iPS cells but this peak was not detected at RE1 sites in NPC cells. Nevertheless, nucleosomal chromatin particles remained strongly positioned flanking approximately 50% of the RE1 sites derived in this study, during differentiation from iPS to NPC cells. This suggested that nucleosome positioning at RE1 sites is independent of the binding of REST or the REST complex and that nucleosomes remain positioned at this subset of RE1 sites during neural cell development. This analysis led to the identification specific groups of REST target genes which retain chromatin structure flanking RE1 sites during neural cell differentiation. This suggests that subsets of REST target genes are targeted during differentiation from iPS to NPC cells, thus answering question four.

It was shown here that there are some similarities in the chromatin structure flanking RE1 and CTCF sites in iPS cells. The presence of arrays of positioned nucleosomes flanking RE1 and CTCF sites correlates with the presence of a sub-nucleosomal chromatin particle in iPS cells. This is likely to be the transcription factor or a transcription factor complex. In both cases in iPS cells, these transcription factors might bind and recruit chromatin remodelling complexes that position nucleosomes flanking their binding sites. However after differentiation to NPC cells, the patterns of nucleosome positioning flanking RE1 and CTCF sites differ, suggesting that

there might be different mechanisms of regulation at these two important transcription factor binding sites. Nucleosomes that remain positioned at transcription factor binding sites after early differentiation, might act as an 'epigenetic memory' so that the transcription factor can rebind to the site, affecting gene regulation at a later time in development, for example in the case of REST. Movement of nucleosomes towards the transcription factor site might occlude the site, restricting transcription factor access. In the case of CTCF, restricting access to its binding site might prevent the recruitment of cohesin, resulting in an inability to form chromatin loops. This might result in the repression of pluripotent genes.

The work in this thesis has shown that chromatin structure does change at regulatory sites in human cells during neural cell development, thus answering question 3. The local changes in nucleosome positioning at regulatory sites during early development shown here might be fundamental in affecting the balance of activation and repression of genes involved in the specification of cell fate in neural development.

7.3.3 Changes in chromatin particle positioning during neural development.

I have shown that chromatin particle positioning changes both locally and globally during neural cell development. Local changes in nucleosome positioning were shown to occur at regulatory sites during early neural cell differentiation. This was shown in Chapters 5 and 6. Evidence for global changes in the numbers of positioned chromatin particles was shown in Chapter 3. The number of positioned chromatin particles increased during differentiation from iPS to NPC cells. In addition I have shown that similar high numbers of positioned chromatin particles are found in the published chromatin maps from K562 and GM12878 cells (Kundaje, Kyriazopoulou-Panagiotopoulou et al. 2012). Interestingly, recent work by West *et al* has shown in both mouse and human cells that nucleosome occupancy increases during differentiation from iPS to fibroblast cells (West, Cook et al. 2014). Taken together, these results suggest that chromatin is more organised in differentiated cells than in pluripotent cells and that the presence of high numbers of strongly positioned nucleosomes is a characteristic of differentiated cells.

Observations suggesting that in pluripotent cells, chromatin organisation is both accessible and dynamic, whereas in differentiated cells chromatin is more compact, were discussed in Chapter 1. These observations led to the idea that pluripotent cell chromatin architecture is 'open' and 'fluid' allowing the access of cell machinery to regulatory regions of the genome that are important in specifying cell fate. The evidence presented in this thesis supports these observations and answers question two.

The increase in the number of positioned chromatin particles shown in Chapter 3 was most marked in the nucleosomal chromatin particle size class, suggesting that this class of chromatin particle is the most dynamic during neural cell development. In the human genome histone H1 containing nucleosomes, known as chromatosomes, might be represented by the super-nucleosomal chromatin particle size class. The H1 depleted or core nucleosomes might be represented by the nucleosomal chromatin particle size class. This strong positioning of core nucleosomes in NPC cells suggested a high degree of genome regulation in these cells by this chromatin particle size class. This fits with previous observations discussed in Chapter 1, section 1.3, that histone H1 containing nucleosomes are depleted in active chromatin and that H1 containing nucleosomes increase chromatin compaction, leading to transcriptional repression.

As I have shown in Chapter 3 (Fig 3.8), that the chromatin particle size classes described in this study are largely discrete, it would be interesting to know whether these nucleosomal particles are enriched regions of active chromatin in NPC cells. This could be tested by ChIP-seq using antibodies against active chromatin marks, such as H3K9ac and H3K27ac or against the histone variant H3.3.

7.4 Future work

7.4.1 A targeted approach to chromatin particle positioning analysis.

The size of the human genome means that it is very expensive to undertake sequencing it in its entirety multiple times to make comparisons between cells types. Now that I have constructed and characterised a basic chromatin map from both a pluripotent and a differentiated cell type, it would be possible to take a targeted enrichment approach to studying specific areas of interest in the genome. It might be possible to do this using an approach that combines ChiP with MNase-seq, selecting the regions of the genome of interest using an antibody, then characterising the chromatin structure in that region. Alternatively regions of interest in the genome could be selected after MNase digestion by DNA hybridisation with the region of interest (Yigit, Zhang et al. 2013).

7.4.2 The relationship between changes in nucleosome positioning and human neuro-psychiatric disorders.

Mutations in the genes that encode ATP-dependent chromatin remodelling proteins are highly associated with genetic risk. Specifically, approximately half of the CHD group of chromatin remodellers that can move and evict nucleosomes, (CHD2, 5, 7 and 8) have neuro-psychiatric disease associations in humans. Now it is possible to begin to examine how the regulation of

nucleosome positioning is involved in genetic risk of neuropsychiatric disorders. This could be carried out by integrating maps of human mutations (SNPs and small indels) with the nucleosome positioning data presented in this thesis. This would correlate genome-wide changes in chromatin structure with common genetic variants associated with neuro-psychiatric diseases. This analysis would reveal the regulatory regions of the genome that are targeted by chromatin remodelling proteins.

In relation to the work presented here on the analysis of chromatin structure at RE1 sites, it would be interesting to examine changes in chromatin structure at later stages of neural development by developing human NPC cells to a later stage of differentiation in culture. Also it would be interesting to examine chromatin structure at RE1 sites after seizures. In the latter case, this would involve developing MNase seq methods from cells derived from rat models, for example after kainate-induced seizures (Hellier and Dudek 1999). This would show how chromatin structure relates to specific sets of neuronal genes and further determine the relationship between the regulation of REST and chromatin structure in human disease. These approaches would lead to insight into the mechanisms of regulation of the neural cell genome and ultimately to deciphering the causes human neuro-psychiatric disorders.

7.4.3 Further global analysis of the locations of the positioned chromatin particles.

It is clear that a large amount of information is contained in these chromatin particle positioning maps, hence further analysis will likely reveal answers to some of the additional questions raised during this work. For example, in this work I have made a conservative estimate of the numbers of well positioned chromatin particles in differentiated cells and located their “geographical” locations in the genome. I have mapped these positioned chromatin particles to known regulatory regions in the genome; some positioned chromatin particles are located at TSS and some are in arrays flanking RE1 sites and CTCF sites. The number of positioned chromatin particles surrounding RE1 sites and CTCF sites genome-wide was obtained by multiplying the number of peaks obtained in the average cumulative frequency distributions by the number of sites analysed in the genome in each case. This gives an approximate number of nucleosomes surrounding these binding sites in the genome. As this number was obtained using the results from the average cumulative frequency distributions, it is likely to be an overestimate. However, this does not alter the conclusion that most of the positioned super-nucleosomal chromatin particles in differentiated cells that are accounted for in this study are at CTCF sites and fewer are at RE1 sites and at TSS. This leaves approximately 75% of the positioned super-nucleosomal chromatin particles for which genomic positions have been obtained in this study, not yet accounted for (Fig 7.1). Therefore,

it would be interesting to map the remaining unaccounted for positioned chromatin particles to determine whether they are located at other known genomic features, or at so far undiscovered ones.

7.6 Concluding remarks.

I have characterised changes that occur in chromatin particle positioning during neural cell development. This has provided insight into the role of chromatin structure in the regulation of the human genome both in pluripotent and differentiated cells. In both cell types, only a small proportion of the genome possesses strongly positioned nucleosomes, but in the work of this thesis I have shown that it is possible to detect regulatory regions of the genome where dynamic changes in chromatin structure occur and regions where chromatin structure remain stable during neural cell differentiation. As these regulatory regions of the genome with positioned nucleosomes are small, this information could be used to develop methods of correlating changes in chromatin structure with genetic variants associated with human diseases. Developing epi-genome-wide association studies (EWAS) to correlate genome-wide changes in chromatin structure with common genetic variants associated with neuro-psychiatric diseases will provide a route to understanding the pathways connecting the mis-regulation of chromatin structure and disease risk.

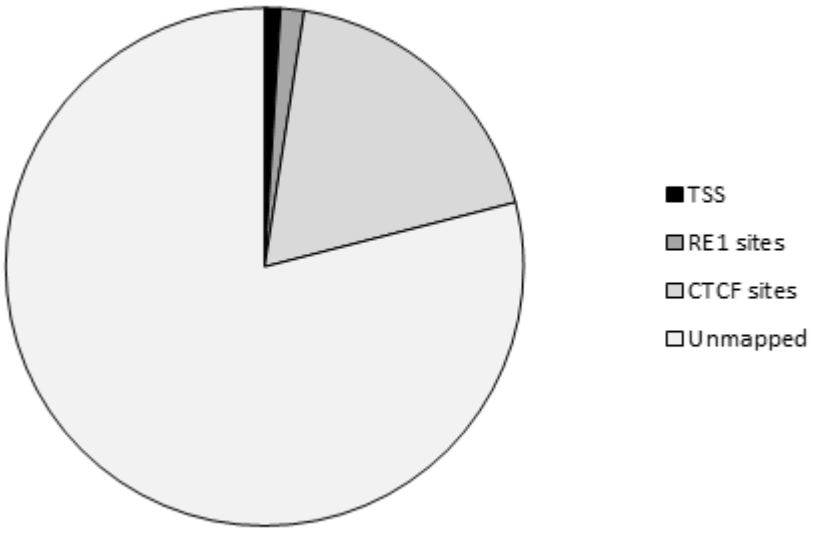


Figure 7.1 The number of super-nucleosomes mapped to the regulatory regions in the human genome in NPC cells.

Chapter 8: Bibliography

AGI (2000). "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana." Nature **408**(6814): 796-815.
Ahmad, K. and S. Henikoff (2002). "The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly." Molecular cell **9**(6): 1191-1200.
Ahmed, K., H. Dehghani, et al. (2010). "Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo." PloS one **5**(5): e10531.
Allan, J., P. G. Hartman, et al. (1980). "The structure of histone H1 and its location in chromatin." Nature **288**(5792): 675-679.
Allfrey, V. G., R. Faulkner, et al. (1964). "Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis." Proceedings of the National Academy of Sciences of the United States of America **51**: 786-794.
Amir, R. E., I. B. Van den Veyver, et al. (1999). "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2." Nature genetics **23**(2): 185-188.
Amit, M., M. K. Carpenter, et al. (2000). "Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture." Developmental biology **227**(2): 271-278.
Andres, M. E., C. Burger, et al. (1999). "CoREST: a functional corepressor required for regulation of neural-specific gene expression." Proceedings of the National Academy of Sciences of the United States of America **96**(17): 9873-9878.
Arents, G. and E. N. Moudrianakis (1995). "The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization." Proceedings of the National Academy of Sciences of the United States of America **92**(24): 11170-11174.
Auger, A., L. Galarneau, et al. (2008). "Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants." Molecular and cellular biology **28**(7): 2257-2270.
Ausio, J. and D. W. Abbott (2002). "The many tales of a tail: carboxyl-terminal tail heterogeneity specializes histone H2A variants for defined chromatin function." Biochemistry **41**(19): 5945-5949.
Axel, R. (1975). "Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease." Biochemistry **14**(13): 2921-2925.
Ballas, N., E. Battaglioli, et al. (2001). "Regulation of neuronal traits by a novel transcriptional complex." Neuron **31**(3): 353-365.
Ballas, N., C. Grunseich, et al. (2005). "REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis." Cell **121**(4): 645-657.
Ballas, N. and G. Mandel (2005). "The many faces of REST oversee epigenetic programming of neuronal genes." Current opinion in neurobiology **15**(5): 500-506.
Bao, L., M. Zhou, et al. (2008). "CTCFBSDB: a CTCF-binding site database for characterization of vertebrate genomic insulators." Nucleic acids research **36**(Database issue): D83-87.
Barak, O., M. A. Lazzaro, et al. (2003). "Isolation of human NURF: a regulator of Engrailed gene expression." The EMBO journal **22**(22): 6089-6100.
Barde, Y. A., D. Edgar, et al. (1982). "Purification of a new neurotrophic factor from mammalian brain." The EMBO journal **1**(5): 549-553.
Basu, A., F. H. Wilkinson, et al. (2014). "YY1 DNA binding and interaction with YAF2 is essential for Polycomb recruitment." Nucleic acids research **42**(4): 2208-2223.
Beketaev, I., Y. Zhang, et al. (2015). "Critical role of YY1 in cardiac morphogenesis." Developmental dynamics : an official publication of the American Association of Anatomists.
Berndsen, C. E. and J. M. Denu (2008). "Catalysis and substrate selection by histone/protein lysine acetyltransferases." Current opinion in structural biology **18**(6): 682-689.
Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-326.

Bersten, D. C., J. A. Wright, et al. (2014). "Regulation of the neuronal transcription factor NPAS4 by REST and microRNAs." Biochimica et biophysica acta **1839**(1): 13-24.
Bessis, A., A. M. Salmon, et al. (1995). "Promoter elements conferring neuron-specific expression of the beta 2-subunit of the neuronal nicotinic acetylcholine receptor studied in vitro and in transgenic mice." Neuroscience **69**(3): 807-819.
Bhinge, A., J. Poschmann, et al. (2014). "MiR-135b is a direct PAX6 target and specifies human neuroectoderm by inhibiting TGF-beta/BMP signaling." The EMBO journal **33**(11): 1271-1283.
Boeger, H., J. Griesenbeck, et al. (2003). "Nucleosomes unfold completely at a transcriptionally active promoter." Molecular cell **11**(6): 1587-1598.
Bolzer, A., G. Kreth, et al. (2005). "Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes." PLoS biology **3**(5): e157.
Bonisch, C., K. Schneider, et al. (2012). "H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization." Nucleic acids research **40**(13): 5951-5964.
Boyer, L. A., R. R. Latek, et al. (2004). "The SANT domain: a unique histone-tail-binding module?" Nature reviews. Molecular cell biology **5**(2): 158-163.
Boyer, L. A., T. I. Lee, et al. (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**(6): 947-956.
Boyle, A. P., L. Song, et al. (2011). "High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells." Genome research **21**(3): 456-464.
Brene, S., C. Messer, et al. (2000). "Regulation of GluR2 promoter activity by neurotrophic factors via a neuron-restrictive silencer element." The European journal of neuroscience **12**(5): 1525-1533.
Brogaard, K., L. Xi, et al. (2012). "A map of nucleosome positions in yeast at base-pair resolution." Nature **486**(7404): 496-501.
Bruce, A. W., I. J. Donaldson, et al. (2004). "Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes." Proceedings of the National Academy of Sciences of the United States of America **101**(28): 10458-10463.
Bruce, A. W., A. J. Lopez-Contreras, et al. (2009). "Functional diversity for REST (NRSF) is defined by in vivo binding affinity hierarchies at the DNA sequence level." Genome research **19**(6): 994-1005.
Burgoyne, L. A. (1985). "A back-to-back zig-zag model for higher order chromatin structure." Cytobios **43**(172-173): 141-147.
Burgoyne, L. A., D. R. Hewish, et al. (1974). "Mammalian chromatin substructure studies with the calcium-magnesium endonuclease and two-dimensional polyacrylamide-gel electrophoresis." The Biochemical journal **143**(1): 67-72.
Burney, M. J., C. Johnston, et al. (2013). "An epigenetic signature of developmental potential in neural stem cells and early neurons." Stem cells **31**(9): 1868-1880.
Cai, H. and M. Levine (1995). "Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo." Nature **376**(6540): 533-536.
Cai, Y., J. Jin, et al. (2007). "YY1 functions with INO80 to activate transcription." Nature structural & molecular biology **14**(9): 872-874.
Cairns, B. R. (2009). "The logic of chromatin architecture and remodelling at promoters." Nature **461**(7261): 193-198.
Cairns, B. R., Y. J. Kim, et al. (1994). "A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast." Proceedings of the National Academy of Sciences of the United States of America **91**(5): 1950-1954.
Cairns, B. R., Y. Lorch, et al. (1996). "RSC, an essential, abundant chromatin-remodeling complex." Cell **87**(7): 1249-1260.

Calderone, A., T. Jover, et al. (2003). "Ischemic insults derepress the gene silencer REST in neurons destined to die." The Journal of neuroscience : the official journal of the Society for Neuroscience **23**(6): 2112-2121.
Carpenter, M. K., X. Cui, et al. (1999). "In vitro expansion of a multipotent population of human neural progenitor cells." Experimental neurology **158**(2): 265-278.
Chambers, S. M., C. A. Fasano, et al. (2009). "Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling." Nature biotechnology **27**(3): 275-280.
Chang, G. S., A. A. Noegel, et al. (2012). "Unusual combinatorial involvement of poly-A/T tracts in organizing genes and chromatin in Dictyostelium." Genome research **22**(6): 1098-1106.
Chen, C. C., J. J. Carson, et al. (2008). "Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair." Cell **134**(2): 231-243.
Chen, H., Y. Tian, et al. (2012). "Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome." PloS one **7**(7): e41374.
Chen, K., Y. Xi, et al. (2013). "DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing." Genome research **23**(2): 341-351.
Chen, L., Y. Cai, et al. (2011). "Subunit organization of the human INO80 chromatin remodeling complex: an evolutionarily conserved core complex catalyzes ATP-dependent nucleosome remodeling." The Journal of biological chemistry **286**(13): 11283-11289.
Chen, L., R. C. Conaway, et al. (2013). "Multiple modes of regulation of the human Ino80 SNF2 ATPase by subunits of the INO80 chromatin-remodeling complex." Proceedings of the National Academy of Sciences of the United States of America **110**(51): 20497-20502.
Chen, Y., J. Lu, et al. (2003). "Association between genetic variation of CACNA1H and childhood absence epilepsy." Annals of neurology **54**(2): 239-243.
Chenier, S., G. Yoon, et al. (2014). "CHD2 haploinsufficiency is associated with developmental delay, intellectual disability, epilepsy and neurobehavioural problems." Journal of neurodevelopmental disorders **6**(1): 9.
Cheutin, T., A. J. McNairn, et al. (2003). "Maintenance of stable heterochromatin domains by dynamic HP1 binding." Science **299**(5607): 721-725.
Chin, M. H., M. J. Mason, et al. (2009). "Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures." Cell stem cell **5**(1): 111-123.
Chodavarapu, R. K., S. Feng, et al. (2010). "Relationship between nucleosome positioning and DNA methylation." Nature **466**(7304): 388-392.
Chow, J. C., Z. Yen, et al. (2005). "Silencing of the mammalian X chromosome." Annual review of genomics and human genetics **6**: 69-92.
Christophorou, M. A., G. Castelo-Branco, et al. (2014). "Citrullination regulates pluripotency and histone H1 binding to chromatin." Nature **507**(7490): 104-108.
Clapier, C. R. and B. R. Cairns (2009). "The biology of chromatin remodeling complexes." Annual review of biochemistry **78**: 273-304.
Collings, C. K., P. J. Waddell, et al. (2013). "Effects of DNA methylation on nucleosome stability." Nucleic acids research **41**(5): 2918-2931.
Compton, J. L., M. Bellard, et al. (1976). "Biochemical evidence of variability in the DNA repeat length in the chromatin of higher eukaryotes." Proceedings of the National Academy of Sciences of the United States of America **73**(12): 4382-4386.
Conaco, C., S. Otto, et al. (2006). "Reciprocal actions of REST and a microRNA promote neuronal identity." Proceedings of the National Academy of Sciences of the United States of America **103**(7): 2422-2427.
Cremer, T. and M. Cremer (2010). "Chromosome territories." Cold Spring Harbor perspectives in biology **2**(3): a003889.
Creyghton, M. P., S. Markoulaki, et al. (2008). "H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment." Cell **135**(4): 649-661.

Cuddapah, S., R. Jothi, et al. (2009). "Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains." Genome research **19**(1): 24-32.
Das, S., S. Jena, et al. (2011). "Alternative splicing produces Nanog protein variants with different capacities for self-renewal and pluripotency in embryonic stem cells." The Journal of biological chemistry **286**(49): 42690-42703.
de Koning, A. P., W. Gu, et al. (2011). "Repetitive elements may comprise over two-thirds of the human genome." PLoS genetics **7**(12): e1002384.
De Rubeis, S., X. He, et al. (2014). "Synaptic, transcriptional and chromatin genes disrupted in autism." Nature **515**(7526): 209-215.
de Wit, E. and W. de Laat (2012). "A decade of 3C technologies: insights into nuclear organization." Genes & development **26**(1): 11-24.
Deindl, S., W. L. Hwang, et al. (2013). "ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps." Cell **152**(3): 442-452.
Dekker, J., M. A. Marti-Renom, et al. (2013). "Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data." Nature reviews. Genetics **14**(6): 390-403.
Delmas, V., D. G. Stokes, et al. (1993). "A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain." Proceedings of the National Academy of Sciences of the United States of America **90**(6): 2414-2418.
Di Cerbo, V., F. Mohn, et al. (2014). "Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription." eLife **3**: e01632.
Ding, N., C. Tomomori-Sato, et al. (2009). "MED19 and MED26 are synergistic functional targets of the RE1 silencing transcription factor in epigenetic silencing of neuronal gene expression." The Journal of biological chemistry **284**(5): 2648-2656.
Dixon, J. R., I. Jung, et al. (2015). "Chromatin architecture reorganization during stem cell differentiation." Nature **518**(7539): 331-336.
Dixon, J. R., S. Selvaraj, et al. (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions." Nature **485**(7398): 376-380.
Donze, D. and R. T. Kamakaka (2001). "RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae." The EMBO journal **20**(3): 520-531.
Drew, H. R. and A. A. Travers (1985). "DNA bending and its relation to nucleosome positioning." Journal of molecular biology **186**(4): 773-790.
Dubochet, J., M. Adrian, et al. (1988). "Cryo-electron microscopy of vitrified specimens." Quarterly reviews of biophysics **21**(2): 129-228.
Dunham, I. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.
Ebersole, T., J. H. Kim, et al. (2011). "tRNA genes protect a reporter gene from epigenetic silencing in mouse cells." Cell cycle **10**(16): 2779-2791.
Egan, C. M., U. Nyman, et al. (2013). "CHD5 is required for neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression." Developmental cell **26**(3): 223-236.
Eltsov, M., K. M. Maclellan, et al. (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 of the United States of America **105**(50): 19732-19737.
Engelhardt, M. (2007). "Choreography for nucleosomes: the conformational freedom of the nucleosomal filament and its limitations." Nucleic acids research **35**(16): e106.
Ernst, J. and M. Kellis (2012). "ChromHMM: automating chromatin-state discovery and characterization." Nature methods **9**(3): 215-216.
Ernst, J. and M. Kellis (2013). "Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types." Genome research **23**(7): 1142-1154.

Ernst, J., P. Kheradpour, et al. (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types." Nature **473**(7345): 43-49.
Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-156.
Fan, Y., T. Nikitina, et al. (2005). "Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation." Cell **123**(7): 1199-1212.
Feil, R. and M. F. Fraga (2011). "Epigenetics and the environment: emerging patterns and implications." Nature reviews. Genetics **13**(2): 97-109.
Filippova, G. N., S. Fagerlie, et al. (1996). "An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes." Molecular and cellular biology **16**(6): 2802-2813.
Finch, J. T. and A. Klug (1976). "Solenoidal model for superstructure in chromatin." Proceedings of the National Academy of Sciences of the United States of America **73**(6): 1897-1901.
Flanagan, J. F., L. Z. Mi, et al. (2005). "Double chromodomains cooperate to recognize the methylated histone H3 tail." Nature **438**(7071): 1181-1185.
Flemming, W. (1882). "Zellsubstanz, Kern und Zelltheilung." F.C.W. Vogel, Leipzig.
Flemming, W. (1965). "Contributions to the Knowledge of the Cell and Its Vital Processes." The Journal of cell biology **25**(1): 3-69.
Floer, M., X. Wang, et al. (2010). "A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding." Cell **141**(3): 407-418.
Fromer, M., A. J. Pocklington, et al. (2014). "De novo mutations in schizophrenia implicate synaptic networks." Nature **506**(7487): 179-184.
Fu, Y., M. Sinha, et al. (2008). "The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome." PLoS genetics **4**(7): e1000138.
Fujita, T., J. Igarashi, et al. (2008). "CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas." Journal of the National Cancer Institute **100**(13): 940-949.
Gaffney, D. J., G. McVicker, et al. (2012). "Controls of nucleosome positioning in the human genome." PLoS genetics **8**(11): e1003036.
Gafni, O., L. Weinberger, et al. (2013). "Derivation of novel human ground state naive pluripotent stem cells." Nature **504**(7479): 282-286.
Garcia, B. A., S. A. Busby, et al. (2004). "Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry." Journal of proteome research **3**(6): 1219-1227.
Gaspar-Maia, A., A. Alajem, et al. (2009). "Chd1 regulates open chromatin and pluripotency of embryonic stem cells." Nature **460**(7257): 863-868.
Geisler, S. and J. Coller (2013). "RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts." Nature reviews. Molecular cell biology **14**(11): 699-712.
Gerstein, M. B., A. Kundaje, et al. (2012). "Architecture of the human regulatory network derived from ENCODE data." Nature **489**(7414): 91-100.
Glaser, T., D. S. Walton, et al. (1992). "Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene." Nature genetics **2**(3): 232-239.
Goffeau, A., B. G. Barrell, et al. (1996). "Life with 6000 genes." Science **274**(5287): 546, 563-547.
Goldberg, A. D., L. A. Banaszynski, et al. (2010). "Distinct factors control histone variant H3.3 localization at specific genomic regions." Cell **140**(5): 678-691.
Gossett, A. J. and J. D. Lieb (2012). "In vivo effects of histone H3 depletion on nucleosome occupancy and position in Saccharomyces cerevisiae." PLoS genetics **8**(6): e1002771.
Greer, E. L., M. A. Blanco, et al. (2015). "DNA Methylation on N(6)-Adenine in C. elegans." Cell **161**(4): 868-878.
Gregor, A., M. Oti, et al. (2013). "De novo mutations in the genome organizer CTCF cause intellectual disability." American journal of human genetics **93**(1): 124-131.

Grune, T., J. Brzeski, et al. (2003). "Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI." Molecular cell **12**(2): 449-460.
Guenther, M. G., G. M. Frampton, et al. (2010). "Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells." Cell stem cell **7**(2): 249-257.
Guillemette, B., A. R. Bataille, et al. (2005). "Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning." PLoS biology **3**(12): e384.
Guo, Y., K. Monahan, et al. (2012). "CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice." Proceedings of the National Academy of Sciences of the United States of America **109**(51): 21081-21086.
Guo, Y., Q. Xu, et al. (2015). "CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function." Cell **162**(4): 900-910.
Guttman, M., J. Donaghey, et al. (2011). "lincRNAs act in the circuitry controlling pluripotency and differentiation." Nature **477**(7364): 295-300.
Hai, T. W., F. Liu, et al. (1989). "Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers." Genes & development **3**(12B): 2083-2090.
Hakimi, M. A., D. A. Bochar, et al. (2002). "A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes." Proceedings of the National Academy of Sciences of the United States of America **99**(11): 7420-7425.
Han, M. and M. Grunstein (1988). "Nucleosome loss activates yeast downstream promoters in vivo." Cell **55**(6): 1137-1145.
Handoko, L., H. Xu, et al. (2011). "CTCF-mediated functional chromatin interactome in pluripotent cells." Nature genetics **43**(7): 630-638.
Hanna, J., A. W. Cheng, et al. (2010). "Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs." Proceedings of the National Academy of Sciences of the United States of America **107**(20): 9222-9227.
Hanson, I. and V. Van Heyningen (1995). "Pax6: more than meets the eye." Trends in genetics : TIG **11**(7): 268-272.
Happel, N. and D. Doenecke (2009). "Histone H1 and its isoforms: contribution to chromatin structure and function." Gene **431**(1-2): 1-12.
Harr, J. C., T. R. Luperchio, et al. (2015). "Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins." The Journal of cell biology **208**(1): 33-52.
Harrow, J., A. Frankish, et al. (2012). "GENCODE: the reference human genome annotation for The ENCODE Project." Genome research **22**(9): 1760-1774.
Hassan, A. H., P. Prochasson, et al. (2002). "Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes." Cell **111**(3): 369-379.
Havas, K., I. Whitehouse, et al. (2001). "ATP-dependent chromatin remodeling activities." Cellular and molecular life sciences : CMLS **58**(5-6): 673-682.
Hawkins, R. D., G. C. Hon, et al. (2010). "Distinct epigenomic landscapes of pluripotent and lineage-committed human cells." Cell stem cell **6**(5): 479-491.
Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.
Hellier, J. L. and F. E. Dudek (1999). "Spontaneous motor seizures of rats with kainate-induced epilepsy: effect of time of day and activity state." Epilepsy research **35**(1): 47-57.
Hellman, A. and A. Chess (2007). "Gene body-specific methylation on the active X chromosome." Science **315**(5815): 1141-1143.
Henikoff, J. G., J. A. Belsky, et al. (2011). "Epigenome characterization at single base-pair resolution." Proceedings of the National Academy of Sciences of the United States of America **108**(45): 18318-18323.

Henikoff, S. (2008). "Nucleosome destabilization in the epigenetic regulation of gene expression." Nature reviews. Genetics **9**(1): 15-26.
Hihara, S., C. G. Pack, et al. (2012). "Local nucleosome dynamics facilitate chromatin accessibility in living mammalian cells." Cell reports **2**(6): 1645-1656.
Hill, D. A. and A. N. Imbalzano (2000). "Human SWI/SNF nucleosome remodeling activity is partially inhibited by linker histone H1." Biochemistry **39**(38): 11649-11656.
Hillier, L. W., A. Coulson, et al. (2005). "Genomics in C. elegans: so many genes, such a little worm." Genome research **15**(12): 1651-1660.
Hiratani, I., T. Ryba, et al. (2008). "Global reorganization of replication domains during embryonic stem cell differentiation." PLoS biology **6**(10): e245.
Hong, L., G. P. Schroth, et al. (1993). "Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA." The Journal of biological chemistry **268**(1): 305-314.
Horn, P. J., L. M. Carruthers, et al. (2002). "Phosphorylation of linker histones regulates ATP-dependent chromatin remodeling enzymes." Nature structural biology **9**(4): 263-267.
Hu, B. Y., J. P. Weick, et al. (2010). "Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency." Proceedings of the National Academy of Sciences of the United States of America **107**(9): 4335-4340.
Hu, G., K. Cui, et al. (2013). "H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation." Cell stem cell **12**(2): 180-192.
Huang, Y., S. J. Myers, et al. (1999). "Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes." Nature neuroscience **2**(10): 867-872.
Huang, Z., Q. Wu, et al. (2011). "Deubiquitylase HAUSP stabilizes REST and promotes maintenance of neural progenitor cells." Nature cell biology **13**(2): 142-152.
Inman, G. J., F. J. Nicolas, et al. (2002). "SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7." Molecular pharmacology **62**(1): 65-74.
Ioshikhes, I. P., I. Albert, et al. (2006). "Nucleosome positions predicted through comparative genomics." Nature genetics **38**(10): 1210-1215.
Iyer, V. and K. Struhl (1995). "Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure." The EMBO journal **14**(11): 2570-2579.
Jiang, C. and B. F. Pugh (2009). "A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome." Genome biology **10**(10): R109.
Jiang, C. and B. F. Pugh (2009). "Nucleosome positioning and gene regulation: advances through genomics." Nature reviews. Genetics **10**(3): 161-172.
Jimenez-Useche, I., J. Ke, et al. (2013). "DNA methylation regulated nucleosome dynamics." Scientific reports **3**: 2121.
Johnson, D. S., A. Mortazavi, et al. (2007). "Genome-wide mapping of in vivo protein-DNA interactions." Science **316**(5830): 1497-1502.
Johnson, R., R. J. Gamblin, et al. (2006). "Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication." Nucleic acids research **34**(14): 3862-3877.
Johnson, R., C. H. Teh, et al. (2008). "REST regulates distinct transcriptional networks in embryonic and neural stem cells." PLoS biology **6**(10): e256.
Jones, P. L., G. J. Veenstra, et al. (1998). "Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription." Nature genetics **19**(2): 187-191.
Jorgensen, H. F., A. Terry, et al. (2009). "REST selectively represses a subset of RE1-containing neuronal genes in mouse embryonic stem cells." Development **136**(5): 715-721.
Jothi, R., S. Cuddapah, et al. (2008). "Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data." Nucleic acids research **36**(16): 5221-5231.

Kagey, M. H., J. J. Newman, et al. (2010). "Mediator and cohesin connect gene expression and chromatin architecture." Nature **467**(7314): 430-435.
Kaji, K., I. M. Caballero, et al. (2006). "The NuRD component Mbd3 is required for pluripotency of embryonic stem cells." Nature cell biology **8**(3): 285-292.
Kallunki, P., G. M. Edelman, et al. (1997). "Tissue-specific expression of the L1 cell adhesion molecule is modulated by the neural restrictive silencer element." The Journal of cell biology **138**(6): 1343-1354.
Kamakaka, R. T. and S. Biggins (2005). "Histone variants: deviants?" Genes & development **19**(3): 295-310.
Kasten, M., H. Szerlong, et al. (2004). "Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14." The EMBO journal **23**(6): 1348-1359.
Kawasaki, H., K. Taira, et al. (2000). "Histone acetyltransferase (HAT) activity of ATF-2 is necessary for the CRE-dependent transcription." Nucleic acids symposium series(44): 259-260.
Kellum, R. and P. Schedl (1991). "A position-effect assay for boundaries of higher order chromosomal domains." Cell **64**(5): 941-950.
Kemp, D. M., J. C. Lin, et al. (2002). "NRSF/REST confers transcriptional repression of the GPR10 gene via a putative NRSE/RE-1 located in the 5' promoter region." FEBS letters **531**(2): 193-198.
Kent, N. A., S. Adams, et al. (2011). "Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing." Nucleic acids research **39**(5): e26.
Kent, N. A. and J. Mellor (1995). "Chromatin structure snap-shots: rapid nuclease digestion of chromatin in yeast." Nucleic acids research **23**(18): 3786-3787.
Khalil, A. M., M. Guttman, et al. (2009). "Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression." Proceedings of the National Academy of Sciences of the United States of America **106**(28): 11667-11672.
Kim, C. S., C. K. Hwang, et al. (2004). "Neuron-restrictive silencer factor (NRSF) functions as a repressor in neuronal cells to regulate the mu opioid receptor gene." The Journal of biological chemistry **279**(45): 46464-46473.
Kim, T. H., Z. K. Abdullaev, et al. (2007). "Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome." Cell **128**(6): 1231-1245.
Kimura, A., K. Matsubara, et al. (2005). "A decade of histone acetylation: marking eukaryotic chromosomes with specific codes." Journal of biochemistry **138**(6): 647-662.
Kleefstra, T., W. A. van Zelst-Stams, et al. (2009). "Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype." Journal of medical genetics **46**(9): 598-606.
Klenova, E. M., R. H. Nicolas, et al. (1993). "CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms." Molecular and cellular biology **13**(12): 7612-7624.
Knapska, E. and L. Kaczmarek (2004). "A gene for neuronal plasticity in the mammalian brain: Zif268/Egr-1/NGFI-A/Krox-24/TIS8/ZENK?" Progress in neurobiology **74**(4): 183-211.
Koenig, M., A. P. Monaco, et al. (1988). "The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein." Cell **53**(2): 219-228.
Konev, A. Y., M. Tribus, et al. (2007). "CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo." Science **317**(5841): 1087-1090.
Korber, P. and W. Horz (2004). "In vitro assembly of the characteristic chromatin organization at the yeast PHO5 promoter by a replication-independent extract system." The Journal of biological chemistry **279**(33): 35113-35120.
Kornberg, R. D. and L. Stryer (1988). "Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism." Nucleic acids research **16**(14A): 6677-6690.

Kraner, S. D., J. A. Chong, et al. (1992). "Silencing the type II sodium channel gene: a model for neural-specific gene regulation." Neuron **9**(1): 37-44.
Kruger, W., C. L. Peterson, et al. (1995). "Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription." Genes & development **9**(22): 2770-2779.
Ku, M., J. D. Jaffe, et al. (2012). "H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions." Genome biology **13**(10): R85.
Kundaje, A., S. Kyriazopoulou-Panagiotopoulou, et al. (2012). "Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements." Genome research **22**(9): 1735-1747.
Kuo, M. H., J. Zhou, et al. (1998). "Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo." Genes & development **12**(5): 627-639.
Kurisaki, A., T. S. Hamazaki, et al. (2005). "Chromatin-related proteins in pluripotent mouse embryonic stem cells are downregulated after removal of leukemia inhibitory factor." Biochemical and biophysical research communications **335**(3): 667-675.
Kuwahara, K., Y. Saito, et al. (2003). "NRSF regulates the fetal cardiac gene program and maintains normal cardiac structure and function." The EMBO journal **22**(23): 6310-6321.
Lachner, M. and T. Jenuwein (2002). "The many faces of histone lysine methylation." Current opinion in cell biology **14**(3): 286-298.
Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Laneve, P., U. Gioia, et al. (2010). "A minicircuitry involving REST and CREB controls miR-9-2 expression during human neuronal differentiation." Nucleic acids research **38**(20): 6895-6905.
Langmead, B., C. Trapnell, et al. (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome biology **10**(3): R25.
Lantermann, A. B., T. Straub, et al. (2010). "Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae." Nature structural & molecular biology **17**(2): 251-257.
Laurent, B. C., I. Treich, et al. (1993). "The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation." Genes & development **7**(4): 583-591.
Lee, B. K. and V. R. Iyer (2012). "Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation." The Journal of biological chemistry **287**(37): 30906-30913.
Lee, D. Y., J. J. Hayes, et al. (1993). "A positive role for histone acetylation in transcription factor access to nucleosomal DNA." Cell **72**(1): 73-84.
Lee, W., D. Tillo, et al. (2007). "A high-resolution atlas of nucleosome occupancy in yeast." Nature genetics **39**(10): 1235-1244.
Lemonde, S., A. Rogaeva, et al. (2004). "Cell type-dependent recruitment of trichostatin A-sensitive repression of the human 5-HT1A receptor gene." Journal of neurochemistry **88**(4): 857-868.
Lever, M. A., J. P. Th'ng, et al. (2000). "Rapid exchange of histone H1.1 on chromatin in living human cells." Nature **408**(6814): 873-876.
Li, G., M. Levitus, et al. (2005). "Rapid spontaneous accessibility of nucleosomal DNA." Nature structural & molecular biology **12**(1): 46-53.
Li, G., S. Liu, et al. (2014). "ISWI proteins participate in the genome-wide nucleosome distribution in Arabidopsis." The Plant journal : for cell and molecular biology **78**(4): 706-714.
Li, W., J. Wu, et al. (2014). "Chd5 orchestrates chromatin remodelling during sperm development." Nature communications **5**: 3812.

Li, Z., J. Schug, et al. (2011). "The nucleosome map of the mammalian liver." Nature structural & molecular biology **18**(6): 742-746.
Liang, Y., M. Cucchetti, et al. (2013). "The lymphoid lineage-specific actin-uncapping protein Rltpr is essential for costimulation via CD28 and the development of regulatory T cells." Nature immunology **14**(8): 858-866.
Liao, J., R. Karnik, et al. (2015). "Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells." Nature genetics **47**(5): 469-478.
Lieberman-Aiden, E., N. L. van Berkum, et al. (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." Science **326**(5950): 289-293.
Lieleg, C., N. Krietenstein, et al. (2015). "Nucleosome positioning in yeasts: methods, maps, and mechanisms." Chromosoma **124**(2): 131-151.
Lilja, T., N. Heldring, et al. (2013). "Like a rolling histone: epigenetic regulation of neural stem cells and brain development by factors controlling histone acetylation and methylation." Biochimica et biophysica acta **1830**(2): 2354-2360.
Lim, L. P., N. C. Lau, et al. (2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs." Nature **433**(7027): 769-773.
Liu, J. C., C. G. Ferreira, et al. (2015). "Human CHD2 is a chromatin assembly ATPase regulated by its chromo- and DNA-binding domains." The Journal of biological chemistry **290**(1): 25-34.
Liu, X., T. A. McEachron, et al. (2014). "Histone H3 mutations in pediatric brain tumors." Cold Spring Harbor perspectives in biology **6**(4): a018689.
Liu, Z., D. R. Scannell, et al. (2011). "Control of embryonic stem cell lineage commitment by core promoter factor, TAF3." Cell **146**(5): 720-731.
Lobanenkov, V. V. and G. G. Gudvin (1989). "[CCCTC-binding protein: a new nuclear protein factor which interaction with 5'-flanking sequence of chicken c-myc oncogene correlates with repression of the gene]." Doklady Akademii nauk SSSR **309**(3): 741-745.
Lobanenkov, V. V., R. H. Nicolas, et al. (1990). "A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene." Oncogene **5**(12): 1743-1753.
Lorch, Y., J. W. LaPointe, et al. (1987). "Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones." Cell **49**(2): 203-210.
Lozzio, B. B. and C. B. Lozzio (1979). "Properties and usefulness of the original K-562 human myelogenous leukemia cell line." Leukemia research **3**(6): 363-370.
Lu, T., L. Aron, et al. (2014). "REST and stress resistance in ageing and Alzheimer's disease." Nature **507**(7493): 448-454.
Luger, K., A. W. Mader, et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.
Lupianez, D. G., K. Kraft, et al. (2015). "Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions." Cell **161**(5): 1012-1025.
Lutz, M., L. J. Burke, et al. (2000). "Transcriptional repression by the insulator protein CTCF involves histone deacetylases." Nucleic acids research **28**(8): 1707-1713.
Luzzati, V. and A. Nicolauieff (1963). "The Structure of Nucleohistones and Nucleoprotamines." Journal of molecular biology **7**: 142-163.
MacPherson, M. J. and P. D. Sadowski (2010). "The CTCF insulator protein forms an unusual DNA structure." BMC molecular biology **11**: 101.
Maeda, R. K. and F. Karch (2007). "Making connections: boundaries and insulators in Drosophila." Current opinion in genetics & development **17**(5): 394-399.
Maeshima, K., S. Hihara, et al. (2010). "Chromatin structure: does the 30-nm fibre exist in vivo?" Current opinion in cell biology **22**(3): 291-297.
Mallon, B. S., J. G. Chenoweth, et al. (2013). "StemCellDB: the human pluripotent stem cell database at the National Institutes of Health." Stem cell research **10**(1): 57-66.

Mallon, B. S., R. S. Hamilton, et al. (2014). "Comparison of the molecular profiles of human embryonic and induced pluripotent stem cells of isogenic origin." Stem cell research **12**(2): 376-386.
Maltby, V. E., B. J. Martin, et al. (2012). "Histone H3 lysine 36 methylation targets the Isw1b remodeling complex to chromatin." Molecular and cellular biology **32**(17): 3479-3485.
Marson, A., S. S. Levine, et al. (2008). "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells." Cell **134**(3): 521-533.
Martin, D., F. Allagnat, et al. (2012). "Specific silencing of the REST target genes in insulin-secreting cells uncovers their participation in beta cell survival." PloS one **7**(9): e45844.
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.
Matsuzaka, Y., K. Okamoto, et al. (2004). "Identification, expression analysis and polymorphism of a novel RLTPR gene encoding a RGD motif, tropomodulin domain and proline/leucine-rich regions." Gene **343**(2): 291-304.
Mavrich, T. N., I. P. Ioshikhes, et al. (2008). "A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome." Genome research **18**(7): 1073-1083.
Mavrich, T. N., C. Jiang, et al. (2008). "Nucleosome organization in the Drosophila genome." Nature **453**(7193): 358-362.
McClelland, S., C. M. Dube, et al. (2011). "Epileptogenesis after prolonged febrile seizures: mechanisms, biomarkers and therapeutic opportunities." Neuroscience letters **497**(3): 155-162.
McClelland, S., C. Flynn, et al. (2011). "Neuron-restrictive silencer factor-mediated hyperpolarization-activated cyclic nucleotide gated channelopathy in experimental temporal lobe epilepsy." Annals of neurology **70**(3): 454-464.
Meissner, A., T. S. Mikkelsen, et al. (2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells." Nature **454**(7205): 766-770.
Meshorer, E. and T. Misteli (2006). "Chromatin in pluripotent embryonic stem cells and differentiation." Nature reviews. Molecular cell biology **7**(7): 540-546.
Meshorer, E., D. Yellajoshula, et al. (2006). "Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells." Developmental cell **10**(1): 105-116.
Mirny, L. A. (2011). "The fractal globule as a model of chromatin architecture in the cell." Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology **19**(1): 37-51.
Misteli, T., A. Gunjan, et al. (2000). "Dynamic binding of histone H1 to chromatin in living cells." Nature **408**(6814): 877-881.
Mito, Y., J. G. Henikoff, et al. (2005). "Genome-scale profiling of histone H3.3 replacement patterns." Nature genetics **37**(10): 1090-1097.
Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-642.
Mizuguchi, G., X. Shen, et al. (2004). "ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex." Science **303**(5656): 343-348.
Moqtaderi, Z. and K. Struhl (2004). "Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes." Molecular and cellular biology **24**(10): 4118-4127.
Moqtaderi, Z., J. Wang, et al. (2010). "Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells." Nature structural & molecular biology **17**(5): 635-640.
Mori, N., C. Schoenherr, et al. (1992). "A common silencer element in the SCG10 and type II Na+ channel genes binds a factor present in nonneuronal cells but not in neuronal cells." Neuron **9**(1): 45-54.

Morrison, A. J. and X. Shen (2009). "Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes." Nature reviews. Molecular cell biology **10**(6): 373-384.
Moyle-Heyrman, G., T. Zaichuk, et al. (2013). "Chemical map of Schizosaccharomyces pombe reveals species-specific features in nucleosome positioning." Proceedings of the National Academy of Sciences of the United States of America **110**(50): 20158-20163.
Mu, W. and D. R. Burt (1999). "Transcriptional regulation of GABAA receptor gamma2 subunit gene." Brain research. Molecular brain research **67**(1): 137-147.
Nagano, T., Y. Lubling, et al. (2013). "Single-cell Hi-C reveals cell-to-cell variability in chromosome structure." Nature **502**(7469): 59-64.
Nakahashi, H., K. R. Kwon, et al. (2013). "A genome-wide map of CTCF multivalency redefines the CTCF code." Cell reports **3**(5): 1678-1689.
Neale, B. M., Y. Kou, et al. (2012). "Patterns and rates of exonic de novo mutations in autism spectrum disorders." Nature **485**(7397): 242-245.
Nedospasov, S. A. and G. P. Georgiev (1980). "Non-random cleavage of SV40 DNA in the compact minichromosome and free in solution by micrococcal nuclease." Biochemical and biophysical research communications **92**(2): 532-539.
Neigeborn, L. and M. Carlson (1984). "Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae." Genetics **108**(4): 845-858.
Nguyen, G. D., S. Gokhan, et al. (2014). "The role of H1 linker histone subtypes in preserving the fidelity of elaboration of mesendodermal and neuroectodermal lineages during embryonic development." PloS one **9**(5): e96858.
Nichols, J. and A. Smith (2009). "Naive and primed pluripotent states." Cell stem cell **4**(6): 487-492.
Nicol, J. W., G. A. Helt, et al. (2009). "The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets." Bioinformatics **25**(20): 2730-2731.
Nie, Y., X. Cheng, et al. (2014). "Nucleosome organization in the vicinity of transcription factor binding sites in the human genome." BMC genomics **15**: 493.
Nishino, Y., M. Eltsov, et al. (2012). "Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure." The EMBO journal **31**(7): 1644-1653.
Niwa, H., J. Miyazaki, et al. (2000). "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells." Nature genetics **24**(4): 372-376.
Noll, M. (1974). "Subunit structure of chromatin." Nature **251**(5472): 249-251.
Nozaki, T., K. Kaizu, et al. (2013). "Flexible and dynamic nucleosome fiber in living mammalian cells." Nucleus **4**(5): 349-356.
O'Roak, B. J., H. A. Stessman, et al. (2014). "Recurrent de novo mutations implicate novel genes underlying simplex autism risk." Nature communications **5**: 5595.
Obri, A., K. Ouararhni, et al. (2014). "ANP32E is a histone chaperone that removes H2A.Z from chromatin." Nature **505**(7485): 648-653.
Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-257.
Oler, A. J., R. K. Alla, et al. (2010). "Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors." Nature structural & molecular biology **17**(5): 620-628.
Olins, A. L., R. D. Carlson, et al. (1976). "Chromatin nu bodies: isolation, subfractionation and physical characterization." Nucleic acids research **3**(12): 3271-3291.
Olins, A. L. and D. E. Olins (1974). "Spheroid chromatin units (v bodies)." Science **183**(4122): 330-332.
Oliver, J. R., R. Kushwah, et al. (2012). "Multiple roles of the epithelium-specific ETS transcription factor, ESE-1, in development and disease." Laboratory investigation; a journal of technical methods and pathology **92**(3): 320-330.

Ong, C. T. and V. G. Corces (2014). "CTCF: an architectural protein bridging genome topology and function." Nature reviews. Genetics **15**(4): 234-246.
Ooi, L., N. D. Belyaev, et al. (2006). "BRG1 chromatin remodeling activity is required for efficient chromatin binding by repressor element 1-silencing transcription factor (REST) and facilitates REST-mediated repression." The Journal of biological chemistry **281**(51): 38974-38980.
Otto, S. J., S. R. McCorkle, et al. (2007). "A new binding motif for the transcriptional repressor REST uncovers large gene networks devoted to neuronal functions." The Journal of neuroscience : the official journal of the Society for Neuroscience **27**(25): 6729-6739.
Oudet, P., J. E. Germond, et al. (1978). "Nucleosome structure." Philosophical transactions of the Royal Society of London. Series B, Biological sciences **283**(997): 241-258.
Oudet, P., M. Gross-Bellard, et al. (1975). "Electron microscopic and biochemical evidence that chromatin structure is a repeating unit." Cell **4**(4): 281-300.
Ozsolak, F., J. S. Song, et al. (2007). "High-throughput mapping of the chromatin structure of human promoters." Nature biotechnology **25**(2): 244-248.
Packer, A. N., Y. Xing, et al. (2008). "The bifunctional microRNA miR-9/miR-9* regulates REST and CoREST and is downregulated in Huntington's disease." The Journal of neuroscience : the official journal of the Society for Neuroscience **28**(53): 14341-14346.
Padinhateeri, R. and J. F. Marko (2011). "Nucleosome positioning in a model of active chromatin remodeling enzymes." Proceedings of the National Academy of Sciences of the United States of America **108**(19): 7799-7803.
Palm, K., N. Belluardo, et al. (1998). "Neuronal expression of zinc finger transcription factor REST/NRSF/XBR gene." The Journal of neuroscience : the official journal of the Society for Neuroscience **18**(4): 1280-1296.
Pan, G., S. Tian, et al. (2007). "Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells." Cell stem cell **1**(3): 299-312.
Papamichos-Chronakis, M., S. Watanabe, et al. (2011). "Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity." Cell **144**(2): 200-213.
Paquette, A. J., S. E. Perez, et al. (2000). "Constitutive expression of the neuron-restrictive silencer factor (NRSF)/REST in differentiating neurons disrupts neuronal gene expression and causes axon pathfinding errors in vivo." Proceedings of the National Academy of Sciences of the United States of America **97**(22): 12318-12323.
Park, P. J. (2009). "ChIP-seq: advantages and challenges of a maturing technology." Nature reviews. Genetics **10**(10): 669-680.
Park, S. Y., J. B. Kim, et al. (2007). "REST is a key regulator in brain-specific homeobox gene expression during neuronal differentiation." Journal of neurochemistry **103**(6): 2565-2574.
Paweletz, N. (2001). "Walther Flemming: pioneer of mitosis research." Nature reviews. Molecular cell biology **2**(1): 72-75.
Pearson, A. G., M. A. Curtis, et al. (2005). "Activating transcription factor 2 expression in the adult human brain: association with both neurodegeneration and neurogenesis." Neuroscience **133**(2): 437-451.
Pekowska, A., T. Benoukraf, et al. (2011). "H3K4 tri-methylation provides an epigenetic signature of active enhancers." The EMBO journal **30**(20): 4198-4210.
Perkowski, J. J. and G. G. Murphy (2011). "Deletion of the mouse homolog of KCNAB2, a gene linked to monosomy 1p36, results in associative memory impairments and amygdala hyperexcitability." The Journal of neuroscience : the official journal of the Society for Neuroscience **31**(1): 46-54.
PGC (2013). "Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis." Lancet **381**(9875): 1371-1379.

Phillips-Cremins, J. E. and V. G. Corces (2013). "Chromatin insulators: linking genome organization to cellular function." Molecular cell **50**(4): 461-474.
Phillips-Cremins, J. E., M. E. Sauria, et al. (2013). "Architectural protein subclasses shape 3D organization of genomes during lineage commitment." Cell **153**(6): 1281-1295.
Phillips, J. E. and V. G. Corces (2009). "CTCF: master weaver of the genome." Cell **137**(7): 1194-1211.
Pischedda, F., J. Szczurkowska, et al. (2014). "A cell surface biotinylation assay to reveal membrane-associated neuronal cues: Negr1 regulates dendritic arborization." Molecular & cellular proteomics : MCP **13**(3): 733-748.
Plasschaert, R. N., S. Vigneau, et al. (2014). "CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation." Nucleic acids research **42**(2): 774-789.
Platt, J. L. (2013). "A study of the Chd family of ATP-dependent chromatin remodellers in Dictyostelium discoideum." http://orca.cf.ac.uk/53993/.
Platt, J. L., B. J. Rogers, et al. (2013). "Different CHD chromatin remodelers are required for expression of distinct gene sets and specific stages during development of Dictyostelium discoideum." Development **140**(24): 4926-4936.
Pointner, J., J. Persson, et al. (2012). "CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe." The EMBO journal **31**(23): 4388-4403.
Pope, B. D., T. Ryba, et al. (2014). "Topologically associating domains are stable units of replication-timing regulation." Nature **515**(7527): 402-405.
Pray-Grant, M. G., J. A. Daniel, et al. (2005). "Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation." Nature **433**(7024): 434-438.
Raab, J. R., J. Chiu, et al. (2012). "Human tRNA genes function as chromatin insulators." The EMBO journal **31**(2): 330-350.
Ramachandran, A., M. Omar, et al. (2003). "Linker histone H1 modulates nucleosome remodeling by human SWI/SNF." The Journal of biological chemistry **278**(49): 48590-48601.
Rao, S. S., M. H. Huntley, et al. (2014). "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." Cell **159**(7): 1665-1680.
Razin, A. (1998). "CpG methylation, chromatin structure and gene silencing-a three-way connection." The EMBO journal **17**(17): 4905-4908.
Read, C. M. and C. Crane-Robinson (1985). "The structure of sub-nucleosomal particles. The octameric (H3/H4)4--125-base-pair-DNA complex." European journal of biochemistry / FEBS **152**(1): 143-150.
Reimold, A. M., M. J. Grusby, et al. (1996). "Chondrodysplasia and neurological abnormalities in ATF-2-deficient mice." Nature **379**(6562): 262-265.
Ren, B., F. Robert, et al. (2000). "Genome-wide location and function of DNA binding proteins." Science **290**(5500): 2306-2309.
Reubinoff, B. E., M. F. Pera, et al. (2000). "Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro." Nature biotechnology **18**(4): 399-404.
Rhee, H. S., A. R. Bataille, et al. (2014). "Subnucleosomal structures and nucleosome asymmetry across a genome." Cell **159**(6): 1377-1388.
Rhee, H. S. and B. F. Pugh (2011). "Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution." Cell **147**(6): 1408-1419.
Rockowitz, S., W. H. Lien, et al. (2014). "Comparison of REST cistromes across human cell types reveals common and context-specific functions." PLoS computational biology **10**(6): e1003671.
Rockowitz, S. and D. Zheng (2015). "Significant expansion of the REST/NRSF cistrome in human versus mouse embryonic stem cells: potential implications for neural development." Nucleic acids research **43**(12): 5730-5743.
Roopra, A., R. Dingledine, et al. (2012). "Epigenetics and epilepsy." Epilepsia **53 Suppl 9**: 2-10.

Roopra, A., R. Qazi, et al. (2004). "Localized domains of G9a-mediated histone methylation are required for silencing of neuronal genes." Molecular cell **14**(6): 727-738.
Rosenberg, M., L. Hui, et al. (1997). "Characterization of short tandem repeats from thirty-one human telomeres." Genome research **7**(9): 917-923.
Rubio, E. D., D. J. Reiss, et al. (2008). "CTCF physically links cohesin to chromatin." Proceedings of the National Academy of Sciences of the United States of America **105**(24): 8309-8314.
Ruhl, D. D., J. Jin, et al. (2006). "Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes." Biochemistry **45**(17): 5671-5677.
Rydberg, B., W. R. Holley, et al. (1998). "Chromatin conformation in living cells: support for a zig-zag model of the 30 nm chromatin fiber." Journal of molecular biology **284**(1): 71-84.
Sajan, S. A. and R. D. Hawkins (2012). "Methods for identifying higher-order chromatin structure." Annual review of genomics and human genetics **13**: 59-82.
Scheffer, M. P., M. Eltsov, et al. (2011). "Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei." Proceedings of the National Academy of Sciences of the United States of America **108**(41): 16992-16997.
Schoenherr, C. J. and D. J. Anderson (1995). "The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes." Science **267**(5202): 1360-1363.
Schoenherr, C. J., A. J. Paquette, et al. (1996). "Identification of potential target genes for the neuron-restrictive silencer factor." Proceedings of the National Academy of Sciences of the United States of America **93**(18): 9881-9886.
Scholl, T., M. B. Stevens, et al. (1996). "A zinc finger protein that represses transcription of the human MHC class II gene, DPA." Journal of immunology **156**(4): 1448-1457.
Schones, D. E., K. Cui, et al. (2008). "Dynamic regulation of nucleosome positioning in the human genome." Cell **132**(5): 887-898.
Schulze, J. M., J. Jackson, et al. (2009). "Linking cell cycle to histone modifications: SBF and H2B monoubiquitination machinery and cell-cycle regulation of H3K79 dimethylation." Molecular cell **35**(5): 626-641.
Schwanbeck, R., H. Xiao, et al. (2004). "Spatial contacts and nucleosome step movements induced by the NURF chromatin remodeling complex." The Journal of biological chemistry **279**(38): 39933-39941.
Schwartz, B. E. and K. Ahmad (2005). "Transcriptional activation triggers deposition and removal of the histone variant H3.3." Genes & development **19**(7): 804-814.
Schwartzentruber, J., A. Korshunov, et al. (2012). "Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma." Nature **482**(7384): 226-231.
Segal, E., Y. Fondufe-Mittendorf, et al. (2006). "A genomic code for nucleosome positioning." Nature **442**(7104): 772-778.
Segal, E. and J. Widom (2009). "Poly(dA:dT) tracts: major determinants of nucleosome organization." Current opinion in structural biology **19**(1): 65-71.
Shah, A., A. Oldenburg, et al. (2014). "A hyper-dynamic nature of bivalent promoter states underlies coordinated developmental gene expression modules." BMC genomics **15**: 1186.
Shatnawi, A., J. D. Norris, et al. (2014). "ELF3 is a repressor of androgen receptor action in prostate cancer cells." Oncogene **33**(7): 862-871.
Shaytan, A. K., D. Landsman, et al. (2015). "Nucleosome adaptability conferred by sequence and structural variations in histone H2A-H2B dimers." Current opinion in structural biology **32C**: 48-57.
Sheik Mohamed, J., P. M. Gaughwin, et al. (2010). "Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells." RNA **16**(2): 324-337.

Shen, X., G. Mizuguchi, et al. (2000). "A chromatin remodelling complex involved in transcription and DNA processing." Nature **406**(6795): 541-544.
Shi, Y., F. Lan, et al. (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." Cell **119**(7): 941-953.
Siggens, L., L. Cordeddu, et al. (2015). "Transcription-coupled recruitment of human CHD1 and CHD2 influences chromatin accessibility and histone H3 and H3.3 occupancy at active chromatin regions." Epigenetics & chromatin **8**(1): 4.
Simms, T. A., S. L. Dugas, et al. (2008). "TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in Saccharomyces cerevisiae." Eukaryotic cell **7**(12): 2078-2086.
Simpson, R. T. (1978). "Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones." Biochemistry **17**(25): 5524-5531.
Sims, R. J., 3rd, C. F. Chen, et al. (2005). "Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains." The Journal of biological chemistry **280**(51): 41789-41792.
Singh, M., S. K. Bag, et al. (2015). "Global nucleosome positioning regulates salicylic acid mediated transcription in Arabidopsis thaliana." BMC plant biology **15**: 13.
Singh, S. K., M. N. Kagalwala, et al. (2008). "REST maintains self-renewal and pluripotency of embryonic stem cells." Nature **453**(7192): 223-227.
Skene, P. J., A. E. Hernandez, et al. (2014). "The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1." eLife **3**: e02042.
Smith, A. G., J. K. Heath, et al. (1988). "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides." Nature **336**(6200): 688-690.
Smith, J. R., L. Vallier, et al. (2008). "Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm." Developmental biology **313**(1): 107-117.
Smith, Z. D., M. M. Chan, et al. (2012). "A unique regulatory phase of DNA methylation in the early mammalian embryo." Nature **484**(7394): 339-344.
Smith, Z. D. and A. Meissner (2013). "DNA methylation: roles in mammalian development." Nature reviews. Genetics **14**(3): 204-220.
Song, F., P. Chen, et al. (2014). "Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units." Science **344**(6182): 376-380.
Splinter, E., H. Heath, et al. (2006). "CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus." Genes & development **20**(17): 2349-2354.
Stappert, L., L. Borghese, et al. (2013). "MicroRNA-based promotion of human neuronal differentiation and subtype specification." PloS one **8**(3): e59011.
Stern, M., R. Jensen, et al. (1984). "Five SWI genes are required for expression of the HO gene in yeast." Journal of molecular biology **178**(4): 853-868.
Stewart, A. F., A. Reik, et al. (1991). "A simpler and better method to cleave chromatin with DNase 1 for hypersensitive site analyses." Nucleic acids research **19**(11): 3157.
Stockdale, C., A. Flaus, et al. (2006). "Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes." The Journal of biological chemistry **281**(24): 16279-16288.
Stormo, G. D., T. D. Schneider, et al. (1982). "Use of the 'Perceptron' algorithm to distinguish translational initiation sites in E. coli." Nucleic acids research **10**(9): 2997-3011.
Stoykova, A., D. Treichel, et al. (2000). "Pax6 modulates the dorsoventral patterning of the mammalian telencephalon." The Journal of neuroscience : the official journal of the Society for Neuroscience **20**(21): 8042-8050.
Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-45.
Struhl, K. and E. Segal (2013). "Determinants of nucleosome positioning." Nature structural & molecular biology **20**(3): 267-273.

Sun, Y. M., M. Cooper, et al. (2008). "Rest-mediated regulation of extracellular matrix is crucial for neural development." PloS one **3**(11): e3656.
Sun, Y. M., D. J. Greenway, et al. (2005). "Distinct profiles of REST interactions with its target genes at different stages of neuronal development." Molecular biology of the cell **16**(12): 5630-5638.
Suto, R. K., M. J. Clarkson, et al. (2000). "Crystal structure of a nucleosome core particle containing the variant histone H2A.Z." Nature structural biology **7**(12): 1121-1124.
Tachibana, M., K. Sugimoto, et al. (2001). "Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3." The Journal of biological chemistry **276**(27): 25309-25317.
Tai, H. H., M. Geisterfer, et al. (2003). "CHD1 associates with NCoR and histone deacetylase as well as with RNA splicing proteins." Biochemical and biophysical research communications **308**(1): 170-176.
Takahashi, K., K. Tanabe, et al. (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-872.
Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.
Tazi, J. and A. Bird (1990). "Alternative chromatin structure at CpG islands." Cell **60**(6): 909-920.
Teif, V. B., D. A. Beshnova, et al. (2014). "Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development." Genome research **24**(8): 1285-1295.
Teif, V. B., Y. Vainshtein, et al. (2012). "Genome-wide nucleosome positioning during embryonic stem cell development." Nature structural & molecular biology **19**(11): 1185-1192.
Tesar, P. J., J. G. Chenoweth, et al. (2007). "New cell lines from mouse epiblast share defining features with human embryonic stem cells." Nature **448**(7150): 196-199.
THDCRG (1993). "A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group." Cell **72**(6): 971-983.
Thoma, F. and T. Koller (1977). "Influence of histone H1 on chromatin structure." Cell **12**(1): 101-107.
Thoma, F., T. Koller, et al. (1979). "Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin." The Journal of cell biology **83**(2 Pt 1): 403-427.
Thomas, J. O. and P. J. Butler (1977). "Characterization of the octamer of histones free in solution." Journal of molecular biology **116**(4): 769-781.
Thomas, J. O. and R. D. Kornberg (1975). "An octamer of histones in chromatin and free in solution." Proceedings of the National Academy of Sciences of the United States of America **72**(7): 2626-2630.
Thompson, B. A., V. Tremblay, et al. (2008). "CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes." Molecular and cellular biology **28**(12): 3894-3904.
Thompson, P. M., T. Gotoh, et al. (2003). "CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system." Oncogene **22**(7): 1002-1011.
Thomson, J. A., J. Itskovitz-Eldor, et al. (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-1147.
Tilgner, H., C. Nikolaou, et al. (2009). "Nucleosome positioning as a determinant of exon recognition." Nature structural & molecular biology **16**(9): 996-1001.
Tillo, D., N. Kaplan, et al. (2010). "High nucleosome occupancy is encoded at human regulatory sequences." PloS one **5**(2): e9129.

Tirosh, I. and N. Barkai (2008). "Two strategies for gene regulation by promoter nucleosomes." Genome research **18**(7): 1084-1091.
Tolstorukov, M. Y., P. V. Kharchenko, et al. (2009). "Comparative analysis of H2A.Z nucleosome organization in the human and yeast genomes." Genome research **19**(6): 967-977.
Toshiyuki, N. and M. Ichiro (2004). "Molecular mechanisms regulating cell type specific expression of BMP/RA Inducible Neural-specific Protein-1 that suppresses cell cycle progression: roles of NRSF/REST and DNA methylation." Brain research. Molecular brain research **125**(1-2): 47-59.
Tosi, A., C. Haas, et al. (2013). "Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex." Cell **154**(6): 1207-1219.
Trynka, G. and S. Raychaudhuri (2013). "Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases." Current opinion in genetics & development **23**(6): 635-641.
Tsai, M. C., O. Manor, et al. (2010). "Long noncoding RNA as modular scaffold of histone modification complexes." Science **329**(5992): 689-693.
Tse, C., T. Sera, et al. (1998). "Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III." Molecular and cellular biology **18**(8): 4629-4638.
Tsukiyama, T., J. Palmer, et al. (1999). "Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in Saccharomyces cerevisiae." Genes & development **13**(6): 686-697.
Tucker, P. W., E. E. Hazen, Jr., et al. (1978). "Staphylococcal nuclease reviewed: a prototypic study in contemporary enzymology. I. Isolation; physical and enzymatic properties." Molecular and cellular biochemistry **22**(2-3): 67-77.
Udugama, M., A. Sabri, et al. (2011). "The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor." Molecular and cellular biology **31**(4): 662-673.
Unnikrishnan, A., P. R. Gafken, et al. (2010). "Dynamic changes in histone acetylation regulate origins of DNA replication." Nature structural & molecular biology **17**(4): 430-437.
Valenzuela, L., N. Dhillon, et al. (2009). "Transcription independent insulation at TFIIIC-dependent insulators." Genetics **183**(1): 131-148.
Vallier, L., M. Alexander, et al. (2005). "Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells." Journal of cell science **118**(Pt 19): 4495-4509.
Valouev, A., J. Ichikawa, et al. (2008). "A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning." Genome research **18**(7): 1051-1063.
Valouev, A., S. M. Johnson, et al. (2011). "Determinants of nucleosome organization in primary human cells." Nature **474**(7352): 516-520.
Van Bortle, K. and V. G. Corces (2012). "tDNA insulators and the emerging role of TFIIIC in genome organization." Transcription **3**(6): 277-284.
Verdaasdonk, J. S. and K. Bloom (2011). "Centromeres: unique chromatin structures that drive chromosome segregation." Nature reviews. Molecular cell biology **12**(5): 320-332.
Vestin, A. and A. A. Mills (2013). "The tumor suppressor Chd5 is induced during neuronal differentiation in the developing mouse brain." Gene expression patterns : GEP **13**(8): 482-489.
Vietri Rudan, M., C. Barrington, et al. (2015). "Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture." Cell reports **10**(8): 1297-1309.
Vissers, L. E., C. M. van Ravenswaaij, et al. (2004). "Mutations in a new member of the chromodomain gene family cause CHARGE syndrome." Nature genetics **36**(9): 955-957.
Vogelstein, B., N. Papadopoulos, et al. (2013). "Cancer genome landscapes." Science **339**(6127): 1546-1558.

Voigt, P., W. W. Tee, et al. (2013). "A double take on bivalent promoters." Genes & development **27**(12): 1318-1338.
Wang, G. G., C. D. Allis, et al. (2007). "Chromatin remodeling and cancer, Part I: Covalent histone modifications." Trends in molecular medicine **13**(9): 363-372.
Wang, G. G., C. D. Allis, et al. (2007). "Chromatin remodeling and cancer, Part II: ATP-dependent chromatin remodeling." Trends in molecular medicine **13**(9): 373-380.
Wang, H., M. T. Maurano, et al. (2012). "Widespread plasticity in CTCF occupancy linked to DNA methylation." Genome research **22**(9): 1680-1688.
Wang, J., J. Zhuang, et al. (2012). "Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors." Genome research **22**(9): 1798-1812.
Wang, J., J. Zhuang, et al. (2013). "Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium." Nucleic acids research **41**(Database issue): D171-176.
Wang, J. C. (1979). "Helical repeat of DNA in solution." Proceedings of the National Academy of Sciences of the United States of America **76**(1): 200-203.
Wang, W., J. Cote, et al. (1996). "Purification and biochemical heterogeneity of the mammalian SWI-SNF complex." The EMBO journal **15**(19): 5370-5382.
Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-562.
Watson, L. A., X. Wang, et al. (2014). "Dual effect of CTCF loss on neuroprogenitor differentiation and survival." The Journal of neuroscience : the official journal of the Society for Neuroscience **34**(8): 2860-2870.
Weiner, A., A. Hughes, et al. (2010). "High-resolution nucleosome mapping reveals transcription-dependent promoter packaging." Genome research **20**(1): 90-100.
Wells, T., K. Rough, et al. (2011). "Transcription Mapping of Embryonic Rat Brain Reveals EGR-1 Induction in SOX2 Neural Progenitor Cells." Frontiers in molecular neuroscience **4**: 6.
West, J. A., A. Cook, et al. (2014). "Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming." Nature communications **5**: 4719.
Westbrook, T. F., E. S. Martin, et al. (2005). "A genetic screen for candidate tumor suppressors identifies REST." Cell **121**(6): 837-848.
Whitehouse, I., A. Flaus, et al. (1999). "Nucleosome mobilization catalysed by the yeast SWI/SNF complex." Nature **400**(6746): 784-787.
Whitehouse, I. and T. Tsukiyama (2006). "Antagonistic forces that position nucleosomes in vivo." Nature structural & molecular biology **13**(7): 633-640.
Whitlock, J. P., Jr. and R. T. Simpson (1976). "Removal of histone H1 exposes a fifty base pair DNA segment between nucleosomes." Biochemistry **15**(15): 3307-3314.
Willemsen, R., J. Levenga, et al. (2011). "CGG repeat in the FMR1 gene: size matters." Clinical genetics **80**(3): 214-225.
Woodcock, C. L., L. L. Frado, et al. (1984). "The higher-order structure of chromatin: evidence for a helical ribbon arrangement." The Journal of cell biology **99**(1 Pt 1): 42-52.
Wu, J. and X. Xie (2006). "Comparative sequence analysis reveals an intricate network among REST, CREB and miRNA in mediating neuronal gene expression." Genome biology **7**(9): R85.
Wu, J. Q., L. Habegger, et al. (2010). "Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing." Proceedings of the National Academy of Sciences of the United States of America **107**(11): 5254-5259.
Wu, Y., W. Zhang, et al. (2014). "Genome-wide nucleosome positioning is orchestrated by genomic regions associated with DNase I hypersensitivity in rice." PLoS genetics **10**(5): e1004378.

Xie, X., T. S. Mikkelsen, et al. (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.
Xu, C., M. S. Inokuma, et al. (2001). "Feeder-free growth of undifferentiated human embryonic stem cells." Nature biotechnology **19**(10): 971-974.
Xue, Y., J. C. Canman, et al. (2000). "The human SWI/SNF-B chromatin-remodeling complex is related to yeast rsc and localizes at kinetochores of mitotic chromosomes." Proceedings of the National Academy of Sciences of the United States of America **97**(24): 13015-13020.
Yao, Y. L., W. M. Yang, et al. (2001). "Regulation of transcription factor YY1 by acetylation and deacetylation." Molecular and cellular biology **21**(17): 5979-5991.
Yen, K., V. Vinayachandran, et al. (2013). "SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes." Cell **154**(6): 1246-1256.
Yigit, E., Q. Zhang, et al. (2013). "High-resolution nucleosome mapping of targeted regions using BAC-based enrichment." Nucleic acids research **41**(7): e87.
Ying, Q. L., J. Nichols, et al. (2003). "BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3." Cell **115**(3): 281-292.
Ying, Q. L., J. Wray, et al. (2008). "The ground state of embryonic stem cell self-renewal." Nature **453**(7194): 519-523.
Yu, J., M. A. Vodyanik, et al. (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-1920.
Yuan, G. C., Y. J. Liu, et al. (2005). "Genome-scale identification of nucleosome positions in S. cerevisiae." Science **309**(5734): 626-630.
Zaugg, J. B. and N. M. Luscombe (2012). "A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast." Genome research **22**(1): 84-94.
Zhang, C., Z. Xuan, et al. (2006). "A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome." Nucleic acids research **34**(8): 2238-2246.
Zhang, G., H. Huang, et al. (2015). "N(6)-methyladenine DNA modification in Drosophila." Cell **161**(4): 893-906.
Zhang, H., W. Jiao, et al. (2013). "Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming." Cell stem cell **13**(1): 30-35.
Zhang, P., M. J. Pazin, et al. (2008). "Nontelomeric TRF2-REST interaction modulates neuronal gene silencing and fate of tumor and stem cells." Current biology : CB **18**(19): 1489-1494.
Zhang, T., W. Zhang, et al. (2015). "Genome-Wide Nucleosome Occupancy and Positioning and Their Impact on Gene Expression and Evolution in Plants." Plant physiology **168**(4): 1406-1416.
Zhang, X., C. T. Huang, et al. (2010). "Pax6 is a human neuroectoderm cell fate determinant." Cell stem cell **7**(1): 90-100.
Zhang, Z. and B. F. Pugh (2011). "High-resolution genome-wide mapping of the primary structure of chromatin." Cell **144**(2): 175-186.
Zhao, H., Y. Xing, et al. (2015). "GAA triplet-repeats cause nucleosome depletion in the human genome." Genomics.
Zhao, X. D., X. Han, et al. (2007). "Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells." Cell stem cell **1**(3): 286-298.
Ziller, M. J., H. Gu, et al. (2013). "Charting a dynamic DNA methylation landscape of the human genome." Nature **500**(7463): 477-481.
Zlatanova, J. and P. Caiafa (2009). "CTCF and its protein partners: divide and rule?" Journal of cell science **122**(Pt 9): 1275-1284.

Zuccato, C., N. Belyaev, et al. (2007). "Widespread disruption of repressor element-1 silencing transcription factor/neuron-restrictive silencer factor occupancy at its target genes in Huntington's disease." The Journal of neuroscience : the official journal of the Society for Neuroscience **27**(26): 6972-6983.

Zuccato, C., M. Tartari, et al. (2003). "Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes." Nature genetics **35**(1): 76-83.

Appendix

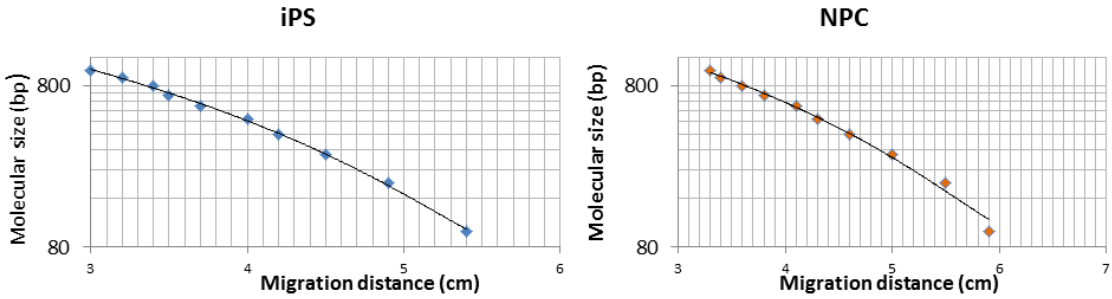


Fig A.1. The nucleosome repeat length of bulk human chromatin. The nucleosome repeat length of bulk chromatin was determined by plotting a calibration curve of the log of the molecular size in base pairs against the distance migrated by the DNA in the gel and then determining the size of the furthest migrating band in the MNase ladder (the lane labelled 'total').

	iPS total reads	NPC total reads
Chr1	420776083	364285425
Chr2	441804789	392419117
Chr3	366564871	314634891
Chr4	366702163	316969669
Chr5	340552189	290172689
Chr6	317809861	273489877
Chr7	275763613	244278269
Chr8	254998311	229631515
Chr9	208618097	186833925
Chr10	231362991	218011409
Chr11	225935261	204772369
Chr12	231607777	203298289
Chr13	180987539	154538319
Chr14	152985929	136978517
Chr15	130978709	121739567
Chr16	111163801	118742655
Chr17	111798077	110612079
Chr18	136758805	122692857
Chr19	65330679	70790417
Chr20	88772815	86239715
Chr21	53849863	56054629
Chr22	42337965	43664621
ChrX	244412315	228311355
ChrY	9243763	14560079
	5011116266	
Total aligned reads for whole genome	5011116266	4503722254
Total aligned reads for autosomal genome	4757460188	4260850820
Ratio of autosomal aligned reads iPS/NPC	1.116551691	
Total no of aligned paired reads for whole genome	2505558133	2251861127

Table A.1 Sequence reads derived from MNase resistant DNA in iPS cell and NPC cultures show broadly similar counts for each chromosome.

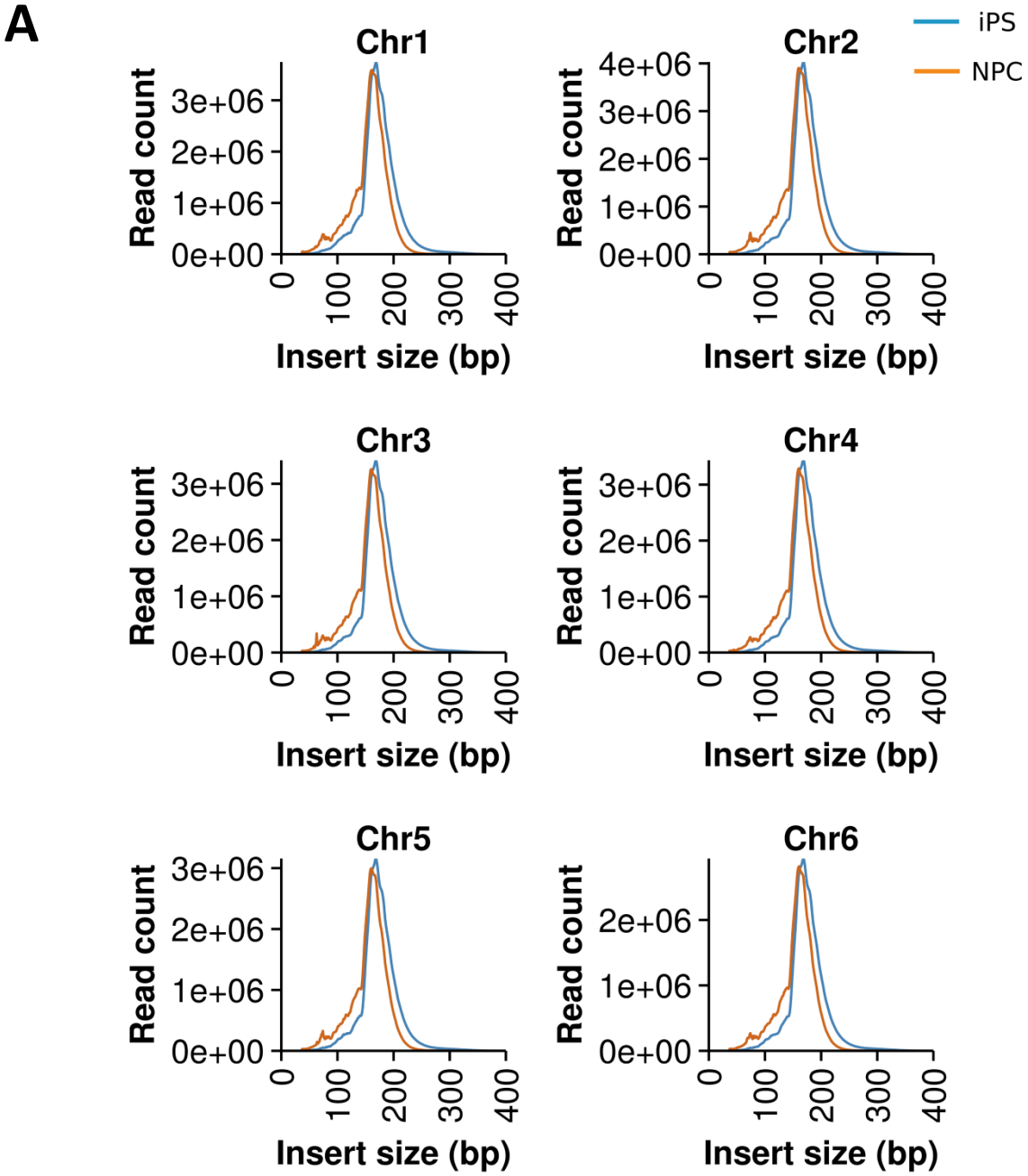


Fig A.2 The number and size of the aligned paired-end sequencing reads for individual chromosomes was determined single base pair resolution. Frequency distributions for chromosomes 1-6 are shown for both iPS (blue) and NPC (orange) cells.

B

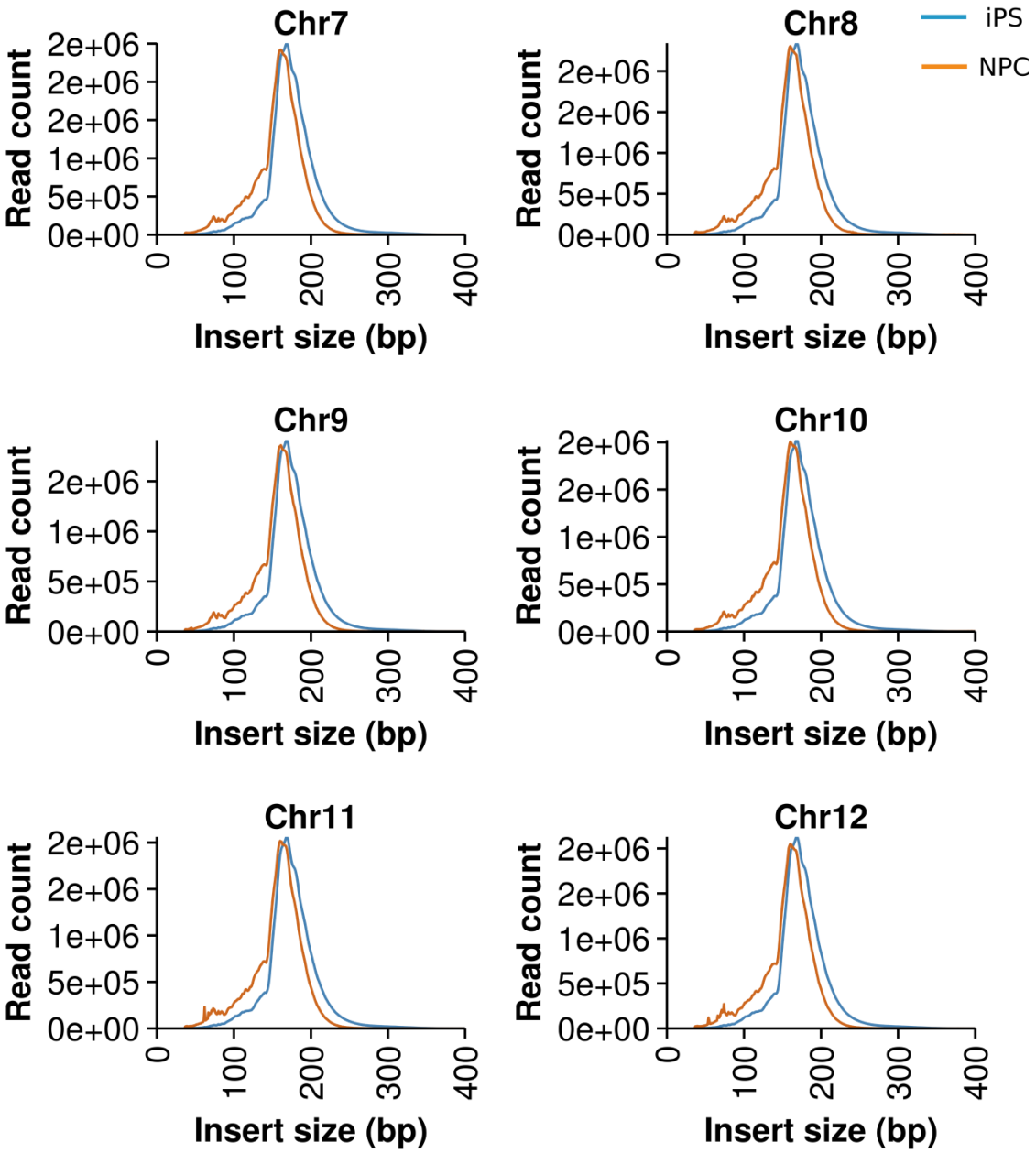


Fig A.2 The number and size of the aligned paired-end sequencing reads for individual chromosomes was determined single base pair resolution. Frequency distributions for chromosomes 7-12 are shown for both iPS (blue) and NPC (orange) cells.

C

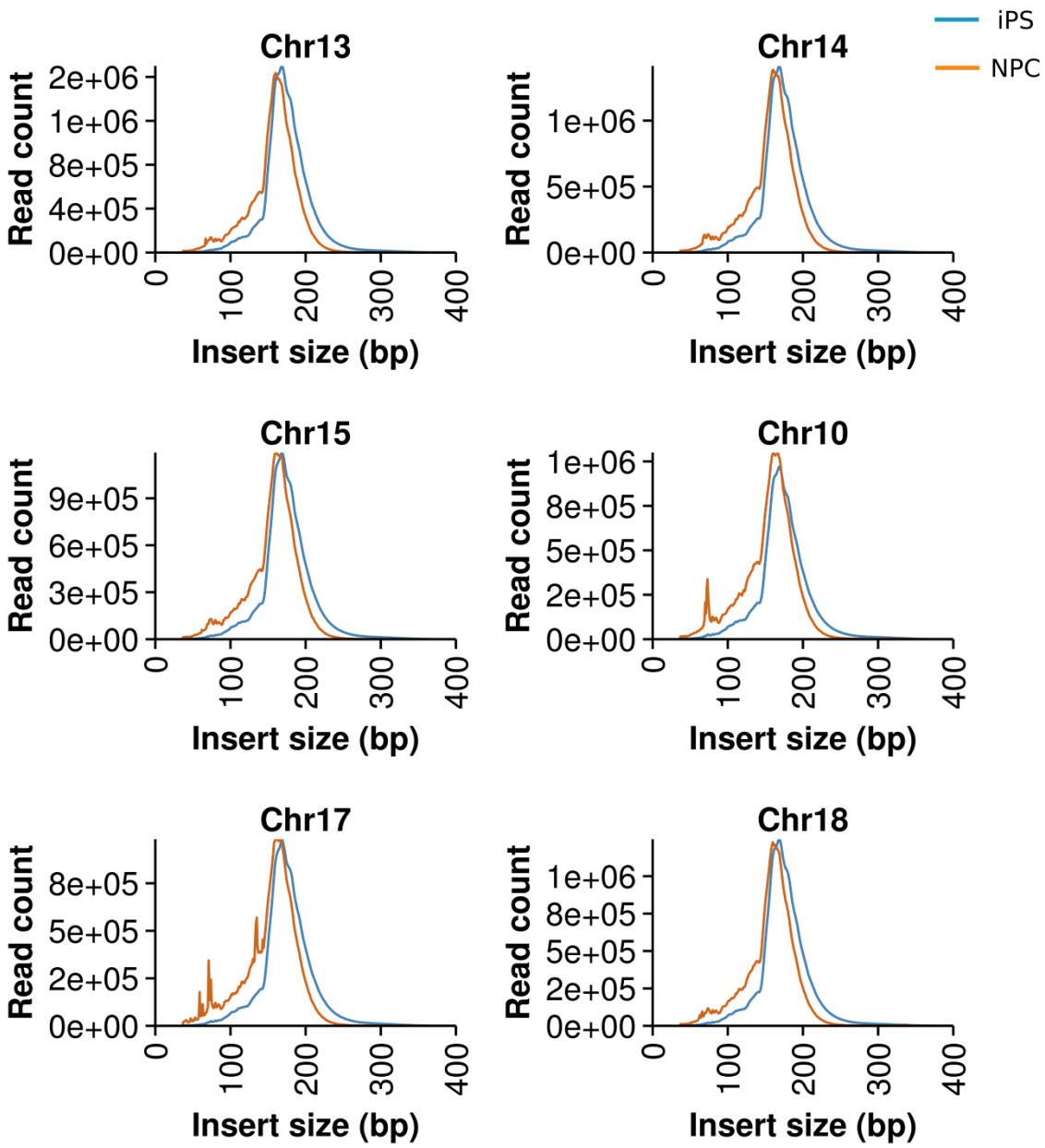


Fig A.2 The number and size of the aligned paired-end sequencing reads for individual chromosomes was determined single base pair resolution. Frequency distributions for chromosomes 13-18 are shown for both iPS (blue) and NPC (orange) cells.

D

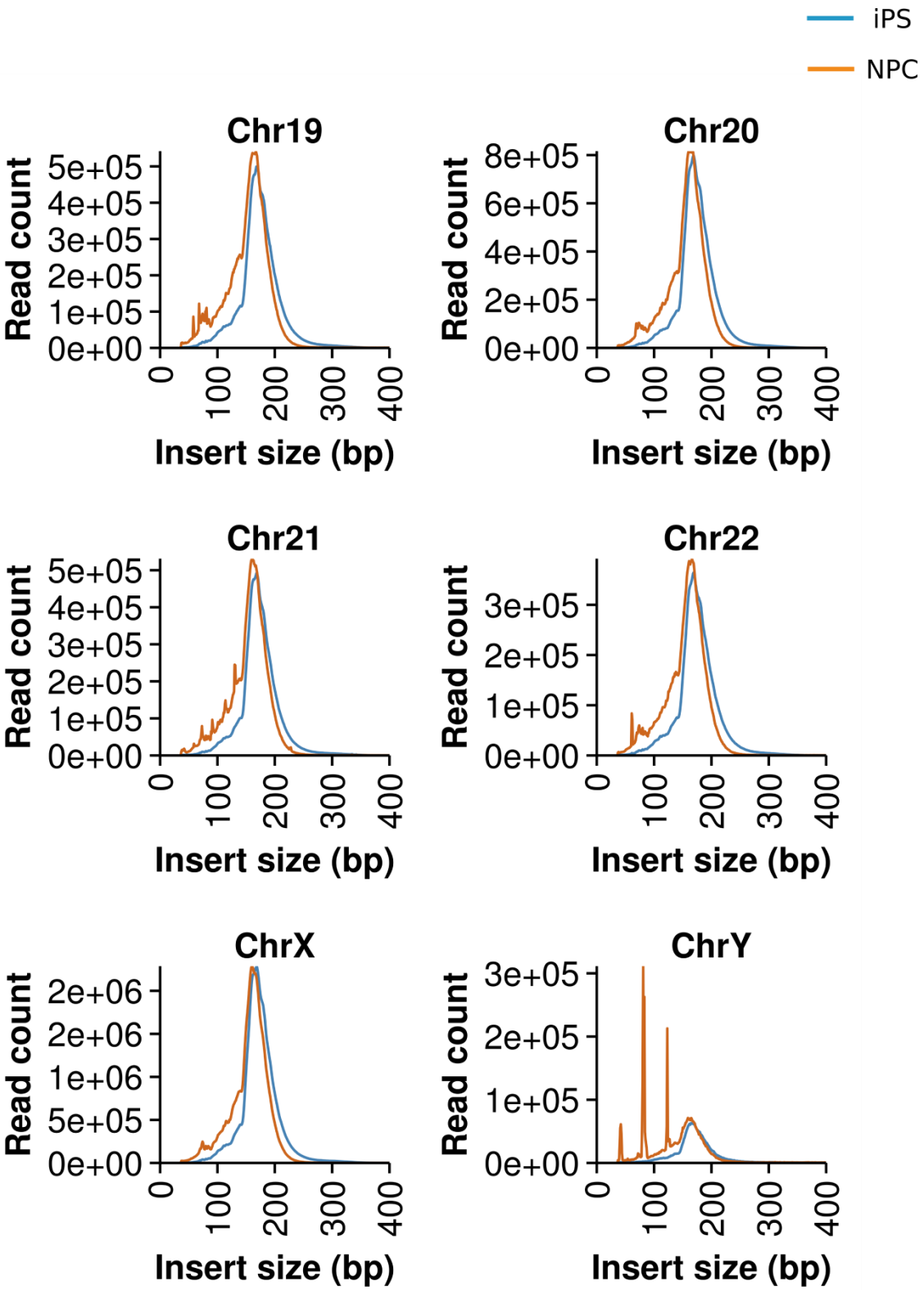


Fig A.2. The number and size of the aligned paired-end sequencing reads for individual chromosomes was determined single base pair resolution. Frequency distributions for chromosomes 19-Y are shown for both iPS (blue) and NPC (orange) cells.

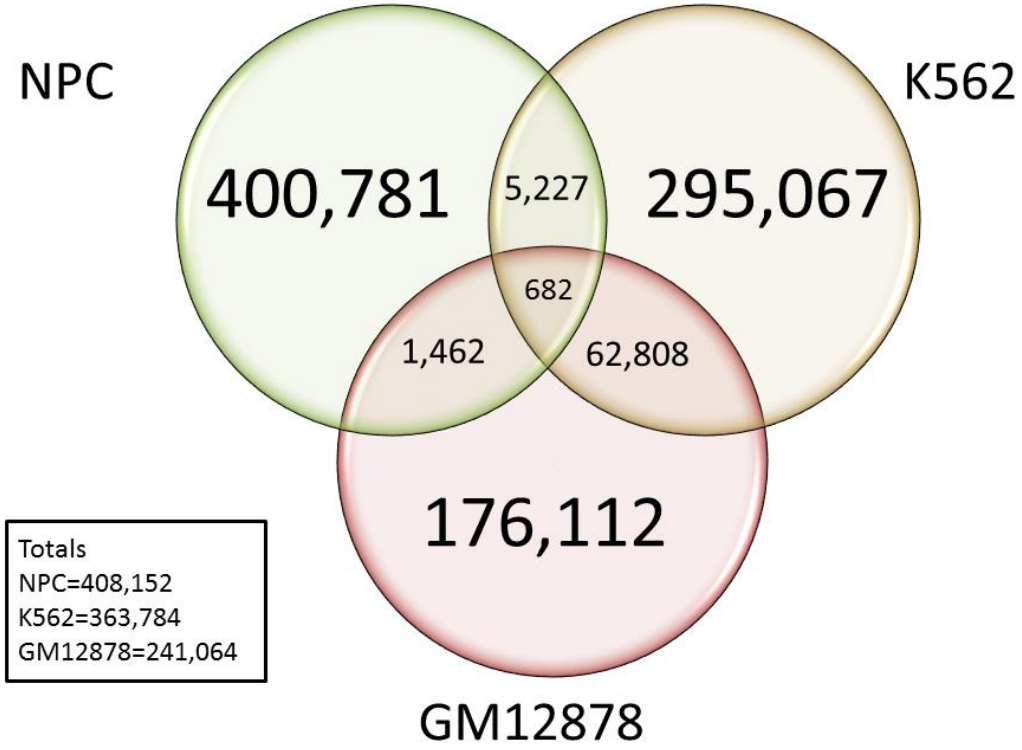


Figure A.3 A comparison of the positions of chromatin particles mapped by MNase-seq in the nucleosomal chromatin particle size class from three differentiated cell types NPC, K562 and GM12878. Chromatin particles, defined as peaks in the genomic distribution of MNase-seq sequence read mid-points, were given explicit genome positions using the heuristic peak-finding algorithm (PeakFinder). The Venn diagram shows the number of particle positions determined for the nucleosomal size class (138-161 bp) of sequence read in NPC cells and the frequency of the overlap of the positions within +/- 10 bp for K562 and GM12878 cells.

GENE NAME	REFERENCE
ACTA1	(Schoenherr, Paquette et al. 1996)
ADAM23	(Sun, Greenway et al. 2005) (Otto, McCorkle et al. 2007)
APBA2	(Sun, Greenway et al. 2005)
ARC	(Sun, Greenway et al. 2005)
ASIC2	(Sun, Greenway et al. 2005)
ATP2B2	(Sun, Greenway et al. 2005)
BDNF	(Schoenherr and Anderson 1995)
BRINP1	(Toshiyuki and Ichiro 2004)
BSX	(Park, Kim et al. 2007)
CACNA1A	(Johnson, Gamblin et al. 2006)
CACNA2D3	(Sun, Greenway et al. 2005)
CACNG2	(Bruce, Donaldson et al. 2004)
CALB1	(Ballas, Grunseich et al. 2005)
CDH4	(Sun, Greenway et al. 2005)
CDK5R2	(Sun, Greenway et al. 2005; Martin, Allagnat et al. 2012)
CHRNB2	(Bessis, Salmon et al. 1995)
CREBBP	(Sun, Greenway et al. 2005)
CRH	(Bruce, Donaldson et al. 2004)
CTNND2	(Sun, Cooper et al. 2008)
CYP11B1	(Schoenherr, Paquette et al. 1996)
DDR2	(Sun, Greenway et al. 2005)
EXTL3	(Sun, Greenway et al. 2005)
GABRG2	(Mu and Burt 1999)
GJD2	(Martin, Allagnat et al. 2012)
GLRA1	(Schoenherr, Paquette et al. 1996)
GRIA2	(Brene, Messer et al. 2000)
GRIN1	(Schoenherr, Paquette et al. 1996)
GRIN2A	(Bruce, Donaldson et al. 2004)
HTR1A	(Lemonde, Rogaeva et al. 2004)
KCNH1	(Sun, Greenway et al. 2005)
KIRREL3	(Sun, Greenway et al. 2005)
L1CAM	(Kallunki, Edelman et al. 1997)
LHX3	(Bruce, Donaldson et al. 2004)
LSAMP	(Zhang, Pazin et al. 2008)
MAPK8IP1	(Martin, Allagnat et al. 2012)
MBP	(Sun, Greenway et al. 2005)
NEFH	(Bruce, Donaldson et al. 2004)
NEFM	(Schoenherr, Paquette et al. 1996)
NPAS4	(Bersten, Wright et al. 2014)
NPPA	(Schoenherr, Paquette et al. 1996)
NPTXR	(Bruce, Donaldson et al. 2004)

NRXN3	(Bruce, Donaldson et al. 2004)
NTRK3	(Bruce, Donaldson et al. 2004)
OPRM1	(Kim, Hwang et al. 2004)
PDGFRA	(Sun, Greenway et al. 2005)
PPP2R2C	(Sun, Greenway et al. 2005)
PRLHR	(Kemp, Lin et al. 2002)
PTPRN	(Martin, Allagnat et al. 2012)
RAB4A	(Sun, Greenway et al. 2005)
RESP18	(Sun, Greenway et al. 2005)
RGS7	(Bruce, Donaldson et al. 2004)
SCN3A	(Zhang, Pazin et al. 2008)
SF1	(Sun, Greenway et al. 2005)
SHANK2	(Sun, Greenway et al. 2005)
SNAP25	(Bruce, Donaldson et al. 2004)
STMN2	(Schoenherr and Anderson 1995)
SYN1	(Schoenherr and Anderson 1995)
SYP	(Schoenherr, Paquette et al. 1996)
SYT2	(Sun, Greenway et al. 2005)
SYT7	(Sun, Greenway et al. 2005)
TUBB3	(Zhang, Pazin et al. 2008)
UNC5D	(Sun, Greenway et al. 2005)
VRK3	(Sun, Greenway et al. 2005)
ZFM1	(Sun, Greenway et al. 2005)

Table A.2 Genes that are known to be regulated by REST.

Experimentally validated genes that are targeted by REST were derived from the literature. The bibliography is shown on the next page.

Appendix Bibliography

Ballas, N., C. Grunseich, et al. (2005). "REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis." Cell **121**(4): 645-657.

Bersten, D. C., J. A. Wright, et al. (2014). "Regulation of the neuronal transcription factor NPAS4 by REST and microRNAs." Biochimica et biophysica acta **1839**(1): 13-24.

Bessis, A., A. M. Salmon, et al. (1995). "Promoter elements conferring neuron-specific expression of the beta 2-subunit of the neuronal nicotinic acetylcholine receptor studied in vitro and in transgenic mice." Neuroscience **69**(3): 807-819.

Brene, S., C. Messer, et al. (2000). "Regulation of GluR2 promoter activity by neurotrophic factors via a neuron-restrictive silencer element." The European journal of neuroscience **12**(5): 1525-1533.

Bruce, A. W., I. J. Donaldson, et al. (2004). "Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes." Proceedings of the National Academy of Sciences of the United States of America **101**(28): 10458-10463.

Johnson, R., R. J. Gamblin, et al. (2006). "Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication." Nucleic acids research **34**(14): 3862-3877.

Kallunki, P., G. M. Edelman, et al. (1997). "Tissue-specific expression of the L1 cell adhesion molecule is modulated by the neural restrictive silencer element." The Journal of cell biology **138**(6): 1343-1354.

Kemp, D. M., J. C. Lin, et al. (2002). "NRSF/REST confers transcriptional repression of the GPR10 gene via a putative NRSE/RE-1 located in the 5' promoter region." FEBS letters **531**(2): 193-198.

Kim, C. S., C. K. Hwang, et al. (2004). "Neuron-restrictive silencer factor (NRSF) functions as a repressor in neuronal cells to regulate the mu opioid receptor gene." The Journal of biological chemistry **279**(45): 46464-46473.

Lemonde, S., A. Rogaeva, et al. (2004). "Cell type-dependent recruitment of trichostatin A-sensitive repression of the human 5-HT1A receptor gene." Journal of neurochemistry **88**(4): 857-868.

Martin, D., F. Allagnat, et al. (2012). "Specific silencing of the REST target genes in insulin-secreting cells uncovers their participation in beta cell survival." PloS one **7**(9): e45844.

Mu, W. and D. R. Burt (1999). "Transcriptional regulation of GABAA receptor gamma2 subunit gene." Brain research. Molecular brain research **67**(1): 137-147.

Otto, S. J., S. R. McCorkle, et al. (2007). "A new binding motif for the transcriptional repressor REST uncovers large gene networks devoted to neuronal functions." The Journal of neuroscience : the official journal of the Society for Neuroscience **27**(25): 6729-6739.

Park, S. Y., J. B. Kim, et al. (2007). "REST is a key regulator in brain-specific homeobox gene expression during neuronal differentiation." Journal of neurochemistry **103**(6): 2565-2574.

Schoenherr, C. J. and D. J. Anderson (1995). "The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes." Science **267**(5202): 1360-1363.

Schoenherr, C. J., A. J. Paquette, et al. (1996). "Identification of potential target genes for the neuron-restrictive silencer factor." Proceedings of the National Academy of Sciences of the United States of America **93**(18): 9881-9886.

Sun, Y. M., M. Cooper, et al. (2008). "Rest-mediated regulation of extracellular matrix is crucial for neural development." PloS one **3**(11): e3656.

Sun, Y. M., D. J. Greenway, et al. (2005). "Distinct profiles of REST interactions with its target genes at different stages of neuronal development." Molecular biology of the cell **16**(12): 5630-5638.

Toshiyuki, N. and M. Ichiro (2004). "Molecular mechanisms regulating cell type specific expression of BMP/RA Inducible Neural-specific Protein-1 that suppresses cell cycle progression: roles of NRSF/REST and DNA methylation." Brain research. Molecular brain research **125**(1-2): 47-59.

Zhang, P., M. J. Pazin, et al. (2008). "Nontelomeric TRF2-REST interaction modulates neuronal gene silencing and fate of tumor and stem cells." Current biology : CB **18**(19): 1489-1494.

Publications from this work

Maruyama, H., J. C. Harwood, et al. (2013). "An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis." EMBO reports **14**(8): 711-717.