CHANGES IN GLOBAL NUCLEOSOME POSITIONING OF HUMAN AND MOUSE EMBRYONIC STEM CELLS

by

Wenjuan Zhang

(Under the Direction of Shaying Zhao)

ABSTRACT

Nucleosome positioning (NP) is a fundamental parameter in chromatin packing/unpacking, playing key roles in transcriptional regulation and maintenance of genomic integrity. To better understand NP dynamics in ESCs, we studied two systems: a human embryonic stem cell (hESCs) differentiation system and a mouse embryonic stem cell (mESCs) knockout system.

We investigated how NP changes in a hESC differentiation system: WA09 hESCs → *ISL1*+ nascent mesoderm (INM) → smooth muscle cells (SMCs), by paired-end sequencing of mononucleosomal DNA fragments generated by micrococcal nuclease (MNase)-digestion (MNase-seq). The analysis reveals that at the promoter and gene body, NP is correlated primarily with transcriptional activity and secondarily with the GC content of the sequence. Pluripotent hESCs also exhibit a more dynamic NP than their differentiated derivatives, indicating that more genes are in the poised state and can be readily activated or silenced once differentiation starts. Surprisingly, the study finds mononucleosomal DNA of hESC to be ~10bp (one helix turn) longer than its differentiated WA09-SMC. Moreover, as fragment length increases, both the GC content

and %CpG also increase. Thus, longer nucleosomal DNA, possibly arisen from a larger histone core, could be instrumental in maintaining the remarkable genomic integrity of hESC by reducing mutations such as C→T changes.

We also studied the NP dynamics in promoters of mESCs and *Ext1*$^{-/-}$ ESCs, with customized high density Comparative Genomic Hybridization (CGH) arrays. We observed the similar pattern as hESCs that at the promoters, NP is correlated primarily with transcriptional activity and secondarily with the GC content of the sequence. In addition, we observed that genes with decreased expression had corresponding NP change while NP remain unchanged for genes have increased expression in the *Ext1*$^{-/-}$ ESCs. Interestingly, genes activated in *Ext1*$^{-/-}$ ESCs have significantly longer mRNA half-life than those silenced on average.

Both studies revealed a fundamental level of epigenetic control that was not previously recognized and adds to the complexity of epigenetic mechanisms. In addition, our studies provided a systematic tool of studying global NP with high-throughput approaches including the next generation sequencing and high density CGH arrays.

INDEX WORDS: Nucleosome positioning, ESC differentiation, Heparan Sulfate, MNase-seq, CGH array, sequence mutation, genomic stability, nucleosome core DNA, GC content

CHANGES IN GLOBAL NUCLEOSOME POSITIONING OF HUMAN AND MOUSE EMBRYONIC STEM CELLS

by

Wenjuan Zhang

BS, Anhui Medical University, China, 2005

MS, University of Science and Technology of China, China, 2008

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

CHANGES IN GLOBAL NUCLEOSOME POSITIONING OF HUMAN AND MOUSE EMBRYONIC STEM CELLS

by

Wenjuan Zhang

Major Professor: Shaying Zhao

Committee: Stephen Dalton
Liming Cai
Kevin Dobbin

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2013

DEDICATION

I would like to dedicate this thesis to my parents.

ACKNOWLEDGEMENTS

I am grateful to my advisor, Professor Shaying Zhao for her support and guidance over the years. She patiently provided the vision, encouragement and advice necessary for me to proceed through the doctorial program and complete my dissertation. Without her guidance and persistent help this dissertation would not have been possible.

I would also like to thank the members of my advisory committee: Drs. Stephen Dalton, Liming Cai and Kevin Dobbin, for their support, guidance and helpful suggestions during my graduate training. I would like to thank all past and present members of the Zhao lab for their friendship and assistance. I would like to thank my husband Jie Tang for standing beside me throughout my Ph.D. studies.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

INTRODUCTION OF NUCLEOSOMES

The chromatin

Eukaryotic genomes are packaged into a nucleoprotein complex known as chromatin, which affects most processes that occur on DNA. There are different levels of chromatin organizations: 1. Heterchromatin, such as those present at centromeres and telomeres, is in a very highly condensed state that resembles the chromatin of cells undergoing mitosis. Heterochromatin is transcriptionally inactive and contains highly repeated DNA sequences. 2. Euchromatin refers to the chromatin that appears less condensed in the microscope. Euchromatin participates in the active transcription of DNA to mRNA products. The unfolded structure allows gene regulatory proteins and RNA polymerase complexes to bind to the DNA sequence, which can then initiate the transcription process. Rearrangements of chromatin structure are of essential importance in regulating gene expression and other nuclear processes, such as DNA replication, recombination and repair, but also cell cycle progression and developmental transitions (Loidl 2001).

The nucleosomes

The basic structural unit of chromatin is nucleosome (Kornberg 1974; Kornberg and Thomas 1974), which was first defined over 30 years ago through a combination of biochemistry, microscopy, and X-ray crystallography. Under electron microscopy, polynucleosomal tracts appear as “beads on a string”, where nucleosomes are seen as beads, and the linker DNA is the string (Olins and Olins 1974; Woodcock et al. 1976). Core particles of nucleosome was initially predicted by nuclease digestion and electron microscopy and shown with a DNA repeat length of 200bp and four core histones, with DNA on the outside of the particle (Noll 1974b; Noll 1974a; Van Holde et al. 1974). Six years later, Klug et al. published a 25Å crystal structure defined the shape of histone core and the path of DNA around the nucleosome outer surface (Klug et al. 1980). Later, the crystal structure was solved at 7Å (Richmond et al. 1984) and 2.8Å (Luger et al. 1997), showing the details of the histone globular domains and the interactions of histone-histone and histone-DNA.

With the fine tuned crystal structure of nucleosomes, it has been recognized that nucleosome consists of 147 base pairs of DNA wrapped 1.7 times around an octamer of histone proteins (two copies each of histones H2A, H2B, H3, and H4) through an interface of ionic and hydrogen bonds (Luger et al. 1997; Richmond and Davey 2003). These histones are arranged into four stable heterodimers, two H2A/H2B and two H3/H4. A heterotetramer of H3 and H4 forms the backbone of the core particle and tightly binds the central DNA with symmetry about the nucleosome dyad (Wolffe and Guschin 2000). Histones are constantly disassembled and reassembled in a stepwise fashion where the assembly begins with H3/H4 dimer binds the nucleosome dyad region, then followed by the incorporation of H2A/H2B dimers.

There are several histone variants in eukaryotes, which can affect both the structure of individual nucleosomes and the ability of nucleosomes to form higher order chromatin structure (Downs et al. 2007). These variants are different in amino acid sequence and associated with particular biological processes and exhibit specific expression patterns (Talbert and Henikoff 2010). It was well characterized that the nucleosomes near the TSS usually contain the H2A.Z variant instead of histone H2A (Albert et al. 2007). Histone variants can also alter the bonds and interactions with DNA, which could change the nucleosome positions and DNA sequence preferences. For example, CENP-A, which was found replace H3 in centric heterchromatin, contains an insertion of two amino acids in a loop involved in the DNA interaction, moreover, evidences shown that CENP-A possibly even forming non-octametric particles with right-handed DNA supercoiling and wrap less than the canonical 147bp (Dalal et al. 2007; Furuyama and Henikoff 2009). Meanwhile, nucleosomes containing the H2A variant H2A.Bdb similarly organizes <147bp (Bao et al. 2004). Nucleosomes with the core DNA length ranged from 100bp to170bp, depending upon the histone core composition, have been reported (Zlatanova et al. 2009; Bonisch and Hake 2012; Hasson et al. 2013).

In addition, linker histones (H1) also affect the architecture and the degree of chromatin compaction (Downs et al. 2007). Linker histones associate with DNA influence the orientation of DNA relative to the nucleosomes at the entry/exit sites of the nucleosomes (Hamiche et al. 1996; Syed et al. 2010). The length of linker DNA varies between 20-90bp depending upon the species of the organism, tissues, and the region of the genomes (Szerlong and Hansen 2011). Linker histone binds DNA that facilitates the folding of chromatin into ~30nm fibers (Thoma et al. 1979; Yao et al. 1991).

THE ORGANIZATION OF NUCLEOSOMES ON GENES

Nucleosome positioning

Aside from compacting DNA, nucleosomes and chromatin also serve important functions in regulating the DNA-related process such as transcription, replication, recombination and DNA repair (Arya et al. 2010). Nucleosome positioning (NP) is defined as the probability that a nucleosome starts at a given base pair in the genome (Segal and Widom 2009). It affects the accessibility of a binding site for a particulate factor which plays a crucial role in regulating transcriptional activity. NP may be characterized with four descriptors: Nucleosome repeat length (NRL), Nucleosome fuzziness, Nucleosome occupancy and Nucleosome phasing (Arya et al. 2010). To date, NP has been mapped on many genomes including yeast, fly, worm and human (Yuan et al. 2005; Mavrich et al. 2008b; Schones et al. 2008; Shivaswamy et al. 2008; Valouev et al. 2008; Kaplan et al. 2009; Westenberger et al. 2009; Lantermann et al. 2010). By summarizing the characteristic features of NP in different regions of genome, people found the NP across the genome is far from random. The most striking feature reveled in the global NP mapping is the nucleosome density differences between the regulatory regions and transcribed sequences. Over 90% of the promoters in budding yeast contain fragments of DNA with very low nucleosome occupancy (Yuan et al. 2005; Lee et al.

2007). NP patterns were not only found in promoters but also found at the transcriptional termination regions and at the boundaries of active and repressed chromatin domains (Cuddapah et al. 2009). Surprisingly, NP pattern was also found at exon/intron junctions and RNA polyadenylation sites that suggest the link between RNA processing and chromatin organization (Schwartz et al. 2009; Spies et al. 2009; Tilgner et al. 2009).

NP in promoter regions

Studies of NP suggest two architectures of promoters, open and closed.

- Open promoters: The promoters typically have a clear and large (~150bp) nucleosome depleted region (NDR), where the transcription factor binding sites (TFBSs) are enriched, immediately upstream of the transcription start site (TSS). Even though the NDR region often referred as nucleosome free region, there is actually a gradient of nucleosome depletion, where the depletion is correlated with the transcriptional activity (Schones et al. 2008; Weiner et al. 2010). The NDR region is usually flanked by well positioned nucleosomes, where the nucleosome located in the downstream is commonly referred as the "+1" nucleosome and the one upstream is referred as the "-1" nucleosome (Yuan et al. 2005; Schones et al. 2008; Cairns 2009). Interestingly, human and fly promoters have the +1 nucleosome shifted downstream (~40-60bp) of the TSS compared to yeast. It has been observed that the nucleosome exhibit stronger phasing downstream than upstream of the TSS, and the average nucleosome occupancy is often higher in the downstream of TSS (Yuan et al. 2005; Mavrich et al. 2008b; Schones et al. 2008). The nucleosomes around TSS at open promoters generally contain the histone H2A variant H2A.Z (Albert et al. 2007).

- Closed promoters: The promoters do not contain a NDR signature but are rather have nucleosomes cover the TSS and the regions flanking the TSS. Most of the transcriptional activators binding sites are also covered by nucleosomes (Lee et al. 2007; Schones et al. 2008). However, study of yeast NP shows that at least one binding site, either in the linker DNA or in the entry/exit sites of the core DNA is exposed (Struhl 1985), where the exposed site allows the access of a pioneer transcription factor to the promoter.

NP at exons and 3' termination sites

Studies have been suggested that intron/exon junctions contain DNA sequences that promote the positioning of nucleosomes (Beckmann and Trifonov 1991; Kogan and Trifonov 2005). In the meantime, well positioned nucleosomes at most internal exons are independent of expression (Andersson et al. 2009). Nucleosomes positioned at these junctions were proposed to protect the splice sites of exons from mutations (Kogan and Trifonov 2005).

Recently, it has been demonstrated that the nucleosomes cover a gene transcription termination sites (TTS) with NDR signature. Specifically, a clear NDR was observed following a well-positioned nucleosomes occupies the TTS (Jiang and Pugh 2009). However, whether the positioned nucleosome located at 3' TTS of gene contributes to the termination is unclear.

CHROMATIN ARCHITECTURE AT PROMOTERS

NP dynamics at promoters

Noticeable changes of NP have been observed upon gene activation or repression. Generally, nucleosomes are evicted upon gene activation and nucleosomes are populated upon gene repression (Shivaswamy et al. 2008). However, changes in NP are restricted to few nucleosomes instead of large scale variations. For example, in CD4+ T-cells, the expressed gene show typical NP pattern in promoters while the unexpressed genes only show strong positioning of +1 nucleosomes (Schones et al. 2008). The histone variant H2A.Z replaces canonical H2A at the -3, -2, +1, +2, +3 when the -1 H2A.Z is lost upon gene activation. The +1 nucleosome also slides downstream of TSS from its initial position a few dozen base pairs (Lomvardas and Thanos 2001; Schones et al. 2008). It is possible that NP changes occur when the transcriptional activity of a gene remain unchanged, probably in preparation for future process. Genes in this state are proposed to be “poised” for transcription with RNA polymerase II either stalled at TSS and produce full-length transcripts with elongation hallmarks or produce short, aborted transcripts (Guenther et al. 2007).

Histone variants and modifications at promoters

The histone variant H2A.Z differs from canonical H2A in its amino-terminal tail sequence and may also affects nucleosome stability by affecting the interactions with H3/H4 tetramer and with itself. The distribution of H2A.Z varies within the promoter in different organism, although certain themes have emerged. In yeast, H2A.Z occupied the +1 and -1 nucleosomes, with lower amount at +2 (Guillemette et al. 2005; Raisner et al.

2005; Zhang et al. 2005). In flies, H2A.Z is absent at the -1 position but present in +1 and several downstream nucleosomes (Mavrich et al. 2008b). In humans, H2A.Z localize from -3 to +3 in genes with low expression (Schones et al. 2008). Functional studies in yeast indicate that H2A.Z is promoting activation and H2A.Z containing nucleosomes are lost upon transcription activation (Guillemette et al. 2005; Raisner et al. 2005; Zhang et al. 2005), especially in human, H2A.Z is lost at -1 position (Schones et al. 2008). This might be caused by the destabilizing effect of H2A.Z on nucleosome core so that H2A.Z containing nucleosomes can be removed more easily (Jin and Felsenfeld 2007).

The nucleosomes at promoters are also subjected to posttranslational modifications that corresponding with transcriptional states. In open promoters, histone markers such as H3K4me3, H2BK5me1, H3K9ac and H3K14ac are enriched significantly in the nucleosome surrounding the TSS (Roh et al. 2006; Barski et al. 2007; Guenther et al. 2007; Kouzarides 2007). Closed promoter often associates with repressive markers such as H3K27me3 (Barski et al. 2007) as well as DNA CpG methylation. Poised promoters are known to contain both H3K4me3 and H3K27me3, which usually referred as “bivalent markers” (Bernstein et al. 2006).

DNA methylation

Methylation at cytosine of CpG has an important role in regulating gene transcription in higher eukaryotes. Several studies point to a strong correlation between DNA methylation status and NP, where experiments indicate that DNA methylation can induce a more compact and rigid nucleosome structure (Choy et al. 2010). In Arabidopsis and human, nucleosomal DNA was found more highly methylated than flanking DNA, suggesting that NP influences DNA methylation patterning throughout the genome and that DNA methyltransferases preferentially target nucleosome-bound DNA (Chodavarapu et al. 2010).

DETERMINANTS OF NP

DNA sequences

Nucleosomes have different affinities for differing DNA sequence that varies in 5000 fold range (Thastrom et al. 2004; Gencheva et al. 2006). It was recognized that the sequence of DNA affects its ability to form nucleosomes before the genomic era (Satchwell et al. 1986; Ioshikhes et al. 1996). The nucleosome preferred DNA sequences does not arise from its interactions with particular DNA bases but rather arise from the sequence-dependent mechanics of the wrapped DNA itself (Widom 2001).

Early studies shown that intrinsic DNA sequence preferences plays a role in NP *in vivo* (Zhurkin et al. 1979; Trifonov and Sussman 1980; Lowary and Widom 1998). Subsequent analyses suggested that *in vitro* nucleosome preferences mirror *in vivo* positions (Gencheva et al. 2006) and that NP *in vivo* can be predicted based on DNA sequence (Segal et al. 2006; Yuan and Liu 2008; Kaplan et al. 2009).

In 1980s, study of a few hundred nucleosomal sequences showed that the most favorable DNA sequences for positioning nucleosomes are WW (W is A or T) and SS (S is C or G) dinucleotides occurring at 10bp intervals (Satchwell et al. 1986; Ioshikhes et al.

2006), where the SS dinucleotides is offset by 5bp compared with the WW patterns. Thousands of nucleosomal DNAs show the same pattern in earlier studies (Satchwell et al. 1986; Widom 2001; Albert et al. 2007; Peckham et al. 2007; Mavrich et al. 2008a; Miele et al. 2008; Valouev et al. 2008). These dinucleotide preferences are now partially understood. The 10bp periodical presence of certain dinucleotides provides a sharp bending every DNA helical repeat, because the WW dinucleotides tend to expand the major groove and face outward of the histone octamer, whereas the SS dinucleotides tend to contract the major groove of DNA 5bp away with the opposite direction when the backbone face inward (Satchwell et al. 1986; Widom 2001; Segal et al. 2006). In addition, histone arginine side chains could potentially provide additional 10bp periodic base specificity by insert into the minor groove every DNA helical repeat (Richmond and Davey 2003). Other sequence combinations including many different 5-mers are disfavored by nucleosomes, certain motifs are disfavored from being located anywhere inside the nucleosome (Field 2008; Kaplan et al. 2009). In addition, GC rich sequences are believed to facilitate the assembly of nucleosomes by increasing DNA flexibility (Peckham et al. 2007; Chung and Vingron 2009; Tillo and Hughes 2009), while poly-AT sequences disfavor the nucleosome formation (Field 2008; Mavrich 2008; Kaplan et al. 2009).

Computational models have been developed to predict the genome-wide location of nucleosomes, based on the existence of sequence preferences of nucleosome in combination with the preferences of nucleosome for different 5-mers. The models correctly predicted 74% of the nucleosome positions in yeast and 60% of the nucleosome positions in worms (Kaplan et al. 2009). Furthermore, it also reproduced the NP pattern at gene promoters obtained by experiments.

Two groups studied the correlation between the *in vitro* and *in vivo* nucleosome positions in yeast and generated different conclusions. Kaplan *et al.* (Kaplan et al. 2009) favored the idea that DNA sequence encodes the genomic NP *in vivo* intrinsically, while Zhang *et al.* (Zhang et al. 2009) stated that NP *in vivo* was not determined by DNA-nucleosome interaction significantly. This debate was resolved in the subsequent analyses by defining nucleosome occupancy and translational NP (the nucleosome extent measured when aligned relative to a given position). The correlation between *in vitro* and *in vivo* nucleosome maps is high (about 70%) when nucleosome occupancy is measured and much lower (15-20%) when translational NP is measured specifically (Radman-Livaja and Rando 2010).

Transcription factors

Transcription factors (TFs) can directly influence the NP *in vivo* by competing with nucleosomes for the access of DNA. Studies of protein-DNA complex structure have revealed that many site specific DNA-binding proteins cannot occupy their target sites when the native state nucleosomes present (Segal and Widom 2009); whereas nucleosomes must be repositioned at nucleosome occupy TF sites to enable TFs to access their sites (Yu and Morse 1999; Adkins and Tyler 2006). The detailed competition between TFs and nucleosomes of DNA target sites is not known. It could result from the partially unwrapped DNA from one end of the nucleosomes, thus enabling TF access

(Polach and Widom 1995; Anderson et al. 2002). Both *in vitro* and *in vivo* studies indicate that the binding of multiple proximal TFs will be inherently cooperative: upon the removal of the occupied nucleosomes and binding of one TF, subsequent TFs will bind the now nucleosome-free DNA more easily (Vashee et al. 1998; Miller and Widom 2003). It was suggested that the equilibrium between TFs and nucleosome binding account for the drastically nucleosome change across different cellular events (Segal and Widom 2009). Furthermore, it was also proposed that histone modifications and DNA methylations lead to the drastic change of some nucleosome positions, but not strong enough to affect the balance for other nucleosomes (Arya et al. 2010).

Chromatin remodeler

ATP-dependent chromatin remodelers use the energy of ATP hydrolysis to move nucleosomes along the DNA or to disassemble nucleosomes and remove them from DNA (Boeger et al. 2008; Hartley and Madhani 2009). Especially, the spacing complexes of the ISW (imitation switch) family, such as ACF (ATP-dependent chromatin assembly and remodeling factor) and CHRAC (chromatin accessibility complex) are known to establish regular nucleosome spacing (Ito et al. 1997; Saha et al. 2006; Gangaraju and Bartholomew 2007). Moreover, an essential remodeler RSC (Chromatin structure remodeling complex) is required for normal NP of the nucleosomes flanking the NDR (Hartley and Madhani 2009).

CHROMATIN AND EMBRYONIC STEM CELL DIFFERENTIATION

Embryonic stem cells (ESCs) are pluripotent, self-renewing cells derived from inner cell mass (ICM) of the developing blastocyst-stage embryo (Thomson et al. 1998). They have nearly unlimited self-renewal capability *in vitro* and have the ability to differentiation into all three germ layers -ectoderm; mesoderm and endoderm- and any fully differentiate cell of the body (Ameen et al. 2008). Once the differentiation initiated, lineage-specification occur though the establishment of a unique genome-wide transcriptional profile. Chromatin state including histone modifications and other proteins that pack the genome, has been considered relate closely with the cell fate decisions (Kouzarides 2007; Surani et al. 2007).

ESCs seem to have a distinct high-order chromatin structure that are richer in euchromatin and, accumulate highly condensed heterochromatin regions as the differentiation progresses (Francastel et al. 2000; Arney and Fisher 2004). Histone modification patterns also changes accompanied with these chromatins changes. Studies have shown that ESCs chromatin is more associated with more active marks. When ESC differentiate, silenced chromatin mark H3K9me3 increased while active chromatin mark -acetylated histones H3 and H4- decreased (Lee et al. 2004; Kurisaki et al. 2005). Besides, regions with both active mark H3K4me3 and repressive mark H3K27me3 have been defined as “bivalent domains”, which were found in many critical regions involved in pluripotency and differentiation of ESCs (Bernstein et al. 2006). The changes of the relative levels of two modifications can discriminate the expression status in ESCs (Mikkelsen et al. 2007). Regions with both H3K4me3 and H3K27me3 modifications can be resolved to monovalent modification when the ESCs committed to a specific lineage (Azuara et al. 2006; Bernstein et al. 2006).

Recently, NP study during lineage commitment in mouse ESCs have been suggested that the nucleosome occupancy was correlated with certain histone modifications throughout the genome. Moreover, the average nucleosome repeat length is 5-7 base pair longer in the differentiated cells, indicating the involvement of NP in the regulatory layer of cell differentiation (Teif et al. 2012).

OUR STUDY

ESCs and differentiated cells share the same genomic DNA sequences but have distinct morphology and cellular functions. Many of the underlying cell-fate decisions occur through gene expression alterations which regulated by epigenetic changes that involved in chromatin features. Current epigenetic research in ESC and its differentiation has greatly focused on histone modifications and DNA methylation, leaving another equally fundamental epigenetic mechanism, NP (Ozsolak et al. 2007; Schones et al. 2008; Jiang and Pugh 2009; Prendergast and Semple 2011; Valouev et al. 2011; Gaffney et al. 2012; Teif et al. 2012; Struhl and Segal 2013), relatively understudied. Therefore, our studies focused on how NP involved in the determination of cellular functions during human ESCs lineage commitment and how NP rearranged with the inactivation of heparan sulfate (HS) expression in mouse ESCs. HS was recognized by playing key roles in cell growth and development (Kraushaar et al. 2010).

For human ESCs, we investigated NP in a well-defined differentiation system: WA09 ESCs → *ISL1*+ nascent mesoderm (INM) → smooth muscle cells (SMCs), by paired-end sequencing of their mononucleosomal DNA fragments generated by micrococcal nuclease (MNase)-digestion (MNase-seq), and for mouse ESCs, we studied promoter NP with high density customized high density Comparative Genomic Hybridization (CGH) arrays hybridized with MNase-digested mononucleosomal DNA and control DNA from mouse *Ext1*$^{+/+}$ and *Ext1* $^{-/-}$ ESCs. In both human and mouse ESCs, we observed the correlations between NP dynamics and transcription activity as well as sequence GC content. In addition, we explored the potential relationship between NP and mRNA half-life, which was investigated the first time. Furthermore, we integrated our NP analysis with other epigenetic markers in human ESCs, such as histone modifications and DNA methlylations. Our results suggested that NP was involved in the regulatory layer of ESC differentiation with correlation of histone modification and DNA methylation patterns.

In summary, both studies revealed a fundamental level of epigenetic control –NP- that was not previously recognized as much as histone modifications and DNA menthylations and add to the complexity of epigenetic mechanisms that contributed to the pluripotency and self-renewal of ESCs. In addition, our studies provided a systematic tool of studying global NP with high-throughput approaches including the next generation sequencing and high density CGH arrays.

REFERENCE

Adkins MW, Tyler JK. 2006. Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Molecular cell* **21**(3): 405-416.

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF. 2007. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature* **446**(7135): 572-576.

Ameen C, Strehl R, Bjorquist P, Lindahl A, Hyllner J, Sartipy P. 2008. Human embryonic stem cells: current technologies and emerging industrial applications. *Critical reviews in oncology/hematology* **65**(1): 54-80.

Anderson JD, Thastrom A, Widom J. 2002. Spontaneous access of proteins to buried nucleosomal DNA target sites occurs via a mechanism that is distinct from nucleosome translocation. *Molecular and cellular biology* **22**(20): 7147-7157.

Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. 2009. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome research* **19**(10): 1732-1741.

Arney KL, Fisher AG. 2004. Epigenetic aspects of differentiation. *Journal of cell science* **117**(Pt 19): 4355-4363.

Arya G, Maitra A, Grigoryev SA. 2010. A structural perspective on the where, how, why, and what of nucleosome positioning. *Journal of biomolecular structure & dynamics* **27**(6): 803-820.

Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M et al. 2006. Chromatin signatures of pluripotent cell lines. *Nature cell biology* **8**(5): 532-538.

Bao Y, Konesky K, Park YJ, Rosu S, Dyer PN, Rangasamy D, Tremethick DJ, Laybourn PJ, Luger K. 2004. Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *The EMBO journal* **23**(16): 3314-3324.

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

Beckmann JS, Trifonov EN. 1991. Splice junctions follow a 205-base ladder. *Proceedings of the National Academy of Sciences of the United States of America* **88**(6): 2380-2383.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**(2): 315-326.

Boeger H, Griesenbeck J, Kornberg RD. 2008. Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell* **133**(4): 716-726.

Bonisch C, Hake SB. 2012. Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic acids research* **40**(21): 10719-10741.

Cairns BR. 2009. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**(7261): 193-198.

Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* **466**(7304): 388-392.

Choy JS, Wei S, Lee JY, Tan S, Chu S, Lee TH. 2010. DNA methylation increases nucleosome compaction and rigidity. *Journal of the American Chemical Society* **132**(6): 1782-1783.

Chung HR, Vingron M. 2009. Sequence-dependent nucleosome positioning. *Journal of molecular biology* **386**(5): 1411-1422.

Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. 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.

Dalal Y, Wang H, Lindsay S, Henikoff S. 2007. Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS biology* **5**(8): e218.

Downs JA, Nussenzweig MC, Nussenzweig A. 2007. Chromatin dynamics and the preservation of genetic information. *Nature* **447**(7147): 951-958.

Field Y. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol* **4**: e1000216.

Francastel C, Schubeler D, Martin DI, Groudine M. 2000. Nuclear compartmentalization and gene activity. *Nature reviews Molecular cell biology* **1**(2): 137-143.

Furuyama T, Henikoff S. 2009. Centromeric nucleosomes induce positive DNA supercoils. *Cell* **138**(1): 104-113.

Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, Widom J, Gilad Y, Pritchard JK. 2012. Controls of nucleosome positioning in the human genome. *PLoS genetics* **8**(11): e1003036.

Gangaraju VK, Bartholomew B. 2007. Mechanisms of ATP dependent chromatin remodeling. *Mutation research* **618**(1-2): 3-17.

Gencheva M, Boa S, Fraser R, Simmen MW, CB AW, Allan J. 2006. In Vitro and in Vivo nucleosome positioning on the ovine beta-lactoglobulin gene are related. *Journal of molecular biology* **361**(2): 216-230.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**(1): 77-88.

Guillemette B, Bataille AR, Gevry N, Adam M, Blanchette M, Robert F, Gaudreau L. 2005. Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS biology* **3**(12): e384.

Hamiche A, Schultz P, Ramakrishnan V, Oudet P, Prunell A. 1996. Linker histone-dependent DNA structure in linear mononucleosomes. *Journal of molecular biology* **257**(1): 30-42.

Hartley PD, Madhani HD. 2009. Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**(3): 445-458.

Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N, Alonso A, Warburton PE, Black BE. 2013. The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nature structural & molecular biology*.

Ioshikhes I, Bolshoy A, Derenshteyn K, Borodovsky M, Trifonov EN. 1996. Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *Journal of molecular biology* **262**(2): 129-139.

Ioshikhes IP, Albert I, Zanton SJ, Pugh BF. 2006. Nucleosome positions predicted through comparative genomics. *Nature Genet* **38**: 1210-1215.

Ito T, Bulger M, Pazin MJ, Kobayashi R, Kadonaga JT. 1997. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* **90**(1): 145-155.

Jiang C, Pugh BF. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nature reviews Genetics* **10**(3): 161-172.

Jin C, Felsenfeld G. 2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & development* **21**(12): 1519-1529.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**(7236): 362-366.

Klug A, Rhodes D, Smith J, Finch JT, Thomas JO. 1980. A low resolution structure for the histone core of the nucleosome. *Nature* **287**(5782): 509-516.

Kogan S, Trifonov EN. 2005. Gene splice sites correlate with nucleosome positions. *Gene* **352**: 57-62.

Kornberg RD. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**(4139): 868-871.

Kornberg RD, Thomas JO. 1974. Chromatin structure; oligomers of the histones. *Science* **184**(4139): 865-868.

Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**(4): 693-705.

Kraushaar DC, Yamaguchi Y, Wang L. 2010. Heparan sulfate is required for embryonic stem cells to exit from self-renewal. *The Journal of biological chemistry* **285**(8): 5907-5916.

Kurisaki A, Hamazaki TS, Okabayashi K, Iida T, Nishine T, Chonan R, Kido H, Tsunasawa S, Nishimura O, Asashima M 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.

Lantermann AB, Straub T, Stralfors A, Yuan GC, Ekwall K, Korber P. 2010. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nature structural & molecular biology* **17**(2): 251-257.

Lee JH, Hart SR, Skalnik DG. 2004. Histone deacetylase activity is required for embryonic stem cell differentiation. *Genesis* **38**(1): 32-38.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics* **39**(10): 1235-1244.

Loidl P. 2001. Introduction: assembly, remodeling and modification of chromatin. *Cellular and molecular life sciences : CMLS* **58**(5-6): 663-664.

Lomvardas S, Thanos D. 2001. Nucleosome sliding via TBP DNA binding in vivo. *Cell* **106**(6): 685-696.

Lowary PT, Widom J. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of molecular biology* **276**(1): 19-42.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**(6648): 251-260.

Mavrich TN. 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* **18**: 1073-1083.

Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF. 2008a. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome research* **18**(7): 1073-1083.

Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC et al. 2008b. Nucleosome organization in the Drosophila genome. *Nature* **453**(7193): 358-362.

Miele V, Vaillant C, d'Aubenton-Carafa Y, Thermes C, Grange T. 2008. DNA physical properties determine nucleosome occupancy from yeast to fly. *Nucleic acids research* **36**(11): 3746-3756.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153): 553-560.

Miller JA, Widom J. 2003. Collaborative competition mechanism for gene activation in vivo. *Molecular and cellular biology* **23**(5): 1623-1632.

Noll M. 1974a. Internal structure of the chromatin subunit. *Nucleic acids research* **1**(11): 1573-1578.

-. 1974b. Subunit structure of chromatin. *Nature* **251**(5472): 249-251.

Olins AL, Olins DE. 1974. Spheroid chromatin units (v bodies). *Science* **183**(4122): 330-332.

Ozsolak F, Song JS, Liu XS, Fisher DE. 2007. High-throughput mapping of the chromatin structure of human promoters. *Nature biotechnology* **25**(2): 244-248.

Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K, Weng Z. 2007. Nucleosome positioning signals in genomic DNA. *Genome research* **17**(8): 1170-1177.

Polach KJ, Widom J. 1995. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *Journal of molecular biology* **254**(2): 130-149.

Prendergast JG, Semple CA. 2011. Widespread signatures of recent selection linked to nucleosome positioning in the human lineage. *Genome research* **21**(11): 1777-1787.

Radman-Livaja M, Rando OJ. 2010. Nucleosome positioning: how is it established, and why does it matter? *Developmental biology* **339**(2): 258-266.

Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD. 2005. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**(2): 233-248.

Richmond TJ, Davey CA. 2003. The structure of DNA in the nucleosome core. *Nature* **423**(6936): 145-150.

Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A. 1984. Structure of the nucleosome core particle at 7 A resolution. *Nature* **311**(5986): 532-537.

Roh TY, Cuddapah S, Cui K, Zhao K. 2006. The genomic landscape of histone modifications in human T cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**(43): 15782-15787.

Saha A, Wittmeyer J, Cairns BR. 2006. Mechanisms for nucleosome movement by ATP-dependent chromatin remodeling complexes. *Results and problems in cell differentiation* **41**: 127-148.

Satchwell SC, Drew HR, Travers AA. 1986. Sequence periodicities in chicken nucleosome core DNA. *Journal of molecular biology* **191**(4): 659-675.

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**(5): 887-898.

Schwartz S, Meshorer E, Ast G. 2009. Chromatin organization marks exon-intron structure. *Nature structural & molecular biology* **16**(9): 990-995.

Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J. 2006. A genomic code for nucleosome positioning. *Nature* **442**(7104): 772-778.

Segal E, Widom J. 2009. What controls nucleosome positions? *Trends in genetics : TIG* **25**(8): 335-343.

Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR. 2008. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS biology* **6**(3): e65.

Spies N, Nielsen CB, Padgett RA, Burge CB. 2009. Biased chromatin signatures around polyadenylation sites and exons. *Molecular cell* **36**(2): 245-254.

Struhl K. 1985. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **82**(24): 8419-8423.

Struhl K, Segal E. 2013. Determinants of nucleosome positioning. *Nature structural & molecular biology* **20**(3): 267-273.

Surani MA, Hayashi K, Hajkova P. 2007. Genetic and epigenetic regulators of pluripotency. *Cell* **128**(4): 747-762.

Syed SH, Goutte-Gattat D, Becker N, Meyer S, Shukla MS, Hayes JJ, Everaers R, Angelov D, Bednar J, Dimitrov S. 2010. Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. *Proceedings of the National Academy of Sciences of the United States of America* **107**(21): 9620-9625.

Szerlong HJ, Hansen JC. 2011. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **89**(1): 24-34.

Talbert PB, Henikoff S. 2010. Histone variants--ancient wrap artists of the epigenome. *Nature reviews Molecular cell biology* **11**(4): 264-275.

Teif VB, Vainshtein Y, Caudron-Herger M, Mallm JP, Marth C, Hofer T, Rippe K. 2012. Genome-wide nucleosome positioning during embryonic stem cell development. *Nature structural & molecular biology* **19**(11): 1185-1192.

Thastrom A, Bingham LM, Widom J. 2004. Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. *Journal of molecular biology* **338**(4): 695-709.

Thoma F, Koller T, Klug A. 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.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* **282**(5391): 1145-1147.

Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J, Guigo R. 2009. Nucleosome positioning as a determinant of exon recognition. *Nature structural & molecular biology* **16**(9): 996-U124.

Tillo D, Hughes TR. 2009. G+C content dominates intrinsic nucleosome occupancy. *BMC bioinformatics* **10**: 442.

Trifonov EN, Sussman JL. 1980. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proceedings of the National Academy of Sciences of the United States of America* **77**(7): 3816-3820.

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K 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, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. *Nature* **474**(7352): 516-520.

Van Holde KE, Sahasrabuddhe CG, Shaw BR. 1974. A model for particulate structure in chromatin. *Nucleic acids research* **1**(11): 1579-1586.

Vashee S, Melcher K, Ding WV, Johnston SA, Kodadek T. 1998. Evidence for two modes of cooperative DNA binding in vivo that do not involve direct protein-protein interactions. *Current biology : CB* **8**(8): 452-458.

Weiner A, Hughes A, Yassour M, Rando OJ, Friedman N. 2010. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome research* **20**(1): 90-100.

Westenberger SJ, Cui L, Dharia N, Winzeler E. 2009. Genome-wide nucleosome mapping of Plasmodium falciparum reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. *BMC genomics* **10**: 610.

Widom J. 2001. Role of DNA sequence in nucleosome stability and dynamics. *Quarterly reviews of biophysics* **34**(3): 269-324.

Wolffe AP, Guschin D. 2000. Review: chromatin structural features and targets that regulate transcription. *Journal of structural biology* **129**(2-3): 102-122.

Woodcock CL, Safer JP, Stanchfield JE. 1976. Structural repeating units in chromatin. I. Evidence for their general occurrence. *Experimental cell research* **97**: 101-110.

Yao J, Lowary PT, Widom J. 1991. Linker DNA bending induced by the core histones of chromatin. *Biochemistry* **30**(34): 8408-8414.

Yu L, Morse RH. 1999. Chromatin opening and transactivator potentiation by RAP1 in Saccharomyces cerevisiae. *Molecular and cellular biology* **19**(8): 5279-5288.

Yuan GC, Liu JS. 2008. Genomic sequence is highly predictive of local nucleosome depletion. *PLoS computational biology* **4**(1): e13.

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science* **309**(5734): 626-630.

Zhang H, Roberts DN, Cairns BR. 2005. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123**(2): 219-231.

Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, Liu XS, Struhl K. 2009. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nature structural & molecular biology* **16**(8): 847-852.

Zhurkin VB, Lysov YP, Ivanov VI. 1979. Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic acids research* **6**(3): 1081-1096.

Zlatanova J, Bishop TC, Victor JM, Jackson V, van Holde K. 2009. The nucleosome family: dynamic and growing. *Structure* **17**(2): 160-171.

CHAPTER 2

CHANGES IN GLOBAL NUCLEOSOME POSITIONING DURING HUMAN EMBRYONIC STEM CELL DIFFERENTIATION[1]

[1]Wenjuan Zhang, Yaping Li, Michael Kulik, Rochelle L. Tiedemann, Keith D. Robertson, Stephen Dalton, Shaying Zhao. Submitted to *Genome Biology*, 05/31/2013.

ABSTRACT

Background

Nucleosome positioning (NP) is a fundamental parameter in chromatin packing/unpacking, playing key roles in transcriptional regulation and maintenance of genomic integrity. However, it is not as well studied as other levels of epigenetic controls such as histone modifications and DNA methylation.

Results

To better understand NP, we investigated how it changes in a human embryonic stem cell (hESC) differentiation system: WA09 hESCs → *ISL1*+ nascent mesoderm (INM) → smooth muscle cells (SMCs), by paired-end sequencing of mononucleosomal DNA fragments generated by micrococcal nuclease (MNase)-digestion (MNase-seq). The analysis reveals that at the promoter and gene body, NP is correlated primarily with transcriptional activity and secondarily with the GC content of the sequence. Pluripotent hESCs also exhibit a more dynamic NP than their differentiated derivatives, indicating that more genes are in the poised state and can be readily activated or silenced once differentiation starts. Surprisingly, the study finds mononucleosomal DNA of hESC to be ~10bp (one helix turn) longer than its differentiated WA09-SMC. Moreover, as fragment length increases, both the GC content and %CpG also increase. Thus, longer nucleosomal DNA, possibly arisen from a larger histone core, could be instrumental in maintaining the remarkable genomic integrity of hESC by reducing mutations such as C→T changes.

Conclusions

The study reveals a fundamental level of epigenetic control that was not previously recognized and adds to the complexity of epigenetic mechanisms that operate in pluripotent cells such as bivalent histone modifications and hydroxymethylation. Importantly, our findings shed light on the mechanisms through which hESCs maintain their remarkable genomic stability.

Keywords

Nucleosome positioning, hESC differentiation, MNase-seq, sequence mutation, genomic stability, nucleosome core DNA, GC content

BACKGROUND

Human embryonic stem cells (hESCs) are pluripotent and therefore have the capacity to differentiate into all cell types of the adult (Thomson et al. 1998; Menendez et al. 2013). If cultured under the appropriate conditions, pluripotent stem cells maintain their genomic integrity, ensuring fidelity in the transmission of genetic information from generation to generation (Maitra et al. 2005; Funk et al. 2012). These unique features could be attributed to epigenetic mechanisms, which are critical in chromatin remodeling and for cell fate specification and cell identity establishment (Gopalakrishnan et al. 2008; Hiratani and Gilbert 2009; Lister et al. 2009; Laurent et al. 2010; Consortium et al. 2012; Xiao et al. 2012; Hu et al. 2013). Indeed, studies have demonstrated that hESCs bear unique chromatin compared to somatic cells, including prevalent bivalent histone modifications and DNA hydroxymethylation (Bernstein et al. 2006; Mikkelsen et al. 2007; Zhao et al. 2007; Cui et al. 2009; Ficz et al. 2011; Koh et al. 2011; Ruzov et al. 2011; Wu et al. 2011). Additionally, changes in DNA methylation and hydroxymethylation, as well as histone modification patterns occur frequently and extensively throughout the course of differentiation (Mikkelsen et al. 2007; Gopalakrishnan et al. 2008; Cui et al. 2009; Lister et al. 2009; Laurent et al. 2010; Ficz et al. 2011; Koh et al. 2011).

Current epigenetic research in hESC and its differentiation has greatly focused on histone modifications and DNA methylation, leaving another equally fundamental epigenetic mechanism, nucleosome positioning (NP) (Ozsolak et al. 2007; Schones et al.

2008; Jiang and Pugh 2009; Prendergast and Semple 2011; Valouev et al. 2011; Gaffney et al. 2012; Teif et al. 2012; Struhl and Segal 2013), relatively understudied.

To more comprehensively understand the chromatin of hESC and its changes during differentiation, we investigated NP in a well-defined differentiation system: WA09 hESCs → *ISL1*+ nascent mesoderm (INM) → smooth muscle cells (SMCs), by paired-end sequencing of their mononucleosomal DNA fragments generated by micrococcal nuclease (MNase)-digestion (MNase-seq). Consistent with published studies at the promoter and gene body (Ozsolak et al. 2007; Schones et al. 2008; Valouev et al. 2011), we noted that NP was correlated primarily with the transcriptional activity and secondarily with the GC content of the sequence, and that WA09-hESCs possessed a more poised and dynamic NP than their differentiated derivatives. An unexpected finding is, however, that mononucleosomal DNA of WA09-hESCs is about 10bp, one helix turn, longer than that of their differentiated WA09-SMCs. Critically, within each cell type, as the DNA fragment length increases, both the GC content and %CpGs also ascend. We hypothesize that the longer mononucleosomal DNA arises from a larger histone core (Zlatanova et al. 2009) and contributes to the maintenance of genomic stability of hESCs by reducing the C→T substitution, the dominating background sequence mutation in the human genome (Venter et al. 2001; Prendergast and Semple 2011; Chen et al. 2012). Hence, our findings represent potentially another unique feature of the chromatin of pluripotent stem cells that may be as fundamental as bivalent histone modifications and DNA hydroxymethylation.

RESULTS

WA09-hESC → WA09-INM → WA09-SMC differentiation

Several analyses support this WA09-hESC → WA09-INM → WA09-SMC differentiation. First, we conducted microarray experiments to examine gene expression changes during the differentiation, of which principle component analysis (PCA) indicated a clear separation among the three cell types (Figure 2.1), supporting the cellular homogeneity for each. Second, we investigated established markers characteristic of each cell type and observed the corresponding changes. These include: 1) silencing of pluripotent markers *SOX2*, *OCT4* and *NANOG* upon WA09-hESC differentiation; 2) high level expression of INM markers *ISL1* and *HAND1* in WA09-INM only; and 3) significant activation of SMC markers such as *ACTA2* in WA09-SMC. These individual marker observations are further corroborated by global gene functional changes during the differentiation. Specifically, at false discovery rate (FDR) of < 0.1, we found a total of 753 genes with expression altered during the first WA09-hESC → WA09-INM differentiation stage. Among them, while known and putative pluripotent markers (20 total) (Galan et al. 2013) were silenced ($p < 2.2e-16$), genes related to development, extracellular matrix (ECM), and focal adhesion were significantly activated (FDR $< 3.5e-8$) (Figure 2.1). The 2nd stage of differentiation, WA09-INM → WA09-SMC, however, was characterized by downregulation of cell adhesion molecules and tight junction genes ($p < 1.0e-04$), as well as upregulation of genes that are consistent with SMC properties, including 29 smooth muscle, actin or calponin-related genes ($p=1.1e-07$). In summary, both individual and global gene expression analyses support the indicated identity and the homogeneity for each cell type of the differentiation.

Mononucleosomal DNA fragment isolation, sequencing and mapping

To isolate mononucleosomal DNA fragments, we treated the cells with MNase that yielded >98% mononucleosomes in each cell type, gel-purified the mononucleosomal DNA band (approximately 150bp), and sequenced from both ends. In total, we generated 205 – 226 million end sequence pairs of 90bp and placed over 94% of them uniquely back onto the human genome properly (both ends on the same chromosome, with the right orientation and a reasonable genomic distance). This resulted in a >10X coverage in both sequence and mononucleosomal fragments. As control, we also sequenced randomly sheared genomic DNA fragments of 150-200bp of WA09-hESC (termed "gWA09-hESC") to a 12X coverage, and achieved the same sequencing and mapping efficiency.

NP changes at the promoter and other parts of the genes during the differentiation

We first focused on the gene promoter and its flanking regions [-5kb to +5kb of the transcription start site (TSS)] for nucleosome occupancy investigation. We observed a nucleosome depleted region (NDR) and well-positioned nucleosomes immediately upstream and downstream of the TSS respectively, the extent of which is directly correlated with the transcriptional activity of the genes (Figure 2.2). In addition, a correlation with the GC content of the sequence, albeit to a much lesser degree, was also noted (Figure 2.2). These correlations also apply to the exon/intron and intron/exon junctions. Lastly, a NDR was detected at the transcription termination site (TTS), which however appear to arise largely from the low GC content of the sequence itself (Figure

2.2). These findings are consistent with published studies (Segal et al. 2006; Ozsolak et al. 2007; Schones et al. 2008).

Another observation from Figure 2.2 is that in WA09-hESC, at both the promoter and gene body, the nucleosome occupancy strength is more or less evenly spaced among the six groups of genes classified based on their expression level. However, in WA09-INM and WA09-SMC, while the two silent groups (with expression intensities of 100-250 and <100 respectively) display similar nucleosome occupancy levels, they are more distant from the expressed groups (with expression intensities of >250). This is consistent with published findings that hESC contains more genes in a poised chromatin (Bernstein et al. 2006; Zhao et al. 2007), readily to resume transcription or to be more irreversibly silenced upon differentiation. The poised chromatin and genes progressively decreased once differentiation started (Figure 2.2).

NP at the promoters matches the published CpG methylation, H3K4me3 and H3K27me3 enrichment status in WA09-hESCs.

We integrated our NP studies with published DNA methylation data for WA09-hESC (Laurent et al. 2010). As shown Figure 2.3, only unmethylated genes display a prominent NDR and as expected, a large portion of the methylated genes were not expressed or expressed at a very low level. These observations are consistent with the notion that NP precedes DNA methylation in gene silencing (Chodavarapu et al. 2010). Meanwhile, we also examined the histone modification data of WA09-hESC from the NIH Roadmap Epigenomics Project. Indeed, significant enrichment with only H3K4me3 is associated with the NP pattern of actively transcribed genes (a prominent NDR and well positioned nucleosomes respectively upstream and downstream of the TSS; see Figures 2 and 3), whereas significant enrichment with only H3K27me3 is associated with that of silent genes. Those with enrichment in both histone marks display a NP pattern of poised genes. Hence, these observations support the accuracy of our NP sequencing and data analysis pipeline.

Genes silenced upon differentiation displayed substantial promoter NP changes, while those activated upon differentiation showed barely any NP changes.

For a total of 118 genes that were actively expressed (with expression intensity of >500) in WA09-hESC but whose expression decreased by at least 2-folds upon differentiation, we observed a clear NP change as demonstrated by the increasingly diminishing NDR and well-positioning nucleosomes flanking the TSS (Figure 2.4). However, for 240 total genes whose expression increased by at least 2-folds upon differentiation to reach an intensity of >500 in both WA09-INM and WA09-SMC, no such a clear change was observed, although the NDR appeared to be slightly more prominent in WA09-SMCs. As a control, we did not observe changes for the genes whose expression remained largely constant for any of the six active or silent groups shown in Figure 2.2.

Puzzled by this NP change difference between gene silencing and activating described above, we explored the mRNA half-life of these genes. The rationale is that gene expression (or mRNA abundance) level, measured via microarray analysis here, is determined by both mRNA production (or transcription which is regulated via the promoter NP and other mechanisms) and degradation (which can be assessed with mRNA half-life). Because mRNA half-life values in hESCs have not been globally determined, we studied their mouse homologues in mouse ESCs already published (Sharova et al. 2009). Interestingly, genes activated upon differentiation were found to have a significantly longer mRNA half-life than those silenced on average ($p<0.01$; see Figure 2.4). This indicates that the expression level of silencing genes (but not activating genes) may be regulated primarily via promoter chromatin modeling, including NP. While consistent with yeast findings which point to an association between a gene's mRNA half-life and its promoter sequence (Bregman et al. 2011; Trcek et al. 2011), the significance and reasons of our observations clearly need further studies.

WA09-hESC's genome appears normal with no large-scale changes found and with a sequence mutation rate as low as normal human genomes

With the sequences of randomly sheared genomic DNA fragments, we investigated potential structural and sequence variations in the genome of WA09-hESC. By examining paired-end sequence read information (see Methods), we did not detect any structural rearrangements such as translocations and inversions. We also explored copy number variations (CNVs) and found comparable amount of CNVs in the genome of WA09-hESC, of an Middle East - East European ancestry (International Stem Cell et al. 2011), as in a normal genome of an European ancestry (NA12892) that was sequenced to approximately the same coverage (11.8X). For point/oligo-base variations, we identified about 4 million single nucleotide polymorphisms (SNPs) when compared to the

NCBI36/hg18 reference human genome (Figure 2.5). This is analogous to two normal genomes derived from blood samples, NA12892 and an Asian genome that was also sequenced to a similar coverage (Wang et al. 2008). Critically, similar to the two normal human genomes, bases transitions (C↔T and G↔A) dominate base transversions (Figure 2.5A). In summary, the analysis reveals no significant structural or sequence changes occurring in the WA09-hESC genome, with the amount of variations comparable to those of the two normal human genomes examined, consistent with other studies (Funk et al. 2012).

Changes in nucleosomes *per se* - mononucleosomal DNA length decreased after the differentiation?

By examining the spanned genomic distance of the mapped sequence read pairs (~200 million per sample), we noticed that the median length of mononucleosomal DNA fragments is 163bp for WA09-hESC, 165bp for WA09-INM, and 155bp for WA09-SMC, and is about the same between autosomes and chromosome X within a cell type (Figure 2.5B). While WA09-hESC and WA09-INM share approximately the same median length ($p=1$), both differ ($p = 0$) from WA09-SMC. Furthermore, the shape of the fragment length distribution of WA09-SMC differs significantly ($p=0.03$) from that of WA09-hESC, but insignificantly ($p>0.08$) from that of WA09-INM (Figure 2.5B). These variations in mononucleosomal DNA length are unlikely due to experimental artifacts (e.g., differences in MNase-digestion), but instead is biologically relevant, as described below. Also note that the fragment lengths of WA09-SMC are similar to those of human lymphoblastoid cell lines undergone similar treatment (Gaffney et al. 2012).

The sequence signature of 147bp canonical nucleosome core.

Canonical nucleosome core sequence is ~147bp with A/T and G/C dinucleotides oscillating at about every 10bp, assisting the winding of the DNA molecule around the histone core that consists of 2 copies each of H3, H4, H2A, and H2B histone proteins (Luger et al. 1997; Richmond and Davey 2003). As shown in Figure 2.6, we indeed observed a clear oscillating pattern of such for MNase-digested fragments of 147bp, but not for randomly sheared fragments. In addition, within the same cell type, nearly identical patterns were observed between autosomes and the X-chromosome. Between the cell types, WA09-hESC resembles WA09-INM but slightly differs from WA09-SMC. Moreover, consistent with the difference in the fragment length distribution described above, WA09-SMCs have 4-5 times more fragments of 147bp (Figure 2.6).

The patterns shown in Figure 2.6 are consistent with those of nonmammalian species (Ercan et al. 2011), with the exception that the frequency of A/T dinucleotides is twice as high as that of G/C ones in our case. This is likely caused by a higher C→T/G→A background mutation rate in the human genome (Venter et al. 2001).

MNase-digested mononucleosomal fragments of all sizes are symmetric in their A/T or G/C dinucleotide oscillating frequency.

As the fragment length increases or decreases away from the canonical nucleosome core length of 147bp, the 10bp periodicity shown in Figure 2.6 gradually diminishes. However, for MNase-digested mononucleosomal fragments of all length examined (135-180bp) in each cell type, the A/T and G/C dinucleotide frequency remains remarkably symmetric from the center of the fragmen. Critically, as the fragment length increases, the G/C dinucleotide frequency also rises from 16% to over 21%, while A/T dinucleotide frequency decreases from 35% to below 28%. For randomly sheared fragments, however, no such symmetry and dinucleotide frequency changes were observed.

The GC content and %CpG increase along with the length of MNase-digested mononucleosomal fragments of 147-200bp.

We further examined the GC content and %CpG in each DNA fragment. Interestingly, for MNase-digested mononucleosomal fragments, both numbers stay constant (at approximately 40% in GC content and 1% CpG) from 90bp to about 147bp (the length of canonical nucleosome core DNA), and then begin to increase until 200bp (reaching ~50% in GC content and 2% CpG in autosomes) (Figure 2.5B). This again differs from randomly sheared fragments, where the GC content and %CpG are both lower than mononucleosomal fragments once the length exceeding 147bp (Figure 2.5B).

Nucleosome and linker length estimation.

We stacked all fragments mapped to a certain genomic window as described (Valouev et al. 2011; Teif et al. 2012). As shown in Figure7, the start and end points of MNase-digested mononucleosomal fragments form clear peaks, which however are completely missing for randomly sheared fragments. Based on these, we estimated the length of nucleosomal DNA to be approximately 165bp, 165bp and 155bp while the linker to be 26bp, 26bp and 36bp long for WA09-hESC, WA09-INM and WA09-SMC, respectively, on average. Thus, in our model the nucleosomal DNA of both WA09-hESC and WA09-INM is about 10bp longer, whereas their linker is about 10bp shorter, when compared to WA09-SMC. Furthermore, the data indicate the same nucleosome repeat length (~191bp) among the three cell types (Figure 2.7), unlike a study with a mouse ESC to neural progenitor differentiation (Teif et al. 2012). The same observations were obtained for fragments mapped to the promoter regions (-5kb to +5kb of the TSS).

Histone genes are downregulated in the WA09-SMCs.

Histone variants have been reported to influence the nucleosome stability and the nucleosomal DNA length (Bonisch and Hake 2012; Li et al. 2012; Hasson et al. 2013; Hu et al. 2013). Consistent with this, we found significantly more histone genes downregulated in WA09-SMC than in WA09-hESC (p=1.46e-07) and WA09-INM (p=0.04), when compared to the entire gene set in the human genome. This change may contributes to the nucleosomal length variation among the three cell types described above (see Discussion).

MATERIALS AND METHODS

WA09-hESC differentiation, mononucleosome preparation, sequencing (MNase-seq) and analysis.

WA09-hESCs were maintained in StemPro defined media (Invitrogen). Differentiation to INM and SMCs was achieved by supplementation of defined media with Wnt3a (25 ng/ml) and BMP4 (50 ng/ml) for 4 and 21 days, respectively. Chromatin was processed by modifying previously published protocols (Ozsolak et al. 2007;

Schones et al. 2008; Spetman et al. 2011; Gaffney et al. 2012) with 200 units of MNase (Worthington Biochemical Corp.) and incubation at 25°C for 5 min to yield >95% mononucleosomal DNA. All DNA samples were required to have a 260/280 absorbance ratio around 1.8 for downstream applications. The about 150bp DNA fragments were gel purified (Spetman et al. 2011), and sequenced from both ends to yield 90bp end sequence pairs using Illumina Genome Analyzer at the BGI. As a control, genomic DNA was extracted following the same protocol but without MNase-digestion, randomly cut to 100-200bp fragments, and sequenced from both ends. All read pairs were then mapped to the human genome (hg18) using the Burrows-Wheeler Aligner (BWA) tool (Li and Durbin 2009) with the default parameters documented in the bwa-0.5.9 version. Over 94% read pairs were uniquely placed to the genome and were used for further analyses.

Permutation analysis of DNA fragment lengths.

Several 1Mb genomic regions were chosen (chr6: 32,000,001-33,000,000bp; chr9: 35,000,001-36,000,000bp; chr15: 20,000,001-21,000,000bp; and chrX: 40,000,001-41,000,000bp), with each having a fragment length distribution representative of its corresponding whole genome distribution (Figure 2.5). Then, the fragment lengths were randomly permutated between the cell types or autosomes and chromosome X within the same cell type, and Wilcoxon signed-rank statistics was calculated for each permutation. P-values were calculated based on 100,000 permutations.

Gene expression microarray study.

RNA was purified from approximately 5 million cells per sample using the Qiagen RNeasy Plus Mini kit (Cat. No. 74134). Then, high quality (a 260/280 absorbance ratio of ~2.0, undegraded, and free of genomic DNA contamination) samples were analyzed using the Affymetrix Human Gene 1.0 ST array with biological replicates. The moderated t-test implemented in the 'limma' package (Smyth 2004) was used to identify differnetnially expressed genes between the cell types, and the p-values were adjusted for multiple-hypothesis testing with the Benjamini and Hochberg method (Benjamini et al. 2001). PCA analysis was performed using R (www.R-project.org). The packages used are available at the Bioconductor (www.bioconductor.org).

Nucleosome occupancy calculation.

The KnownGene annotation (hg18) downloaded from the UCSC genome database (genome.ucsc.edu) was used to match the genes of the Affymetrix Human Gene 1.0 ST array, and 33,271 transcripts (17,592 genes) with more than 90% overlapping coordinates were chosen for the nucleosome occupancy analysis shown in Figure 2.2. Nucleosome occupancy for the base *i* in the genome was estimated by $\log_2 \frac{c_{mi}/\bar{c}_m}{c_{gi}/\bar{c}_g}$, where c_{mi} and c_{gi} respectively represent the total count of mononucleosomal fragments (m) and randomly sheared genomic fragments (g) covering the base i, whereas $\bar{c}_m$and $\bar{c}_g$ represent the corresponding genome-wide count average of c_{mi} and c_{gi}. The TSS, exon-intron, intron-exon junction, and gene end were extracted from the UCSC KnownGene annotation. Average GC content for the corresponding region was calculated based on the hg18 reference genome.

Promoter CpG methylation and histone modification analysis.

Bisulfate sequencing data of WA09-hESC were obtained from a published study (Laurent et al. 2010). Both H3K4me3 (GSM605316) and H3K27me3 (GSM706066 and GSM667622) ChIP-seq data for WA09-hESC were downloaded from NIH Roadmap Epigenomics (www.roadmapepigenomics.org), and mapped to the hg18 genome with BWA using the default parameters. Uniquely mapped reads were selected for further analyses.

Gene functional analysis and mRNA half-life.

Gene functional annotation and enrichment were analyzed by DAVID (Huang da et al. 2009). Mouse mRNA half-life data were obtained from a study (Sharova et al. 2009) and the mouse-human gene conversion were achieved using the Human and Mouse Orthology file obtained from Mouse Genome Database (www.informatics.jax.org). As a result, mRNA half-life was assigned to13,578 human genes.

Mononucleosomal DNA dinucleotide frequency determination.

Mononucleosomal fragments with both reads perfectly matching to the hg18 genome were identified. Then, the corresponding reference sequences of a chosen mononucleosomal length were cut out and aligned. Lastly, fractions of AA/AT/TT/TA and CC/CG/GC/GG dinucleotides at each base position from dyad were calculated and plotted as shown Figure 2.6.

Structural and sequence variation analysis of the WA09-hESC genome.

The sequences of the randomly sheared genomic DNA of WA09-hESC (gWA09-hESC), along with genomic sequences of an European ancestry (NA12892) downloaded from www.1000genomes.org and of an Asian (Wang et al. 2008), were mapped to the hg18 genome with BWA. SNPs, compared to the hg18 reference genome, were identified using SAMtools (Li et al. 2009) and GATK (McKenna et al. 2010). Copy number variations were identified as described (Tang et al. 2010; Abyzov et al. 2011), by normalizing the mapped fragment density at each 200bp window by $\log_2 \frac{d_{ti}/\bar{d}_t}{d_{ai}/\bar{d}_a}$, where d_{ti} and d_{ai} respectively represent the fragment density of the window *i* of the test genome (t, either WA09-hESC or NA12892) and of the normalizing Asian genome, whereas $\bar{d}_t$ and $\bar{d}_a$ represent the corresponding genome-wide average of d_{ti} and d_{ai}.

Length estimation for nucleosome repeat, nucleosomal DNA, and linker.

First, at each base position within a certain genomic region (e.g., 2kb) starting from the 1st base or last base of each mapped fragment, we counted the frequency of other mononucleosomal fragments with their start or end point mapped to the location as described (Valouev et al. 2011; Teif et al. 2012). Both frequency counts formed peaks and valleys, allowing the estimation of nucleosome repeat, nucleosomal DNA, and linker lengths (Figure 2.7). To avoid complication, fragments mapped to within 1Mb regions from the nearest centromeres and telomeres were excluded (Stephens et al. 2009), and for regions with multiple fragments identically mapped, only one fragment was selected as described (Zhang et al. 2008).

Acknowledgement

We thank the Emory Biomarker Core for conducting the microarray experiments, as well as the BGI for the sequencing work. The study was funded by GM085354 (PI: Dr. Stephen Dalton).

FIGURE LEGENDS

Figure 2.1. Gene expression analysis of the three cell types. A) The principle component analysis (PCA) was performed with the entire transcript set (22,089 in total) included in the Affymetrix Human Gene 1.0 ST array. **B**) The enriched functional groups of genes with significant expression changes (red: upregulated; green: downregulated) among the three cell types is shown.

Figure 2.2. Nucleosome occupancy and GC content at gene promoters, exon-intron and intron-exon junctions, and gene ends. Genes were divided into six groups, represented by the six colored lines as shown, based on their expression intensities obtained from the microarray analysis Genes in each group were aligned at the indicated positions (e.g., 5kb upstream and downstream of the TSS). In each panel, plots on the left present the average nucleosome occupancy status, represented by the mononucleosomal fragment density normalized against the density of the randomly sheared genomic DNA fragments (gWA09-hEC) in $\log_2$ scale (see Methods), at each position indicated (the X-axis). Plots on the right indicate the average GC content of the corresponding sequences.

Figure 2.3. Nucleosome occupancy, CpG methylation, and histone modification at promoter regions in WA09-hESC. **A**) Genes were sorted into five groups based on their promoter CpG methylation status as described (Laurent et al. 2010). Specifically, methylated (M), between methylated and partially methylated (M_P), partially methylated (P), between partially methylated and unmethylated (U_P), or unmethylated (U) correspond to >80%, 60-80%, 40-60%, 20-40%, or <20% of CpGs being methylated

respectively. **B)** Genes were sorted into three groups: open, poised and closed, based on their relative enrichment with H3K4me3 and H3K27me3 as shown. In each plot, the Y-axis represents the average nucleosome occupancy status of each group at each base pair position (the X-axis) as described in Figure 2. The tables indicate the number of genes further categorized based on their expression intensity (the top row) for each group shown in the corresponding plot above.

Figure 2.4. Nucleosome occupancy at promoter regions of genes down/up-regulated upon differentiation. A) Top panel presents the nucleosome occupancy for genes with expression intensity of >500 in WA09-hESC (black) but decreased by >2-folds in both WA09-INM (red) and WA09-SMC (green). Bottom panel presents the nucleosome occupancy for genes with expression intensity increased by >2-folds upon differentiation to reach >500 in both WA09-INM and WA09-SMC. **B**) The mRNA half-life distribution of the same downregulated and upregulated genes as shown in **A**. The nucleosome occupancy is presented as described in Figure 2.

Figure 2.5A. Base substitution types of the WA09-hESC genome (gWA09-hESC) in comparison with two other normal human genomes. The base changes were identified by comparing the relevant genomic sequences to the hg18 human genome. EUR: the genome of an European ancestry. Asian: the Asian genome published (Wang et al. 2008_ENREF_50).

Figure 2.5B. Fragment length and sequence content distribution. Uniquely and perfectly mapped fragments were used to plot their average length distribution at a 5bp window from 90bp to 219bp. Average GC content and %CpG were calculated in the same window.

Figure 2.6. A/T and G/C dinucleotide frequencies for 147bp DNA fragments. The 147bp mononucleosomal DNA fragments (WA09-hESC, WA09-INM, and WA09-SMC) or randomly sheared genomic DNA (gWA09-hESC) were aligned. Then, the average frequencies of AA/AT/TA/TT (black) and CC/CG/GC/GG (red) dinucleotides were computed at each base pair position and presented as the left and right Y-axis, respectively. The X-axis represents the relative base pair coordinates from the center of the fragment. **A**: Autosomes; **B**: chromosome X.

Figure 2.7. Estimation of nucleosome and linker lengths of the three cell types. A) **Top**: The plots indicate the frequency (the Y-axis) of the start (solid lines) and end (dashed lines) of randomly sheared gWA09-hESC (the 1st plot), or mononucleosomal fragments (the 2nd and 3rd plots with WA09-hESC in black, WA09-INM in red, and WA09-SMC in green) mapped to each base pair position (the X-axis) within a 1kb region starting from the 1st base of each mapped fragment. Peaks of the solid and dashed lines, which were respectively marked by orange and blue vertical lines in WA09-hESC, allowed the estimation of the length of nucleosomal and linker DNA. **Bottom**: two hypotheses are proposed. Hypothesis I: the nucleosome core DNA of WA09-hESC and WA09-INM is about 10bp longer than WA09-SMC (165b vs. 155bp). Hypothesis II: the nucleosome core DNA length is identical (~155bp) among the three cell types. However,

in WA09-hESC and WA09-INM, about 10bp linker adjacent to the nucleosome is protected by other proteins (pink) from MNase-digestion.

B) The A/T and G/C dinucleotide frequencies were plotted for fragments of 165bp (those mapped to the 378-542bp region in **A**) for WA09-hESC (523,790 in total) and WA09-INM (619,575 in total), as well as those of 155bp (mapped to the 378-532bp region in **A**) for WA09-SMC (585,540 in total) as described in Figure 2.6. As a control, randomly sheared gWA09-hESC DNA fragments mapped to the same regions (65,711 and 16,537 in total mapped to the 378-542bp and 378-532bp regions respectively in **A**) were also plotted.

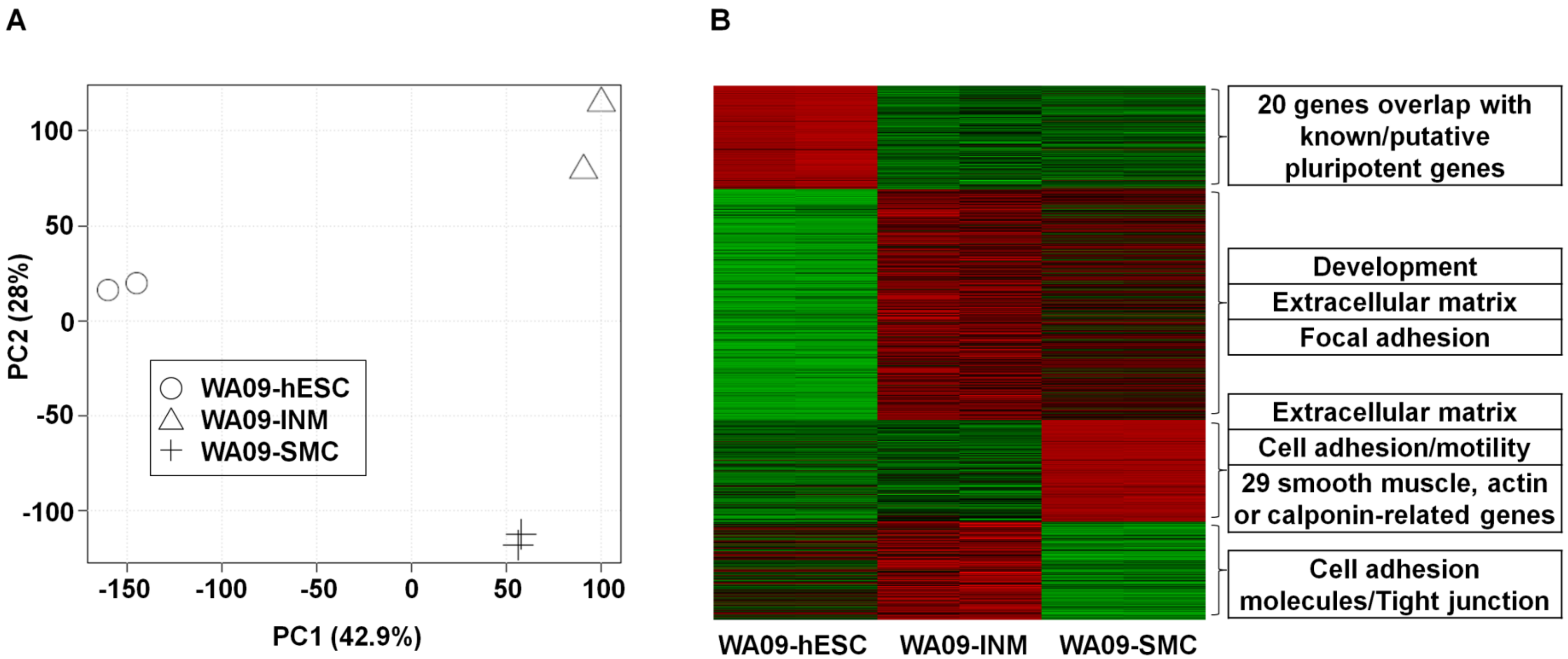
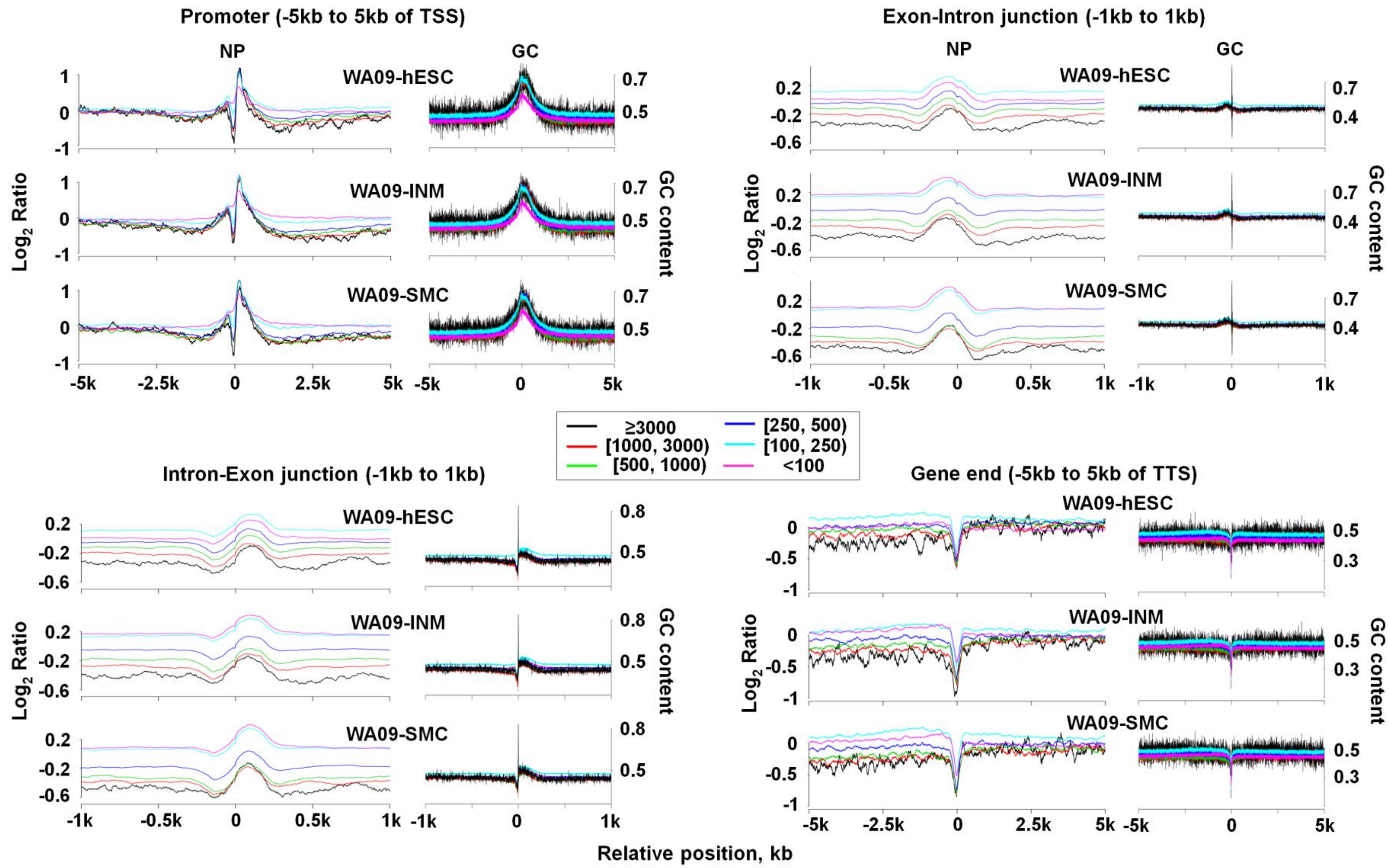

A
B
100
50
0
-50
-100
PC2 (28%)
-150
-100
-50
0
50
100
PC1 (42.9%)
WA09-hESC
WA09-INM
WA09-SMC
20 genes overlap with known/putative pluripotent genes
Development
Extracellular matrix
Focal adhesion
Extracellular matrix
Cell adhesion/motility
29 smooth muscle, actin or calponin-related genes
Cell adhesion molecules/Tight junction
WA09-hESC
WA09-INM
WA09-SMC

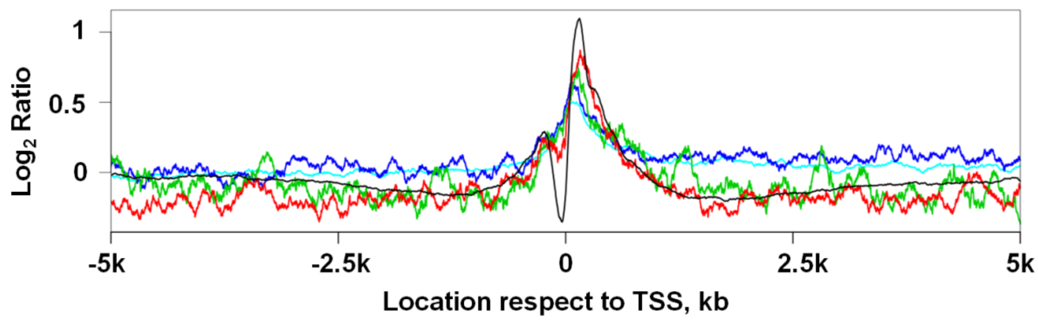
Figure 2.1



Promoter (-5kb to 5kb of TSS)
Exon-Intron junction (-1kb to 1kb)
Intron-Exon junction (-1kb to 1kb)
Gene end (-5kb to 5kb of TTS)
NP
GC
WA09-hESC
WA09-INM
WA09-SMC
Log2 Ratio
GC content
≥3000
[1000, 3000)
[500, 1000)
[250, 500)
[100, 250)
<100
Relative position, kb

Figure 2.2

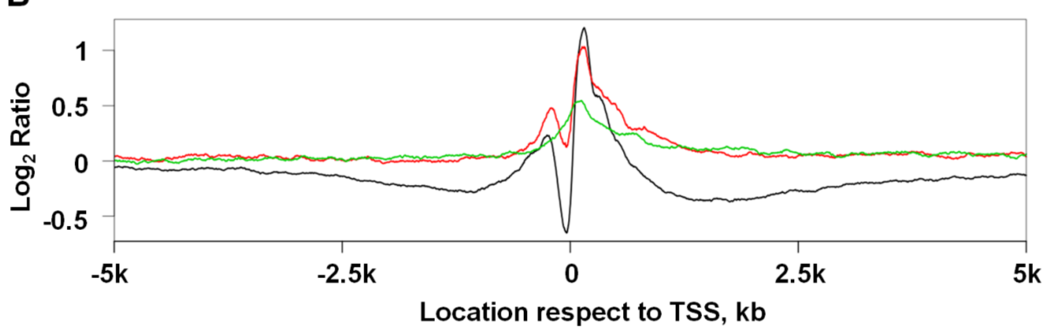
A



Number of genes grouped based on expression and DNA methylation

	≥3000	[1000,3000)	[500,1000)	[250,500)	[100,250)	<100
M	6	28	77	175	666	2906
M_P	1	6	9	15	71	344
P	0	2	9	11	32	76
U_P	2	14	21	56	84	103
U	174	967	1717	2771	4295	3333

B



Number of genes grouped based on expression and histone modification

Log$_2$(H3K4me3/H3K27me3)	≥3000	[1000,3000)	[500,1000)	[250,500)	[100,250)	<100
≥3	161	874	1494	2211	2379	727
[-1,3)	7	94	238	630	1805	1764
<-1	9	46	100	218	1001	4186

Figure 2.3

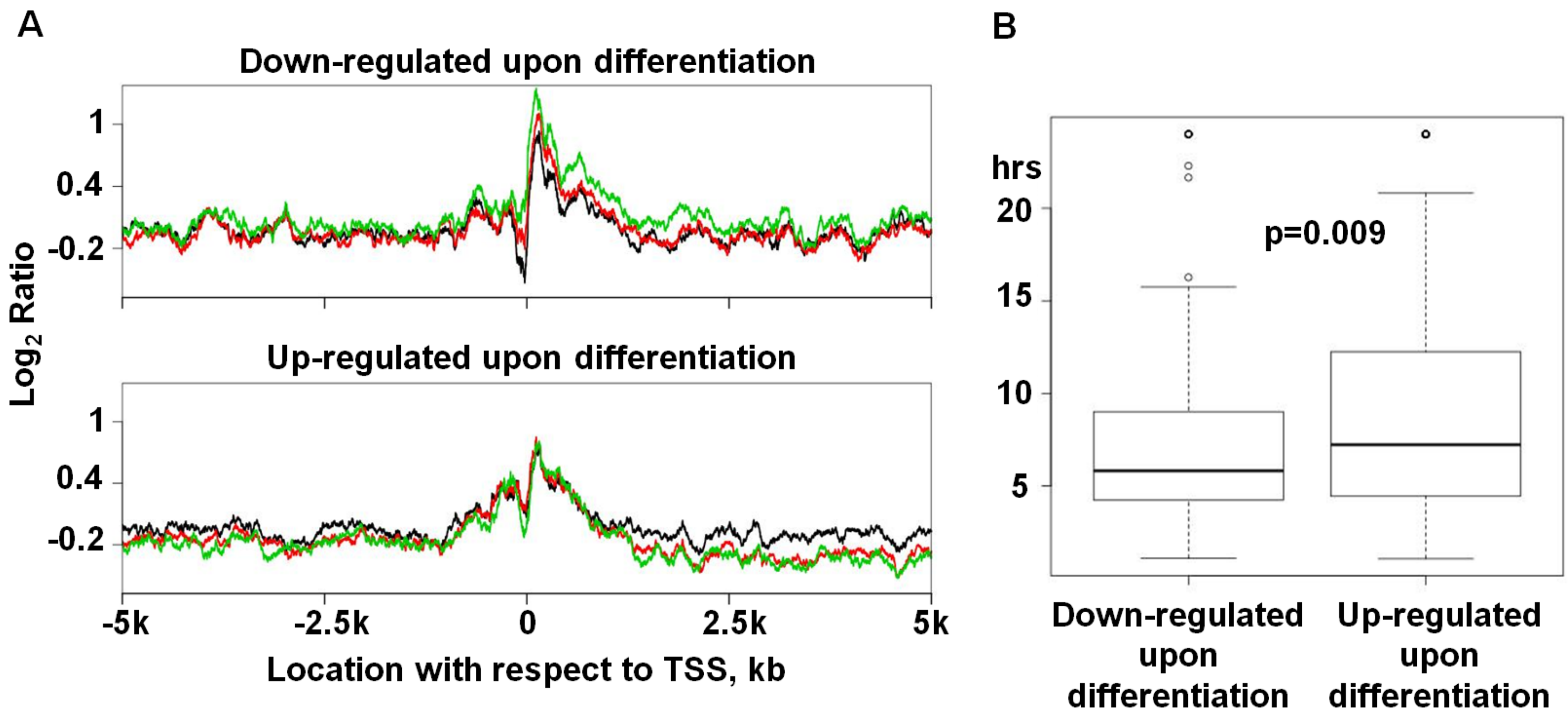

A
Down-regulated upon differentiation
1
0.4
-0.2
Log2 Ratio
Up-regulated upon differentiation
1
0.4
-0.2
-5k
-2.5k
0
2.5k
5k
Location with respect to TSS, kb
B
hrs
20
15
10
5
p=0.009
Down-regulated upon differentiation
Up-regulated upon differentiation

Figure 2.4

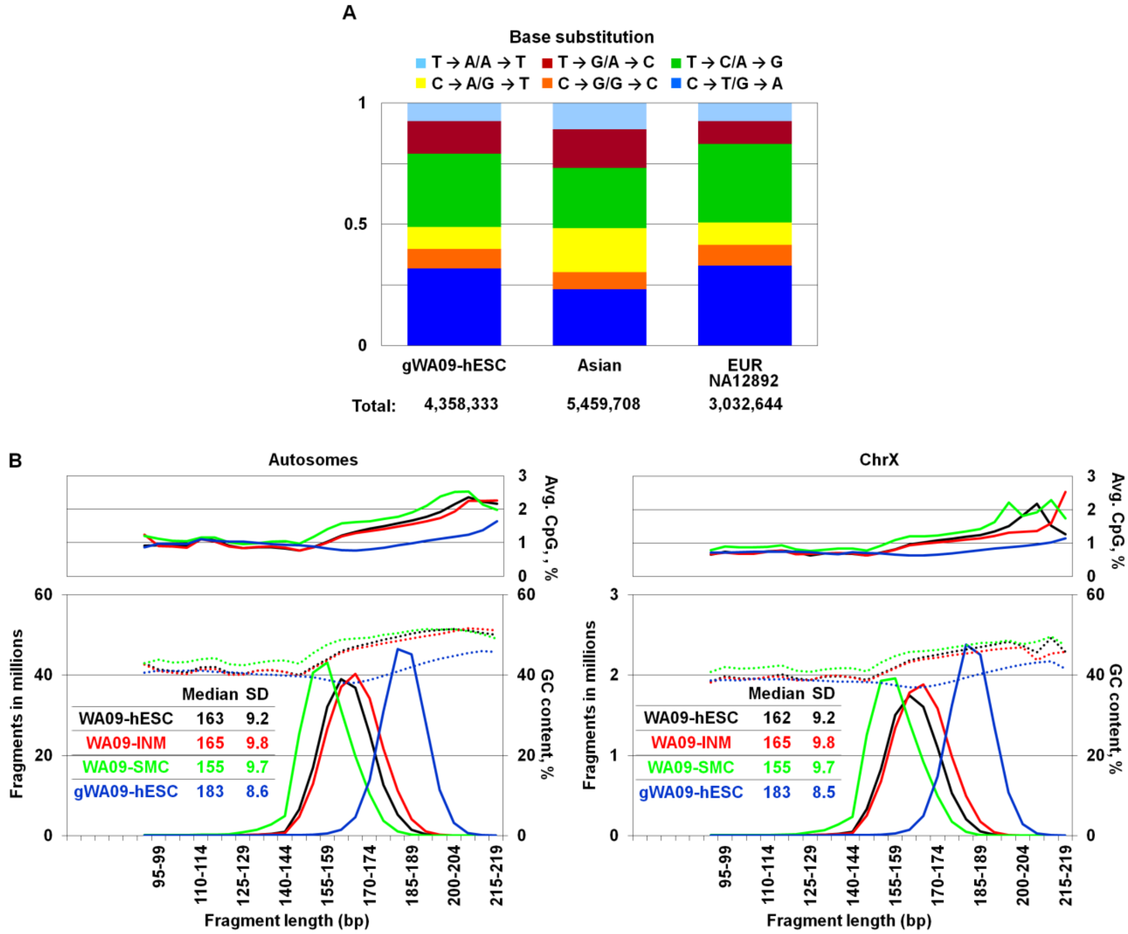

A
Base substitution
T → A/A → T
T → G/A → C
T → C/A → G
C → A/G → T
C → G/G → C
C → T/G → A
1
0.5
0
gWA09-hESC
Asian
EUR NA12892
Total:
4,358,333
5,459,708
3,032,644
B
Autosomes
ChrX
Avg. CpG, %
Fragments in millions
GC content, %
Median SD
WA09-hESC 163 9.2
WA09-INM 165 9.8
WA09-SMC 155 9.7
gWA09-hESC 183 8.6
WA09-hESC 162 9.2
WA09-INM 165 9.8
WA09-SMC 155 9.7
gWA09-hESC 183 8.5
95-99
110-114
125-129
140-144
155-159
170-174
185-189
200-204
215-219
Fragment length (bp)

Figure 2.5

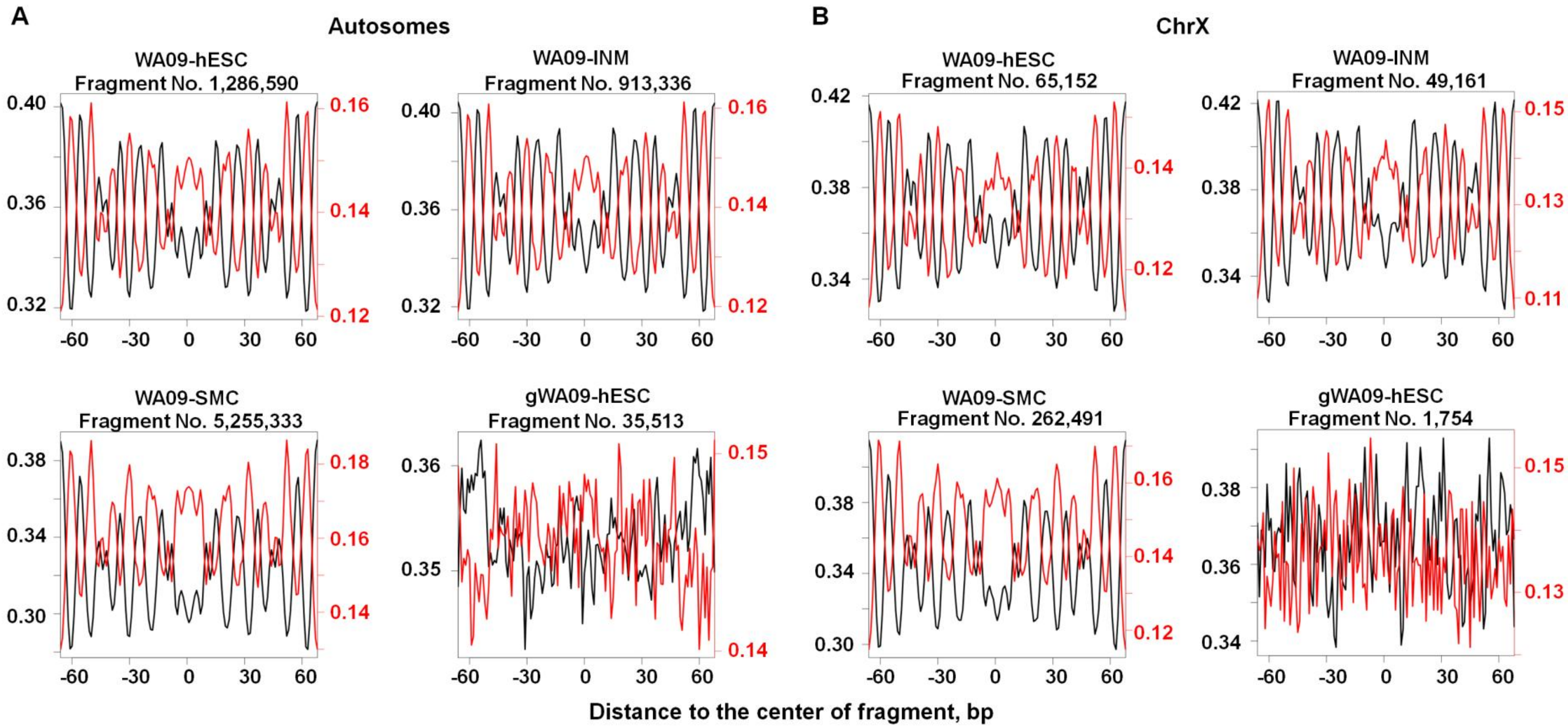

A
Autosomes
B
ChrX
WA09-hESC
Fragment No. 1,286,590
WA09-INM
Fragment No. 913,336
WA09-hESC
Fragment No. 65,152
WA09-INM
Fragment No. 49,161
WA09-SMC
Fragment No. 5,255,333
gWA09-hESC
Fragment No. 35,513
WA09-SMC
Fragment No. 262,491
gWA09-hESC
Fragment No. 1,754
Distance to the center of fragment, bp

Figure 2.6

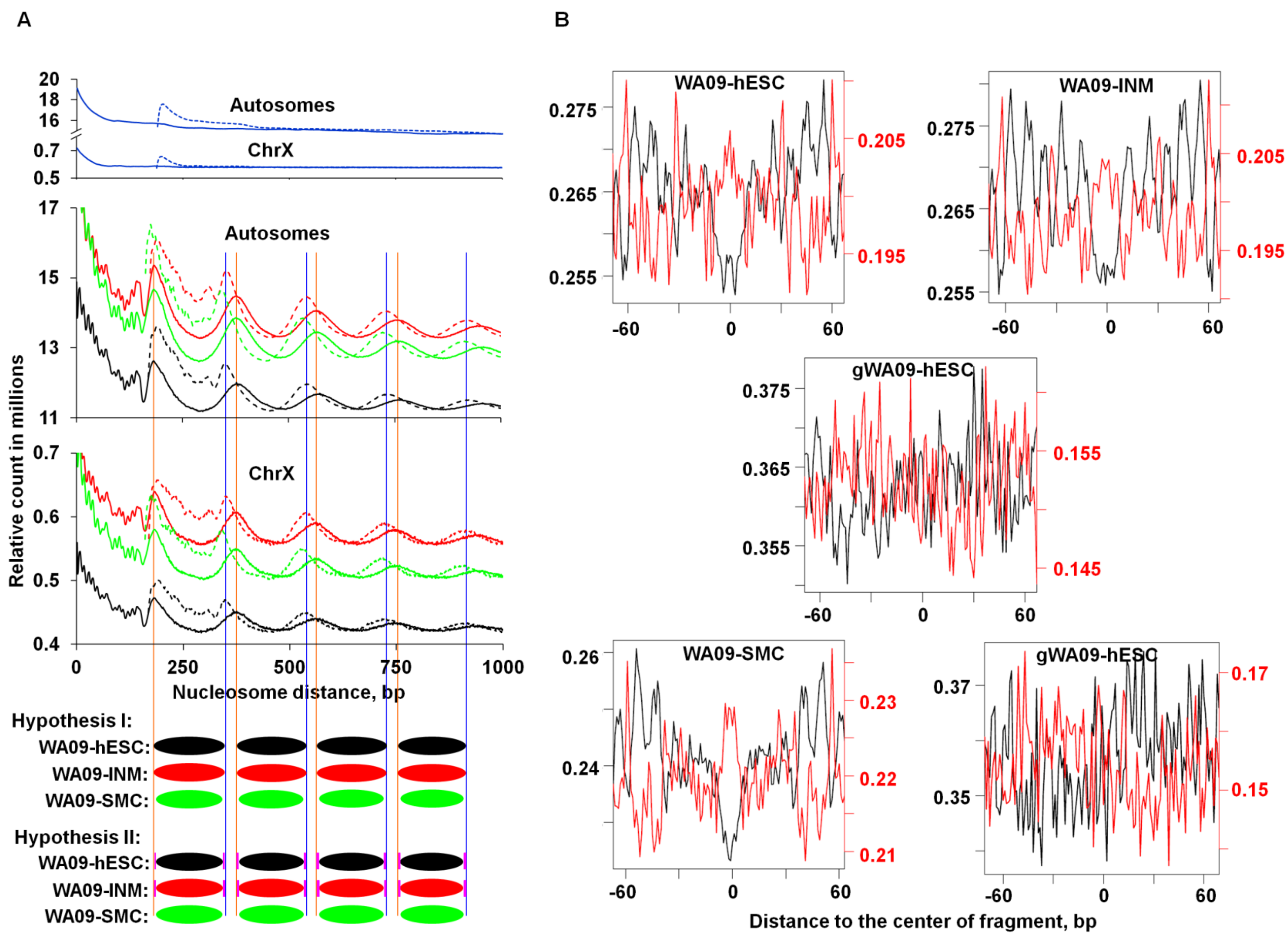

A
B
Autosomes
ChrX
Autosomes
ChrX
Relative count in millions
Nucleosome distance, bp
Hypothesis I:
WA09-hESC:
WA09-INM:
WA09-SMC:
Hypothesis II:
WA09-hESC:
WA09-INM:
WA09-SMC:
WA09-hESC
WA09-INM
gWA09-hESC
WA09-SMC
gWA09-hESC
Distance to the center of fragment, bp

Figure 2.7

REFERENCES

Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome research* **21**(6): 974-984.

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. 2001. Controlling the false discovery rate in behavior genetics research. *Behavioural brain research* **125**(1-2): 279-284.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**(2): 315-326.

Bonisch C, Hake SB. 2012. Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic acids research* **40**(21): 10719-10741.

Bregman A, Avraham-Kelbert M, Barkai O, Duek L, Guterman A, Choder M. 2011. Promoter elements regulate cytoplasmic mRNA decay. *Cell* **147**(7): 1473-1483.

Chen X, Chen Z, Chen H, Su Z, Yang J, Lin F, Shi S, He X. 2012. Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. *Science* **335**(6073): 1235-1238.

Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* **466**(7304): 388-392.

Consortium EP Dunham I Kundaje A Aldred SF Collins PJ Davis CA Doyle F Epstein CB Frietze S Harrow J et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**(7414): 57-74.

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. 2009. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell stem cell* **4**(1): 80-93.

Ercan S, Lubling Y, Segal E, Lieb JD. 2011. High nucleosome occupancy is encoded at X-linked gene promoters in C. elegans. *Genome research* **21**(2): 237-244.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**(7347): 398-402.

Funk WD, Labat I, Sampathkumar J, Gourraud PA, Oksenberg JR, Rosler E, Steiger D, Sheibani N, Caillier S, Stache-Crain B et al. 2012. Evaluating the genomic and sequence integrity of human ES cell lines; comparison to normal genomes. *Stem cell research* **8**(2): 154-164.

Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, Widom J, Gilad Y, Pritchard JK. 2012. Controls of nucleosome positioning in the human genome. *PLoS genetics* **8**(11): e1003036.

Galan A, Diaz-Gimeno P, Poo ME, Valbuena D, Sanchez E, Ruiz V, Dopazo J, Montaner D, Conesa A, Simon C. 2013. Defining the Genomic Signature of Totipotency and Pluripotency during Early Human Development. *PloS one* **8**(4): e62135.

Gopalakrishnan S, Van Emburgh BO, Robertson KD. 2008. DNA methylation in development and human disease. *Mutation research* **647**(1-2): 30-38.

Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N, Alonso A, Warburton PE, Black BE. 2013. The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nature structural & molecular biology*.

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

Hu G, Cui K, Northrup D, Liu C, Wang C, Tang Q, Ge K, Levens D, Crane-Robinson C, Zhao K. 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 da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**(1): 44-57.

International Stem Cell I Amps K Andrews PW Anyfantis G Armstrong L Avery S Baharvand H Baker J Baker D Munoz MB et al. 2011. Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. *Nature biotechnology* **29**(12): 1132-1144.

Jiang C, Pugh BF. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nature reviews Genetics* **10**(3): 161-172.

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G et al. 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell stem cell* **8**(2): 200-213.

Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J et al. 2010. Dynamic changes in the human methylome during differentiation. *Genome research* **20**(3): 320-331.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.

Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J, Li W, Kaestner KH. 2012. Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell* **151**(7): 1608-1616.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**(7271): 315-322.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**(6648): 251-260.

Maitra A, Arking DE, Shivapurkar N, Ikeda M, Stastny V, Kassauei K, Sui G, Cutler DJ, Liu Y, Brimble SN et al. 2005. Genomic alterations in cultured human embryonic stem cells. *Nature genetics* **37**(10): 1099-1103.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**(9): 1297-1303.

Menendez L, Kulik MJ, Page AT, Park SS, Lauderdale JD, Cunningham ML, Dalton S. 2013. Directed differentiation of human pluripotent cells to neural crest stem cells. *Nature protocols* **8**(1): 203-212.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153): 553-560.

Ozsolak F, Song JS, Liu XS, Fisher DE. 2007. High-throughput mapping of the chromatin structure of human promoters. *Nature biotechnology* **25**(2): 244-248.

Prendergast JG, Semple CA. 2011. Widespread signatures of recent selection linked to nucleosome positioning in the human lineage. *Genome research* **21**(11): 1777-1787.

Richmond TJ, Davey CA. 2003. The structure of DNA in the nucleosome core. *Nature* **423**(6936): 145-150.

Ruzov A, Tsenkina Y, Serio A, Dudnakova T, Fletcher J, Bai Y, Chebotareva T, Pells S, Hannoun Z, Sullivan G et al. 2011. Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell research* **21**(9): 1332-1342.

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**(5): 887-898.

Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J. 2006. A genomic code for nucleosome positioning. *Nature* **442**(7104): 772-778.

Sharova LV, Sharov AA, Nedorezov T, Piao Y, Shaik N, Ko MS. 2009. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA research : an* *international journal for rapid publication of reports on genes and genomes* **16**(1): 45-58.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* **3**: Article3.

Spetman B, Lueking S, Roberts B, Dennis JH. 2011. Microarray mapping of nucleosome position. *Epigenetics: A Reference Manual J Craig and N Wong, eds, Horizon Scientific Press, Norwich, UK*.

Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ et al. 2009. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**(7276): 1005-1010.

Struhl K, Segal E. 2013. Determinants of nucleosome positioning. *Nature structural & molecular biology* **20**(3): 267-273.

Tang J, Le S, Sun L, Yan X, Zhang M, Macleod J, Leroy B, Northrup N, Ellis A, Yeatman TJ et al. 2010. Copy number abnormalities in sporadic canine colorectal cancers. *Genome research* **20**(3): 341-350.

Teif VB, Vainshtein Y, Caudron-Herger M, Mallm JP, Marth C, Hofer T, Rippe K. 2012. Genome-wide nucleosome positioning during embryonic stem cell development. *Nature structural & molecular biology* **19**(11): 1185-1192.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* **282**(5391): 1145-1147.

Trcek T, Larson DR, Moldon A, Query CC, Singer RH. 2011. Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell* **147**(7): 1484-1497.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. *Nature* **474**(7352): 516-520.

Venter JC Adams MD Myers EW Li PW Mural RJ Sutton GG Smith HO Yandell M Evans CA Holt RA et al. 2001. The sequence of the human genome. *Science* **291**(5507): 1304-1351.

Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y et al. 2008. The diploid genome sequence of an Asian individual. *Nature* **456**(7218): 60-65.

Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & development* **25**(7): 679-684.

Xiao S, Xie D, Cao X, Yu P, Xing X, Chen CC, Musselman M, Xie M, West FD, Lewin HA et al. 2012. Comparative epigenomic annotation of regulatory DNA. *Cell* **149**(6): 1381-1392.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**(9): R137.

Zhao XD, Han X, Chew JL, Liu J, Chiu KP, Choo A, Orlov YL, Sung WK, Shahab A, Kuznetsov VA 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.

Zlatanova J, Bishop TC, Victor JM, Jackson V, van Holde K. 2009. The nucleosome family: dynamic and growing. *Structure* **17**(2): 160-171.

CHAPTER 3

CHANGES IN PROMOTER NUCLEOSOME POSITIONING IN MOUSE HEPARIN SULFATE DEFICIENT EMBRYONIC STEM CELLS

INTRODUCTION

Embryonic stem cells (ESCs) are pluripotent, self-renewing cells derived from inner cell mass (ICM) of the developing blastocyst-stage embryo. They can be propagated in culture in an undifferentiated state through a process of self-renewal while maintaining the ability to form any fully differentiate cell of the body. (Thomson et al. 1998; Takahashi and Yamanaka 2006). Heparan sulfate (HS) is a highly sulfated glycosaminoglycan molecule that encoded by *Ext1* and *Ext2* genes in mouse ESCs and is abundant of the cell surface in both ESCs and differentiated cells (Holmborn et al. 2004). Several studies suggested that HS is involved in the regulation of ESC differentiation and cell lineage development accompanied by the HS structure change (Johnson et al. 2007; Baldwin et al. 2008). It also has been observed that HS deficient mouse ESCs retain the potential of self-renewal but lost the ability of multilineage differentiation (Kraushaar et al. 2010).

Epigenetic features of ESCs including characteristic histone modifications and DNA methylation patterns have been under intense investigation (Mikkelsen et al. 2007; Gopalakrishnan et al. 2008; Cui et al. 2009; Lister et al. 2009; Laurent et al. 2010; Ficz et al. 2011; Koh et al. 2011). However, another equally fundamental epigenetic mechanism, nucleosome positioning (NP) is relatively understudied (Ozsolak et al. 2007; Schones et al. 2008; Jiang and Pugh 2009; Prendergast and Semple 2011; Valouev et al. 2011; Gaffney et al. 2012; Teif et al. 2012; Struhl and Segal 2013). Recent study of global NP during mouse ESCs differentiation has suggested that NP was also involved in the regulatory layer of ESC differentiation, in correlation with certain histone modifications throughout the genome (Teif et al. 2012).

In previous studies, the NP at a genome scale in yeast (Yuan et al. 2005; Lee et al. 2007) and at a limited genomic regions in the human genome (Dennis et al. 2007; Ozsolak et al. 2007) have been determined using high density DNA arrays with tiling oligonucleotide probes. It has been proved to be a rapid and robust approach to determine the positioned nucleosome in the genome. In this study, we employed the same technique and identified genome wide promoter NP in the mouse genome for *Ext1*$^{+/+}$ and *Ext1*$^{-/-}$ ESCs, and studied how NP dynamics in promoters contribute to the maintenance of ESCs self-renewal after the *Ext1* gene knock out.

RESULTS

Expression analysis for *Ext1*$^{+/+}$ and *Ext1*$^{-/-}$ ESCs.

In order to study how HS biosynthesis deficient affect the expression pattern of mouse ES cells, we conducted microarray experiments with biological replicates to examine gene expression changes before and after the *Ext1* gene knockout, of which principle component analysis (PCA) indicated a clear separation between *Ext1*$^{+/+}$ and *Ext1*$^{-/-}$ ESCs (Figure 3.1A). Furthermore, our examination of individual genes is consistent with previous published study (Kraushaar et al. 2010) that *Ext1*$^{-/-}$ ESCs were deficient in *Ext1* expression but retained normal levels of *Ext2* mRNA, and pluripotent genes such as *Nanog*, *Sox2* and *Pou5f1*, remain in high level of expression in *Ext1*$^{-/-}$ ESCs. More comprehensively, we identified 447 genes with expression altered more than 2 folds in *Ext1*$^{-/-}$ ESCs. Among them, genes related to TGF-β pathway, cell motion, phospholipid binding, control of transcription and response to monosaccharide stimulus were downregulated significantly ($p<0.05$, see Figure 3.1B) while genes related to cell adhesion, ECM/glycoprotein, EFG-related and protease were upregulated significantly ($p<0.05$, see Figure 3.1B). Taken together, the gene expression analysis support the observation that *Ext1*$^{-/-}$ ESCs retained their normal self-renewal ability and HS formation is not require for ESCs self-renewal maintenance.

Examine nucleosome positioning in promoters with customized CGH arrays.

We treated the cells with micrococcal nuclease (MNase) that yielded >98% mononucleosomes, gel-purified the mononucleosomal DNA band (approximately 150bp) and hybridized with customized high density Comparative Genomic Hybridization (CGH) arrays. The CGH array covered the promoter regions (1kb upstream and 1kb downstream of transcriptions start sites (TSSs)) of chosen transcripts with 2.1 million 50-mer probes and 15bp spacing between adjacent probes. Collected array signals were denoised as described in previous study (Ozsolak et al. 2007) and further normalized among arrays with quantile normalization (see Methods).

NP pattern in promoters.

We partitioned promoters into four classes depending on their expression intensities as indicated in Figure 3.2. No obvious NP differences were noted between the same expression group of *Ext1*$^{+/+}$ and *Ext1*$^{-/-}$ ESCs. However, we observed a nucleosome depleted region (NDR) and well-positioned nucleosomes immediately upstream and downstream of the TSS respectively, the extent of which is directly correlated with the transcriptional activity of the genes (Figure 3.2). In addition, a correlation with the GC content of the sequence, albeit to a lesser degree, was also noted. We calculated the correlation between NP and GC content in different promoter regions, and found that GC content is anti-correlated with NP in the upstream of TSS while highly correlated with NP in the downstream of TSS. Our observations of NP in promoter regions are consistent with published studies at the promoter (Ozsolak et al. 2007; Schones et al. 2008; Valouev et al. 2011) that NP was correlated primarily with the transcriptional activity and secondarily with the GC content of the sequence. More specifically, our observations suggested that the major determinant of NP in the upstream of TSS is transcriptional activity whereas GC content contributes more in the downstream of TSS.

No NP change in promoters for gene with expression remains in the same group after *Ext1* gene knockout.

More than 85% of genes were remain in the same expression group after the *Ext1* gene knockout. Meanwhile, we did not find NP changes in these genes' promoters (Figure 3.3A). We observed that NP in active genes (Expression intensity>100) still exhibited a NDR region and positioned nucleosomes in the upstream and downstream of TSS respectively, where the depletion of nucleosomes is correlated with the transcriptional activity of genes. We did not observe clear NP patterns for silent genes (Expression intensity<100), which suggested the randomness of nucleosomes. These observations are consistent with NP patterns shown in Figure 3.2. To explore whether the overall NP distribution is representative of individual genes, we examined NP for individual genes of each expression group (Figure 3.3B): *Gpx1* has expression intensity of 3068±346; *Lmnb2* has expression intensity of 372±34; *Prex2* has expression intensity of 111±7 and *Synpo2* has expression intensity of 54±4 in both *Ext1*$^{+/+}$ and *Ext1*$^{-/-}$ ESCs. Our observation of NP of these individual genes suggested the same pattern as the groups they belong to, and no NP change was found after *Ext1* gene knockout.

Genes silenced after *Ext1* knockout displayed substantial promoter NP changes, while those activated after *Ext1* knockout showed barely any NP changes.

For a total of 278 transcripts that were actively expressed in *Ext1*$^{+/+}$ cells but whose expression decreased by at least 2-folds upon knockout, we observed a clear NP change as demonstrated by the increasingly diminishing NDR and well-positioning nucleosomes flanking the TSS (Figure 3.4A). However, for 189 total transcripts whose expression increased by at least 2-folds upon knockout in *Ext1*$^{-/-}$ ESCs, no such a clear change was found (Figure 3.4A). We explored the values of mRNA half-life of these genes owing to the NP change difference between gene silencing and activating described above. The rationale is that gene expression (or mRNA abundance) level, measured via microarray analysis here, is determined by both mRNA production (or transcription which is regulated via the promoter NP and other mechanisms) and degradation (which can be assessed with mRNA half-life). We integrated the mRNA half-life data that previously published (Sharova et al. 2009) from mouse ESCs. Interestingly, genes activated upon differentiation were found to have a significantly longer mRNA half-life than those silenced on average (Student's t-test: p=1.98e-6; see Figure 3.4B). As genes shown in Figure 3.4C, all four genes' expressions have been upregulated/downregulated more than two folds. However, we observed NP change for both *Anxa3* and *Hsh2d*, whose mRNA half-life are less than 5hrs, whereas we did not find NP change for both *Myo1g* and *Pla1a***,** whose mRNA half-life are longer than 22hrs.These observations indicate that the silencing genes' expression may be regulated primarily via the promoter chromatin modeling (including NP), whereas the activating genes' expression is determined largely through mRNA decay. While consistent with yeast findings which point to an association between the values of gene's mRNA half-life and its promoter sequence (Bregman et al. 2011; Trcek et al. 2011), the significance and reasons of our observations clearly need further studies.

MATERIAL AND METHODS

Gene expression analysis.

RNA was purified from Mouse *Ext1*$^{+/+}$ and *Ext1*$^{-/-}$ ESCs per sample using the Qiagen RNeasy Plus Mini kit (Cat. No. 74134). Then, samples were analyzed using Affymetrix MoGene 1.0ST arrays with biological replicates. PCA analysis was performed using R (www.R-project.org).

TSS verification.

We downloaded both the mouse knownGene (mm9) annotation from the UCSC genome database (genome.ucsc.edu) and mouse transcripts from Ensembl build 58 (useast.ensembl.org), which contained 49,409 and 85,345 transcripts respectively. Then we verified the annotated TSSs of transcripts as following: we downloaded previous published H3K4me3, Pol II and Pol II-Ser5p ChIP-seq data of mouse ES cell (Mikkelsen et al. 2007, Rahl et al. 2010) and mapped those reads to mouse genome (mm9) using Burrows-Wheeler Alignment Tool (BWA) (Li and Durbin 2009) with the default parameters documented in the bwa-0.5.9 version. We identified enriched segments as described before (Tang et al. 2010) and mapped these enriched segments to the promoters (1kb upstream and 1kb downstream of TSSs) of all transcripts. In total, 23,731 and 46,586 knownGene and Ensembl transcripts were verified respectively.

We also identified non-promoter regions with enriched H3K4me3, Pol II or Pol II-ser5p segments. Each region was masked by RepeatMasker (http://www.repeatmasker.org). We required those segments have size greater than 50bp with divergence greater than 25%. Then, we chose segments which have masked region less than 50% and contained neither tRNA nor satellite. We expanded those segments for 750bp on each direction as our array design regions.

Finally, we added 8,705 transcripts from Mouse gene 1.0ST array (Affymetrix) which were not included in our array design in the above steps.

High density CGH array design.

We defined 1kb upstream and 1kb downstream of the TSSs as our promoter regions and merged those overlapping coordinates. In total, we have 30,816 knownGene transcripts, 46,055 Ensembl transcripts with 31,303 unique TSSs and 2,679 segments included in the preliminary array design. Then we designed high density CGH array (Nimblegen) which contained 2.1 million 50-mer probes staggered in 15-base pair (bp) steps covering selected coordinates. Those promoters and segments with less than 50% coverage in the probe picking process were eliminated. The final array design contained 18,771 genomic regions with 30,177 knownGene transcripts, 45,673 Ensembl transcripts and 2,633 non-promoter segments with a total length of 41,157,806bp.

Mononucleosome preparation and preliminary data analysis.

Chromatin was processed by modifying previously published protocols (Ozsolak et al. 2007; Schones et al. 2008; Spetman et al. 2011; Gaffney et al. 2012) with 200 units of MNase (Worthington Biochemical Corp.) and incubation at 25°C for 5 min to yield >98% mononucleosomal DNA. All DNA samples were required to have a 260/280 absorbance ratio around 1.8 for downstream applications. The about 150bp DNA fragments were gel purified (Spetman et al. 2011), and hybridized to CGH arrays with sonicated genomic input DNA from matched cells. After collecting signals for each probe, we performed wavelet data denosing followed by quantile normalization using Wavelet Toolbox in Matlab and ‘limma’ package (Smyth 2004) as described in previous study (Sharova et al. 2009).

Nucleosome occupancy calculation.

We chose 22,371 knowGene transcripts and 17,960 Ensembl transcripts (13,789 genes) which have more than 90% overlapping coordinates with transcripts of the Mogene 1.0ST array for the nucleosome occupancy analysis shown in Figure 3.2. Nucleosome occupancy for each base pair was calculated by averaging the signal of all probes that cover this base pair. Average GC content for the corresponding region was calculated based on the mm9 reference genome.

Gene functional analysis and mRNA half-life.

Gene functional annotation and enrichment were analyzed by DAVID (Huang da et al. 2009) and Panther Classification System (http://www.pantherdb.org/). Mouse mRNA half-life data were obtained from previously published study (Sharova et al. 2009).

FIGURE LEGENDS

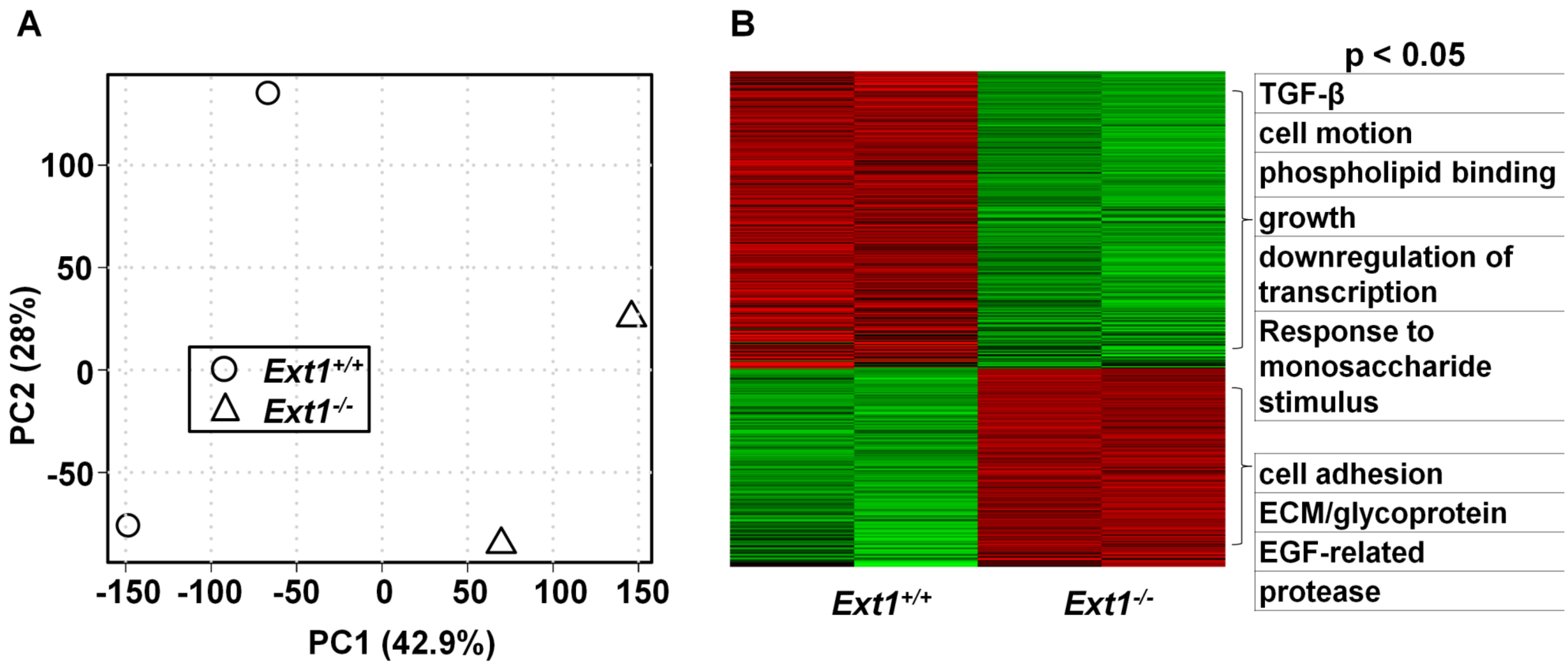
Figure 3.1. Gene expression analysis of the mouse *Ext1*$^{+/+}$ and *Ext1*$^{-/-}$ ESCs. A) The principle component analysis (PCA) was performed with the entire transcript set (25,137 in total) included in the Affymetrix MoGene 1.0ST array. B) The enriched functional groups of genes with more than two folds expression changes (red: upregulated; green: downregualted) between two cell types.

Figure 3.2. Nucleosome occupancy and GC content at gene promoters. Genes were divided into four groups, represented by the four colored lines as shown, based on their expression intensities obtained from the microarray analysis. Genes in each group were aligned at TSSs and average probe signal at each base pair was calculated. X-axis represents relative distance to the TSS; Y-axis on the left side represents the average NP signals in $\log_2$ scale; Y-axis on the right side indicates GC content. Solid lines represent the average nucleosome positioning signals, while dash lines represent average GC content. Ovals depict hypothetical nucleosome positions while the height of the ovals represents how well the nucleosome is positioned (See Figure s1 for biological replicates).

Figure 3.3. Nucleosome occupancy at promoters of genes with no significant change in expression intensities after *Ext1* gene knockout. A) From top to bottom, each plot presents the nucleosome occupancy for one of the four gene groups defined in Figure 2 and without expression changes after the *Ext1* gene knockout. No notable changes in NP were observed for these genes between *Ext1*$^{+/+}$ (red) and *Ext1*$^{-/-}$ (green) cells. **B)** Each

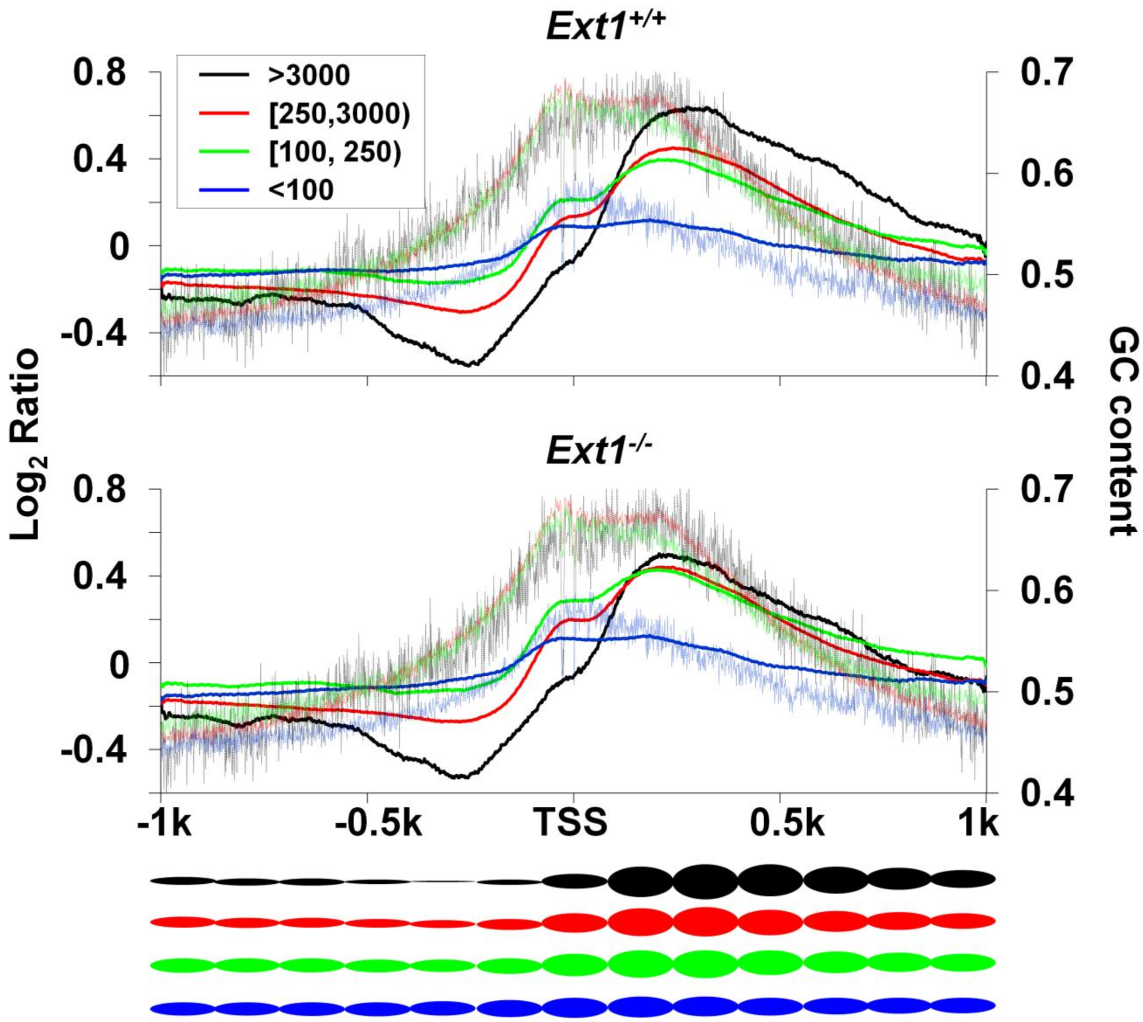
plot presents the nucleosome occupancy for one gene that represents their corresponding gene groups on the **A** panel. In each plot, the Y-axis represents the average nucleosome occupancy at each base pair position (the X-axis) as described in Figure 2. Arrows indicate transcription direction. The ovals depict hypothetical nucleosome positions across the site (See Figure s2 for biological replicates).

Fig 3.4. Nucleosome occupancy at promoter regions of genes down/upregulated after *Ext1* gene knockout. A) Top panel presents the nucleosome occupancy for genes with expression intensity decreased by >2-folds in *Ext1*$^{-/-}$ (green) compared with *Ext1+/+* (red). Bottom panel presents the nucleosome occupancy for genes with expression intensity increased by >2-folds in *Ext1*$^{-/-}$. **B)** The mRNA half-life distribution of the same downregulated and upregulated genes as shown in **A**. The genes, along with their mRNA half-life values and the statistic test, are provided in Table S5. **C)** Anxa3 and Myo1g have expression decreased more than 2 folds in *Ext1*$^{-/-}$ ESCs, Hsh2d and Pla1a have their expression increased more than 2 folds in *Ext1*$^{-/-}$ ESCs. NP has changed for genes with short mRNA half-life (Anxa3 and Hsh2d) while NP did not change for genes with long mRNA half-life (Myo1g and Pla1a). The mRNA half-life of each gene is marked with black color. Expression intensities measured in each cell type were labeled with their corresponding color. The ovals depict hypothetical nucleosome positions in each cell type (Green: *Ext1*$^{-/-}$, Red: *Ext1*$^{+/+}$) across the site (See Figure s3 for biological replicates).



A
100
50
0
-50
-150
-100
-50
0
50
100
150
PC2 (28%)
PC1 (42.9%)
Ext1+/+
Ext1-/-
B
p < 0.05
TGF-β
cell motion
phospholipid binding
growth
downregulation of transcription
Response to monosaccharide stimulus
cell adhesion
ECM/glycoprotein
EGF-related
protease
Ext1+/+
Ext1-/-

Fig 3.1



$Ext1^{+/+}$
>3000
[250,3000)
[100, 250)
<100
0.8
0.4
0
-0.4
0.7
0.6
0.5
0.4
Log_2 Ratio
GC content
$Ext1^{-/-}$
-1k
-0.5k
TSS
0.5k
1k

Fig 3.2

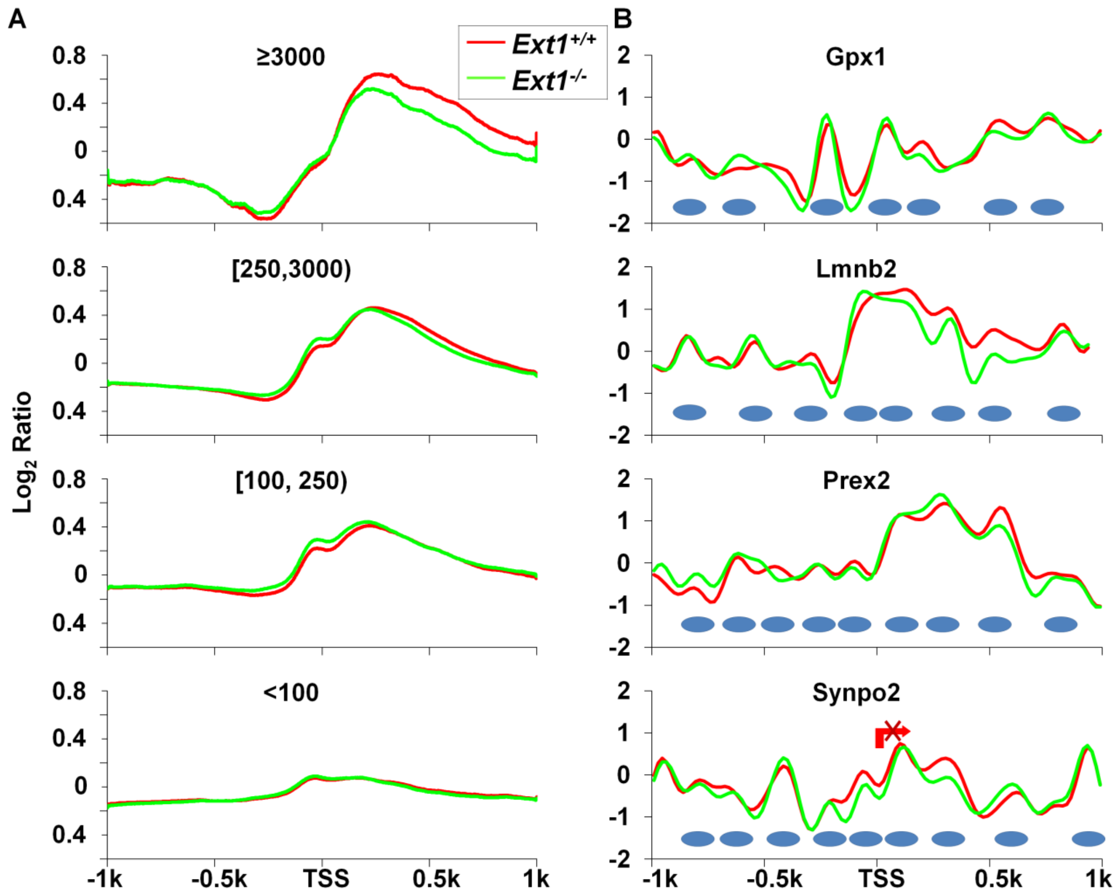

A
B
Ext1+/+
Ext1-/-
≥3000
[250,3000)
[100, 250)
<100
Log2 Ratio
0.8
0.4
0
-1k
-0.5k
TSS
0.5k
1k
Gpx1
Lmnb2
Prex2
Synpo2
2
1
-1
-2

Fig 3.3

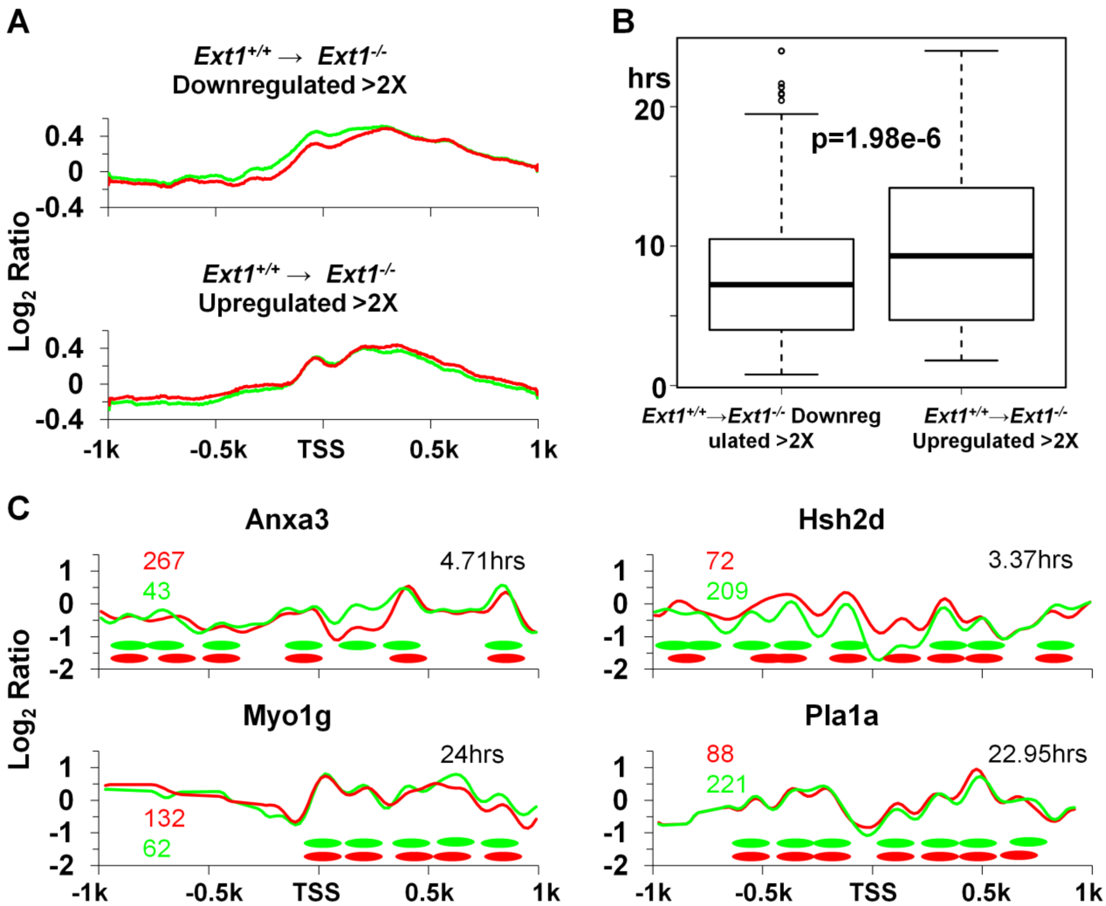

A
Ext1+/+ → Ext1-/-
Downregulated >2X
Ext1+/+ → Ext1-/-
Upregulated >2X
Log2 Ratio
0.4
0
-0.4
-1k
-0.5k
TSS
0.5k
1k
B
hrs
20
10
0
p=1.98e-6
Ext1+/+→Ext1-/- Downregulated >2X
Ext1+/+→Ext1-/- Upregulated >2X
C
Anxa3
267
43
4.71hrs
Hsh2d
72
209
3.37hrs
Myo1g
24hrs
132
62
Pla1a
88
221
22.95hrs
1
0
-1
-2

Fig 3.4

REFERENCE

Baldwin RJ, ten Dam GB, van Kuppevelt TH, Lacaud G, Gallagher JT, Kouskoff V, Merry CL. 2008. A developmentally regulated heparan sulfate epitope defines a subpopulation with increased blood potential during mesodermal differentiation. *Stem Cells* **26**(12): 3108-3118.

Bregman A, Avraham-Kelbert M, Barkai O, Duek L, Guterman A, Choder M. 2011. Promoter elements regulate cytoplasmic mRNA decay. *Cell* **147**(7): 1473-1483.

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. 2009. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell stem cell* **4**(1): 80-93.

Dennis JH, Fan HY, Reynolds SM, Yuan G, Meldrim JC, Richter DJ, Peterson DG, Rando OJ, Noble WS, Kingston RE. 2007. Independent and complementary methods for large-scale structural analysis of mammalian chromatin. *Genome research* **17**(6): 928-939.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**(7347): 398-402.

Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, Widom J, Gilad Y, Pritchard JK. 2012. Controls of nucleosome positioning in the human genome. *PLoS genetics* **8**(11): e1003036.

Gopalakrishnan S, Van Emburgh BO, Robertson KD. 2008. DNA methylation in development and human disease. *Mutation research* **647**(1-2): 30-38.

Holmborn K, Ledin J, Smeds E, Eriksson I, Kusche-Gullberg M, Kjellen L. 2004. Heparan sulfate synthesized by mouse embryonic stem cells deficient in NDST1 and NDST2 is 6-O-sulfated but contains no N-sulfate groups. *The Journal of biological chemistry* **279**(41): 42355-42358.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**(1): 44-57.

Johnson CE, Crawford BE, Stavridis M, Ten Dam G, Wat AL, Rushton G, Ward CM, Wilson V, van Kuppevelt TH, Esko JD et al. 2007. Essential alterations of heparan sulfate during the differentiation of embryonic stem cells to Sox1-enhanced green fluorescent protein-expressing neural progenitor cells. *Stem Cells* **25**(8): 1913-1923.

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic acids research* **32**(Database issue): D493-496.

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G et al. 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell stem cell* **8**(2): 200-213.

Kraushaar DC, Yamaguchi Y, Wang L. 2010. Heparan sulfate is required for embryonic stem cells to exit from self-renewal. *The Journal of biological chemistry* **285**(8): 5907-5916.

Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J et al. 2010. Dynamic changes in the human methylome during differentiation. *Genome research* **20**(3): 320-331.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics* **39**(10): 1235-1244.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**(7271): 315-322.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153): 553-560.

Ozsolak F, Song JS, Liu XS, Fisher DE. 2007. High-throughput mapping of the chromatin structure of human promoters. *Nature biotechnology* **25**(2): 244-248.

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA. 2010. c-Myc regulates transcriptional pause release. *Cell* **141**(3): 432-445.

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**(5): 887-898.

Sharova LV, Sharov AA, Nedorezov T, Piao Y, Shaik N, Ko MS. 2009. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA research : an international journal for rapid publication of reports on genes and genomes* **16**(1): 45-58.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* **3**: Article3.

Spetman B, Lueking S, Roberts B, Dennis JH. 2011. Microarray mapping of nucleosome position. *Epigenetics: A Reference Manual J Craig and N Wong, eds, Horizon Scientific Press, Norwich, UK*.

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

Tang J, Le S, Sun L, Yan X, Zhang M, Macleod J, Leroy B, Northrup N, Ellis A, Yeatman TJ et al. 2010. Copy number abnormalities in sporadic canine colorectal cancers. *Genome research* **20**(3): 341-350.

Teif VB, Vainshtein Y, Caudron-Herger M, Mallm JP, Marth C, Hofer T, Rippe K. 2012. Genome-wide nucleosome positioning during embryonic stem cell development. *Nature structural & molecular biology* **19**(11): 1185-1192.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* **282**(5391): 1145-1147.

Trcek T, Larson DR, Moldon A, Query CC, Singer RH. 2011. Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell* **147**(7): 1484-1497.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. *Nature* **474**(7352): 516-520.

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science* **309**(5734): 626-630.

CHAPTER 4

CONCLUSIONS AND DISSCUSSIONS

Both our studies of human and mouse ESCs reveal a fundamental level of epigenetic control that was not previously recognized and adds to the complexity of epigenetic mechanisms that operate in pluripotent cells such as bivalent histone modification (Bernstein et al. 2006; Mikkelsen et al. 2007; Cui et al. 2009) and hydroxymethylation (Ficz et al. 2011; Koh et al. 2011; Wu et al. 2011). We made several important discoveries by investigating the global NP changes in a well-defined hESC differentiation system: WA09 hESCs → *ISL1*+ nascent mesoderm (INM) → smooth muscle cells (SMCs) and promoter NP dynamics in $Ext1^{+/+}$ and $Ext1^{-/-}$ mouse ESCs, by paired-end MNase-seq and high density CGH arrays respectively. We found in both human and mouse ESCs: First, NP is correlated primarily with transcriptional activity and secondarily with the GC content of the sequence in promoters. Second, our results suggest the correlation between NP and mRNA half-life indicating that the expression level of silencing genes (but not activating genes) may be regulated primarily via promoter chromatin modeling, including NP.

Specifically, we found the nucleosomal DNA of hESC WA09 to be ~10bp longer than their differentiated SMCs. Several lines of evidence indicate that this variation is unlikely to be due to experimental artifacts such as differences in MNase-digestion. First, mononucleosomal fragments of various size in each cell type all carry the signature of

MNase-digestion (Dingwall et al. 1981), with the first and last bases being either A or T at ≥90% of the time. Moreover, clear NP patterns are observed for all three cell types, the non-randomness of which argues against the experimental artifact theory. Second, like the canonical nucleosome core sequence of 147bp, mononucleosomal fragments of all size remain symmetric in their A/T and G/C dinucleotide oscillating frequency. Furthermore, both the GC content and %CpG increase along with the fragment length between 147bp and 200bp. All these features are missing in randomly sheared genomic DNA fragments. Lastly, our promoter and gene body NP results are consistent with published studies, supporting the accuracy of our analysis. A difference of "10bp", a DNA helix turn, observed among the cell types is also against the experimental artifact scenario. A recent study (Gaffney et al. 2012) has reported a 10bp periodicity in DNase I sensitivity extending beyond the putative nucleosome core region into the adjacent linker sequences in human lymphoblastoid cells. Notably, these human somatic cells also have approximately the same mononucleosomal lengths (Gaffney et al. 2012) as our SMCs.

Several mechanisms could conceivably explain the observed variation in mononucleosomal DNA length among the three cell types. First, the histone core may be different, as nucleosomes with the core DNA length ranged from 100bp to170bp, depending upon the histone core composition, have been reported (Zlatanova et al. 2009; Bonisch and Hake 2012; Hasson et al. 2013). It is possible that WA09-hESCs and their 4-day differentiated WA09-INM have a larger histone core, resulting in an extra helix turn of DNA in wrapping, when compared to the 21-day differentiated WA09-SMCs. This model is consistent with the A/T and G/C dinucleotide symmetry shown in Figure 2.7B.

Alternatively, the histone cores of WA09-hESCs and WA09-INM harbor histone variants that stabilize the nucleosomes and protect more DNA from MNase-digestion. Examples of histone variants influencing the nucleosome stability have been reported (Bonisch and Hake 2012; Li et al. 2012; Hasson et al. 2013; Hu et al. 2013). Consistent with these possibilities, significantly more histone genes are downregulated in WA09-SMC, when compared to WA09-hESC and WA09-INM. This may lead to a change in nucleosome histone core composition during differentiation. Second, it is also possible that the three cell types are all similar in nucleosome core composition. However, in both WA09-hESC and WA09-INM, about 10bp linker DNA is protected by proteins such as hMeCP2 (Nikitina et al. 2007) from MNase-digestion. This theory, however, cannot easily explain the observed dinucleotide symmetry shown in Figure 2.7B of chapter 2, unlike the first model. Third, the difference could also come from the DNA itself, e.g., difference in its methylation or hydroxymethylation status. This also seems less likely because we did not observe significant differences between the X-chromosome, whose DNA is more heavily methylated (Nazor et al. 2012), and autosomes, in any of our analyses. Clearly, further studies, e.g., investigating the core histones and linker protein, are needed to distinguish these scenarios.

Importantly, our findings shed light on why hESCs can maintain their remarkable genomic stability, unlike somatic cells. The dominating background mutations in the human genome are C→T changes arising from deamination of cytosine or 5-methylcytosine (Venter et al. 2001). A recent study has reported that in eukaryotes, nucleosome occupancy nearly eliminated cytosine deamination, resulting in a ~50%

decrease in the C→T mutation rate in nucleosomal DNA (Chen et al. 2012). As described above, the nucleosomal DNA of WA09-hESC is 10bp longer than their terminally differentiated SMCs. In addition, both the GC content and %CpG increase along with the mononucleosomal fragment length once exceeding the canonical nucleosomes core length of 147bp. Thus, compard to somatic cells, longer nucleosomal DNA in hESC could be instrumental in maininting the low sequence mutation rate by reducing C→T changes, contributing its remarkable genome integrity.

In summary, our study reveals a fundamental level of epigenetic control that was not previously recognized and adds to the complexity of epigenetic mechanisms that operate in pluripotent cells such as bivalent histone modifications and hydroxymethylation.

Future studies

Both NP studies introduced in this dissertation raised questions that need to be answered in the future. These questions suggest a variety of research directions that need to be pursued to construct overall epigenetic dynamics in embryonic stem cells.

One such direction would be to distinguish two hypotheses raised in Figure 2.7: the ~10bp longer mononucleosomal DNA in hESCs was cause by larger histone core (Zlatanova et al. 2009; Bonisch and Hake 2012; Hasson et al. 2013) or proteins (Nikitina et al. 2007) at the nucleosome entry/exit position protect more DNA from MNase digestion. For the first hypothesis: four histone H2A variants, namely, H2A.X, H2A.Z, macroH2A, and H2A.Bbd and two histone H3 variants, namely, H3.3 and CENP-A, have been identified in mammalian cells (Okuwaki et al. 2005; Sarma and Reinberg 2005). First, classical biochemistry experiments can be done to verify whether the mononucleosomal fragment size differences were caused by histone variant: we can reconstitute nucleosomes *in vitro* from purified core histone protein variants and varies length of DNA fragments by salt gradient dialysis (Peterson 2008). Nucleosomes will be purified and the length of DNA that wrap around nucleosomes can be accessed by gel electrophoresis after digesting the core protein with proteinase. This experiment helps us to verify whether the composition of histone core result in nucleosomal DNA length differences. In addition, the result will also suggest which histone variant leads to the ~165bp and 155bp mononucleosomal DNA in WA09-hESC and WA09-SMC, respectively. Second, genome-wide histone variant distributions can be determined by chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) (Jin et al. 2009; Goldberg et al. 2010). Histone variants are not uniformly distributed in the genome, they also mark chromatin status and are involved in the regulation of gene expression (Torres-Padilla et al. 2006; Pasque et al. 2012). Integration of the nucleosome reconstitution experiment and the genome wide study of histone variant distributions will help us answer the question of whether histone variant cause the differences of DNA fragments in three cell types and how the histone variant dynamics contribute to the hESC differentiation. For the second hypothesis: proteins such as hMeCP2 (Nikitina et al. 2007) et al. 2007) protects ~11bp DNA from MNase-digestion. Whether certain protein exists in the MNase digested fragments can be first determined by measuring the absorbance at 280 nm then verified with biochemistry experiments such as western plot.

The other direction would be the integration of NP dynamics during hESC differentiation with other epigenetic components, including histone modification, DNA methylation etc. We have explored the relationship between NP and histone modification/DNA methylation status in hESCs. However, we did not have the matched either histone modification or DNA methylation data for two other differentiation cell types: WA09-INM and WA09-SMC. In this case, we cannot discuss how NP dynamics corresponding to histone modification and DNA methylation state changes during the differentiation. Once the histone modification and DNA methylation date for WA09-INM and WA09-SMC have been generated, we can investigate how these epigenetic components function cooperatively during the hESC differentiation process. Similarly, we can study the relationship between NP and these two epigenetic markers after HS knockout by generating the histone modification and DNA methylation data for mouse ESCs and mouse *Ext1*$^{-/-}$ ESCs.

The third line of research is to investigate the relationship between NP and mRNA half life values. We observed the similar correlation between NP and mRNA half life values in both hESCs and mESCs that genes which are upregulated/downregulated have distinct distributions of mRNA half life. However, the human study was conducted by mapping mESC mRNA half life to human with homologs, therefore, the correction for hESCs mRNA half lives are needed. Furthermore, quantitative measurements of both NP in promoter and mRNA half life can be measured by perturbation of the promoter sequence for individual gene (Bregman et al. 2011; Trcek et al. 2011). The combination of both accurate hESC mRNA half life values and individual gene investigations may provide us more comprehensive evidences for the mechanism working behind the regulation of transcription.

The fourth line of research is to explore the regulatory network in either hESC differentiation or mESC knockout system. Previous researches have identified the core regulatory network helps maintaining the pluripotency of ES cells (Boyer et al. 2005; Loh et al. 2006), yet other genes such as Myc and Lin28 were also play essential roles in cell fate decisions (Graf and Enver 2009). In our mouse HS deficient ESCs, we found that the expression of *Myc* was decreased ~10folds, which may suggest a *Myc*-centered regulatory network in mESCs. As suggested in other studies, we may use ChIP-seq to identify genome wide *Myc* targets (Chen et al. 2008), and use *in vivo* metabolic biotin tagging method (Kim et al. 2010) to identify *Myc* partners.

REFERENCES

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125(2): 315-326.

Bonisch C, Hake SB. 2012. Histone H2A variants in nucleosomes and chromatin: more or less stable? Nucleic acids research 40(21): 10719-10741.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122(6): 947-956.

Bregman A, Avraham-Kelbert M, Barkai O, Duek L, Guterman A, Choder M. 2011. Promoter elements regulate cytoplasmic mRNA decay. Cell 147(7): 1473-1483.

Chen X, Chen Z, Chen H, Su Z, Yang J, Lin F, Shi S, He X. 2012. Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. Science 335(6073): 1235-1238.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133(6): 1106-1117.

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. 2009. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell stem cell 4(1): 80-93.

Dingwall C, Lomonossoff GP, Laskey RA. 1981. High sequence specificity of micrococcal nuclease. Nucleic acids research 9(12): 2659-2673.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 473(7347): 398-402.

Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, Widom J, Gilad Y, Pritchard JK. 2012. Controls of nucleosome positioning in the human genome. PLoS genetics 8(11): e1003036.

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X et al. 2010. Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140(5): 678-691.

Graf T, Enver T. 2009. Forcing cells to change lineages. Nature 462(7273): 587-594.

Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N, Alonso A, Warburton PE, Black BE. 2013. The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat Struct Mol Biol.

Hu G, Cui K, Northrup D, Liu C, Wang C, Tang Q, Ge K, Levens D, Crane-Robinson C, Zhao K. 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.

Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G. 2009. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nature genetics 41(8): 941-945.

Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, Cantor AB, Orkin SH. 2010. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell 143(2): 313-324.

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G et al. 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell stem cell 8(2): 200-213.

Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J, Li W, Kaestner KH. 2012. Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. Cell 151(7): 1608-1616.

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature genetics 38(4): 431-440.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448(7153): 553-560.

Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Muller FJ, Wang YC, Boscolo FS et al. 2012. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell stem cell 10(5): 620-634.

Nikitina T, Ghosh RP, Horowitz-Scherer RA, Hansen JC, Grigoryev SA, Woodcock CL. 2007. MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. The Journal of biological chemistry 282(38): 28237-28245.

Okuwaki M, Kato K, Shimahara H, Tate S, Nagata K. 2005. Assembly and disassembly of nucleosome core particles containing histone variants by human nucleosome assembly protein I. Molecular and cellular biology 25(23): 10639-10651.

Pasque V, Radzisheuskaya A, Gillich A, Halley-Stott RP, Panamarova M, Zernicka-Goetz M, Surani MA, Silva JC. 2012. Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. Journal of cell science 125(Pt 24): 6094-6104.

Peterson CL. 2008. Salt gradient dialysis reconstitution of nucleosomes. CSH protocols 2008: pdb prot5113.

Sarma K, Reinberg D. 2005. Histone variants meet their match. Nature reviews Molecular cell biology 6(2): 139-149.

Torres-Padilla ME, Bannister AJ, Hurd PJ, Kouzarides T, Zernicka-Goetz M. 2006. Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. The International journal of developmental biology 50(5): 455-461.

Trcek T, Larson DR, Moldon A, Query CC, Singer RH. 2011. Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. Cell 147(7): 1484-1497.

Venter JC Adams MD Myers EW Li PW Mural RJ Sutton GG Smith HO Yandell M Evans CA Holt RA et al. 2001. The sequence of the human genome. Science 291(5507): 1304-1351.

Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes & development 25(7): 679-684.

Zlatanova J, Bishop TC, Victor JM, Jackson V, van Holde K. 2009. The nucleosome family: dynamic and growing. Structure 17(2): 160-171.