Genomic Targeting and Function of Polycomb Repressive Complex 2 and ISWI Chromatin Remodelers

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Darko Barisic

aus Zagreb, Kroatien

Basel, 2017

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Dirk Schübeler und Prof. Dr. Anton Wutz.

Basel, den 18.04.2017

Prof. Dr. Martin Spiess

To Fabian Jenny.

Acknowledgements

First and foremost, I would like to thank my PhD supervisor Dirk Schübeler for being a supportive mentor in countless ways. His tremendous patience and encouragement enabled me to become a better scientist, to think more critically and to independently develop research ideas.

In addition, I thank Anton Wutz and Joerg Betschinger for their helpful feedback and support as my thesis committee advisors.

Furthermore, I want to especially thank Michael Stadler for his highly valuable support and help in computational analysis. Arnaud Krebs was instrumental in the initial stages of the project development and data analysis. Altuna Akalin, Lukas Burger and Anais Bardet provided me with essential bioinformatics support. I thank Sophie Dessus-Babus, Matyas Flemr, Hubertus Kohler and Philip Jermann for their indispensable input and help with technical and methodological aspects of the projects. Furthermore, I thank Christiane Wirbelauer and Leslie Hoerner for technical support within the lab. Moreover, I want to thank the entire Schübeler group for creating such an incredible atmosphere and a fun place to work for the past 4.5 years.

I want to thank Boehringer Ingelheim Fonds for funding my PhD project and supporting my career development. They provided me with numerous opportunities to attend conferences, learn through inspiring courses and seminars, but also to network and meet remarkable people within the program.

Special thank you to Juliane Schmidt, for her incredible friendship and constant support throughout my PhD. Furthermore, I want to thank Rafael Santos for his help in the latter stages of my PhD. Finally, I thank Fabian Jenny, for being my lifetime inspiration and a wonderful influence in everything I do.

Summary

In eukaryotes, chromatin provides a way to compact the genetic material into the confined space of a nucleus. It is also a means to store the same genetic information in different chromatin states. Alteration of these states is enabled by chromatin modifying and remodeling machineries – enzymes that utilize a diverse range of structural changes to chromatin. Despite their apparent importance in gene regulation, it is unclear how they facilitate the transition between chromatin states. Within two distinct projects, we aimed to (1) decipher how chromatin modifying complexes, namely the Polycomb group proteins, are targeted to chromatin and (2) how chromatin remodelers, specifically the ISWI remodeling complexes, change chromatin structure.

Polycomb group proteins assemble as chromatin-modifying complexes that maintain the memory of the silent transcriptional state, in part through methylation of lysine 27 on histone H3. Despite their established importance during development, it is largely unclear how these complexes are recruited to specific target genes and how they impair transcription. In flies, Polycomb is recruited by Polycomb response elements that are abundant in various DNA-binding factor motifs. However, the contribution of individual motifs is not yet resolved. In mammals, equivalents of Polycomb response elements are not yet characterized. Here, we aimed to dissect Polycomb-mediated silencing in the mouse genome by identifying DNA determinants of Polycomb recruitment and investigating the role of Polycomb recruitment in transcriptional silencing. More specifically, we developed an assay to test many DNA sequences with various sequence properties for their ability to drive PRC2 recruitment in mouse embryonic stem cells. The assay enabled integration of hundreds of sequences into a defined genomic location in parallel. We found that high density of unmethylated CG motifs within a synthetic backbone sequence is sufficient to recruit PRC2. Furthermore, to link PRC2 recruitment with transcriptional repression, we used

CRISPR/Cas9 technology to delete the core PRC2 (Eed) component and monitored the transcriptional response by RNA-seq. Upon depletion of global H3K27me3 levels, we observed no significant changes in gene expression in mouse embryonic stem cells but global deregulation of PRC2 targets during differentiation into neuronal progenitors. These results indicate that recruitment of PRC2 and subsequent H3K27 methylation is important for cell-fate transition, but not required for gene repression in mouse embryonic stem cells.

For the second project, we were interested in chromatin remodelers (ISWI) and their role in regulating chromatin structure. Chromatin remodelers are known to use the energy of ATP hydrolysis to evict, slide and reposition nucleosomes, yet we do not fully understand how nucleosome positioning and occupancy affects transcription factor binding. To this aim, we deleted Snf2h, the ATPase subunit of the ISWI chromatin remodeling family, in mouse embryonic stem cells. The Snf2h knockout mouse embryonic stem cells are viable with unchanged expression of pluripotency markers, which is exciting as this is the first viable knockout of an ATPase remodeler subunit. To determine global changes upon deletion of Snf2h, we monitored nucleosome positioning, chromatin accessibility and transcriptional response in Snf2h knockout cells using MNase, ATAC and RNA sequencing, respectively. Extensive data analysis revealed global changes in nucleosome positioning proximal to transcription start sites and transcription factor motifs. Analyzing nucleosome positioning and chromatin accessibility data, we identified transcription factors that require Snf2h to bind their target sites, such as CTCF. It seems that in the absence of Snf2h, nucleosomes cannot be evicted from CTCF motifs, which in turn results in loss of CTCF binding. Taken together, these results indicate that ISWI complexes enable transcription factor binding, at both promoters and distal regulatory regions, by sliding of motif-bound nucleosomes.

Table of Contents

Chapter 1
Introduction

1.1 Brief History of Epigenetics

Shortly after the initial analysis of the human genome sequence (Lander et al., 2001), it became clear that the DNA sequence itself is not sufficient to understand the complexity of regulating a genome. This is, in fact, the central enigma in the field of epigenetics: how do we come from one genome in a single-cell zygote to a whole person made out of hundreds of different cell types and trillions of cells? The idea that different cell types do not possess different genes, rather different ways to regulated them came about already in the 19th century. Scientists were in the search for elements that determine the developmental plan of an organism. The term *epigenetics* was coined in 1942 by Conrad Waddington, a British developmental biologist. At that time, *epiphenotype* was considered to entail all the developmental processes that occur between the genotype and the phenotype (Waddington, 1942). Although Waddington considered epigenetic mechanisms only in the context of embryogenesis and development, he captured the essence of epigenetic regulation by describing the fertilized egg as a form *"… in which all the complexity of the fully developed animal is implicit but not yet present*". Up to this point, it was not clear what was the epigenetic element that carries out developmental decisions. It was not even obvious that somatic cells inherit the complete genetic information from the fertilized egg. Yet, Stedman and Stedman, by comparing chemistry of nuclei from erythrocytes and liver cells, proposed that histone proteins act to suppress activity of particular genes, in a cell-specific manner (Stedman, 1950). "… *each nucleus possesses a basic protein characteristic of the type of cell of which it forms part*" they hypothesized (Stedman, 1950). The notion that different cell types have

different types of histones, or histones with different residues, only started to be investigated year later, with the discovery of histone lysine methylation and acetylation (Allfrey, Faulkner, & Mirsky, 1964; Murray, 1964). Yet, one of the first demonstrations that nucleosomes influence gene transcription *in vivo* was in 1998, by Michael Grunstein and Min Han (M Han, Kim, Kayne, & Grunstein, 1988; Min Han & Grunstein, 1988). Their research associated the depletion of nucleosomes at the *PHO5* gene in *S. cerevisiae* with gene activation. Shortly after, *PHO5* system became a paradigm in the field and paved the way for modern epigenetics. In fact, numerous findings on chromatin-modifying and nucleosome-remodeling factors were derived from experiments in the PHO5 system; they will be discussed in greater details in the chapter 1.3.2. Following discoveries by Grunstein and Han (Min Han & Grunstein, 1988), epigenetics as a field experienced an exponential growth. Yet, 70 years after Edgar & Ellen Stedman started investigating epigenetic mechanisms of gene regulating, we still find ourselves asking the same question they did: "*It has always been a puzzle to us […] how the physiological functions of cell nuclei in the same organism can differ […] from one cell-type to another when they all contain identical chromosomes and hence identical genes* (Stedman, 1950)."

1.2 Evolution of Gene Regulation

In the course of evolution, genome size increased with the emergence of more complex forms of life. In the animal kingdom alone, there is a 3000-fold variability in the genome size between certain species (Gregory et al., 2007). Interestingly, genome size of an organism does not scale with the number of genes in the respective genome. The increase in number of genes is rather small in proportion (Gregory et al., 2007). This observation led to the hypothesis that phenotypic diversity arises from increasing number of regulatory regions. This would enable regulation of the limited number of genes in different ways. Indeed, comparative analysis shows that vertebrates, although phenotypically different, share the same number of genes and tissue types (Brawand et al., 2011; Romero, Ruvinsky, & Gilad, 2012). It seems the phenotypic diversity among vertebrates rather comes from adaptive changes in gene regulation, and not from mutations in protein-coding sequences (Romero et al., 2012). However, it remains an open question what changes in the wiring of gene regulation explain the differences in genes expression levels, and consequently influence the phenotypic differences. This is of particular interest because evolution of regulatory circuits in vertebrates is directly linked to mechanisms of cellular differentiation during development. Transcription factors and chromatin modifying and remodeling complexes are thought to be the two main contributors to this regulatory divergence. Findings and work presented in this thesis aim to investigate and help better understand those core mechanisms of gene regulation in mammals.

1.3 Gene Regulation in the Context of Chromatin

1.3.1 Chromatin Organization in Eukaryotes

The genomes of eukaryotes exist mainly in the form of chromatin. One of many roles of chromatin is to compress 3.2 gigabases of DNA, found in almost every cell of a human body, into a highly restricted space within a nucleus (Gregory et al., 2007). In other words, about 2 meters of DNA in length, if DNA molecules are linearly extended, has to be compressed to 6 μm^3 of the nucleus space (Oudet, Gross-Bellard, & Chambon, 1975). Chromatin is thought to play an important role in that compaction. The fundamental subunit of chromatin is the nucleosome which consists of 147 bp of DNA wrapped around an octamer of two of H2A, H2B, H3 and H4 histone proteins (Luger, Mäder, Richmond, Sargent, & Richmond, 1997; Figure 1.1). Histone proteins are composed of a structured globular domain, an unstructured highly basic N-terminal tail with many lysine and arginine residues, as well as a short basic C-terminal tail (Luger et al., 1997). The interaction between positive charges of histone proteins and the negative charges of the DNA brings stability to the histone-DNA complex (Grunstein, 1997). This compaction of DNA into nucleosomes adds a level of basal repression genome-wide. It also creates an accessibility barrier for DNA-binding factors, such as transcription factors and the replication machinery (Min Han & Grunstein, 1988; Knezetic & Luse, 1986). Having a binding barrier provides an opportunity for differentially regulating accessibility to DNA and creating distinct functional outcomes. It provides a way to regulate cell-type specific programs from identical genomic sequence. Indeed, changes to chromatin structure contribute to dynamic changes in gene expression in the course of development (Allis, 2007). Moreover, chromatin contributes to maintenance of cell fate by providing stable, heritable states of gene expression (Allis, 2007).

Altering chromatin to counter its repressive nature occurs on multiple layers; at the level of (1) histones and DNA in the form of histone variants and

chromatin modification, (2) nucleosome positioning and (3) higher-order chromatin structure (Figure **1.1**). Several variants of the canonical core histones with specific function in different biological contexts have been described. One prominent example is the replacement of canonical histones H2A and H3 with H2A.Z and H3.3 variants, respectively. These replacements happen in the context of actively transcribed regions, however, the precise mechanism by which replacement histones affect transcriptional outcomes is still under debate (Talbert & Henikoff, 2016; see chapter below). It is thought that histone variants tether specific effector proteins to chromatin. Furthermore, they might impact stability of nucleosomes which would affect binding to DNA. Chromatin modifiers also provide a mean to regulate accessibility of DNA. Chromatin modifiers are considered to be *writers* and *erasers* of post-translational modifications, with majority of modifications found at the N-terminal tails of histone proteins. This includes addition or removal of acethyl, methyl, ubiquityl or phosphate functional groups, among others (Allis, 2007). Certain chromatin modifications are associated with repressive chromatin structure, such as tri-methylation of histone H3 on lysine residues 27 (H3K27me3) and 9 (H3K9me3; Zhou, Goren, & Bernstein, 2011). Other modifications, H3K4me3 and H3K27ac for example, are generally found on chromatin permissive for transcription (V. W. Zhou et al., 2011). They affect chromatin structure by (1) reducing the positive charge of histones and therefore loosening the interaction with the DNA or by (2) recruiting chromatin remodelers and other effector proteins that bind the respective modified residue (eg. bromodomain and chromodomain bearing proteins; Kouzarides, 2007; Zhou, Goren, & Bernstein, 2011). Despite their recognized importance in development, it is still unclear whether particular histone modifications are a cause or a consequence of a certain transcriptional state.

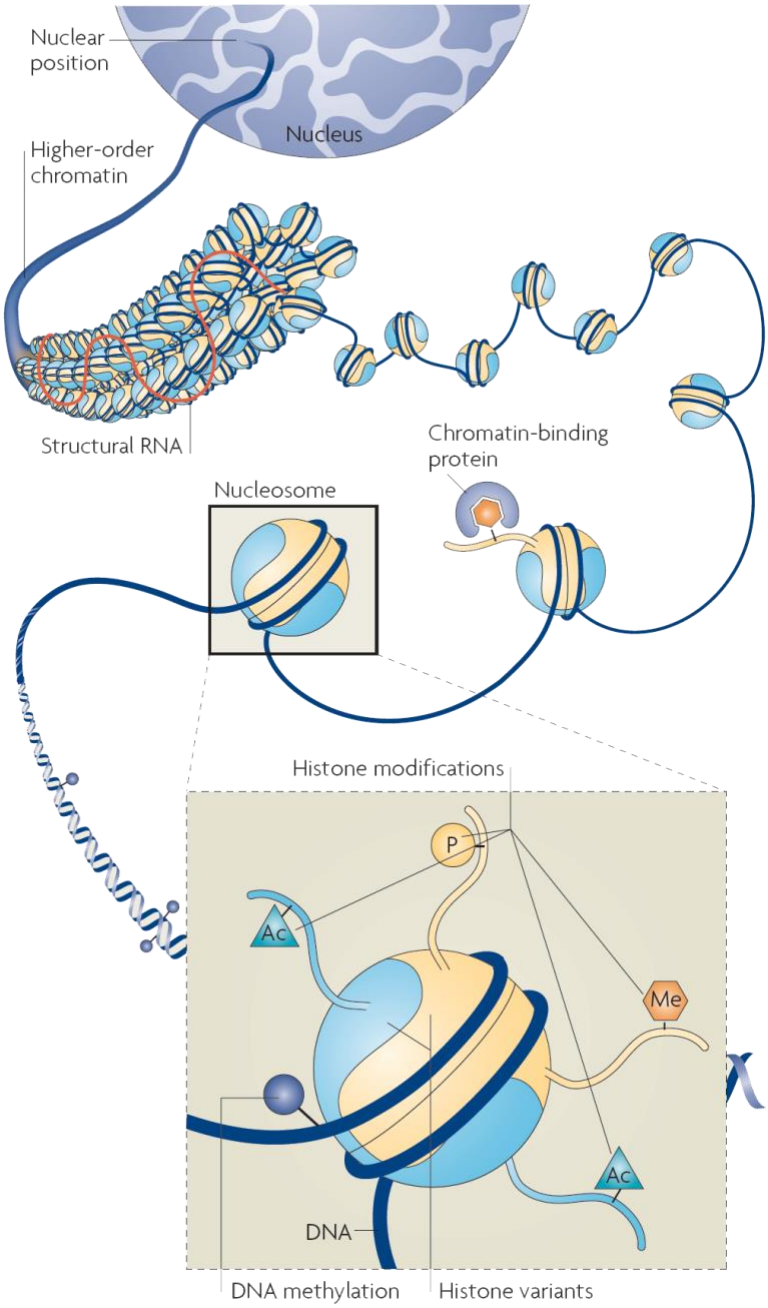


Figure 1.1 Organization of chromatin in eukaryotic genomes. The DNA alpha-helix is wrapped around a histone octamer that is composed of one $(H3\text{-}H4)_2$ tetramer and two H2A-H2B dimers. DNA can be modified at cytosines within CpG dinucleotides. Histones are also subjected to post-translational modifications; methylation (Me), acetylation (Ac), phosphorylation (P), etc. Histone variants add further complexity; variants H3.1 and H3.2 are incorporated in a DNA replication-dependent manner. H3.3, one of the replacement variants, is incorporated in a DNA replication-independent manner. Nucleosomes further fold into higher-order structure resulting in higher compaction, with metaphase chromosomes exhibiting the highest form of compaction (Adapted from Probst et al. 2009).

DNA methylation is another modification of chromatin that is associated with a repressive function (A. Bird, 2002; A. P. Bird & Wolffe, 1999). DNA methylation functions as a repressive mark either by recruiting methyl-CpG binding domain (MBD) proteins or by directly inhibiting binding of methylation-sensitive transcription factors (A. P. Bird & Wolffe, 1999; Schübeler, 2015). Taken together, chromatin modifications and histone variants seem to act as a platform for tethering other complexes. This is the case even when their role in chromatin accessibility is direct, such as in the case of histone acetylation that has an immediate effect on nucleosome stability, or DNA methylation that can block transcription factor binding directly. Both modifications, like many other modifications to histone tails, recruit complexes that further change chromatin structure.

Indeed, chromatin modifications and histone variants are known to recruit chromatin remodelers, large multi-protein assemblies. Remodelers use the energy of ATP hydrolysis to slide and evict nucleosomes or change nucleosome composition (Ho & Crabtree, 2010; Narlikar, Sundaramoorthy, & Owen-Hughes, 2013; Figure 1.2). It is thought that chromatin remodeling exposes the genomic sequence masked by the nucleosome and thereby enables binding of transcription factors and the transcriptional machinery. It has been shown, especially in the context of transcriptional initiation, that chromatin remodelers are required for eviction and repositioning of nucleosomes in order to activate genes (Carey, Li, & Workman, 2006; Parnell, Huff, & Cairns, 2008). Chromatin remodelers are not the only contributors to nucleosome positioning genome-wide. It seems some transcription factors, also known as pioneering transcription factors (eg., FoxA, GATA), are able to bind their target motifs even when it is masked by a nucleosome (Hughes, Jin, Rando, & Struhl, 2012; Soufi et al., 2015). The binding of the pioneering transcription factor will therefore affect nucleosome positioning in proximity to its motif. It is, however, unclear whether pioneering factors act autonomously. They might require activity of chromatin remodelers, or other mean of nucleosome fluidity, for functional binding *in vivo*. Finally, intrinsic properties of DNA sequence itself can

influence nucleosome positioning. The ability of DNA to bend around the histone octamer is considered the main contributing feature. Two types of sequence determinants have been shown to affect DNA bending in yeast; (1) 10-bp periodicity of flexible dinucleotides (AT or TA) and (2) poly(dA:dT) and poly(dG:dC) tracts found in promoters or *S. cervisiae* (Brogaard, Xi, Wang, & Widom, 2012; Segal & Widom, 2009). Optimal nucleosome formation occurs when flexible dinucleotides (AT and TA) are positioned in the 10 bp helical turn that faces the histones (Bai & Morozov, 2010; Jiang & Pugh, 2009; Struhl & Segal, 2013). poly(dA:dT) and poly(dG:dC) are preferentially found outside of nucleosomes as their stiffness inhibits nucleosome formation (Struhl & Segal, 2013). Identification of sequence determinants of nucleosome positioning, however, has been restricted to a limited number of organisms, with most studies performed in yeast. Whether the same mechanism prevails in higher eukaryotes is still a matter of debate. Further details on determinants of nucleosome positioning in mammals will be discussed in following chapters.

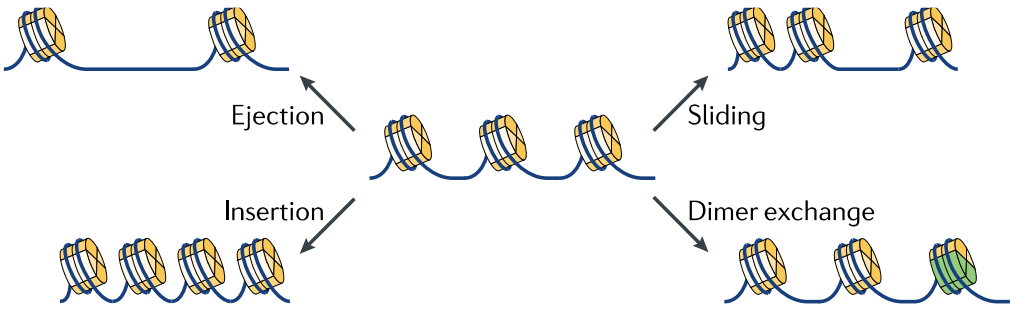


Figure 1.2 Molecular mechanisms of nucleosome remodeling. Four main mechanisms of nucleosome remodeling are depicted. Different chromatin-remodeling families exert those processes (discussed in greater detail in Chapter 1.5).

1.3.2 Gene Regulation by Nucleosome Positioning

Recent advances in DNA sequencing technology made it possible to map nucleosome positions throughout the genome. There are many methods to map nucleosomes genome-wide; from using the Assay for Transposase-Accessible Chromatin Sequencing (ATAC-seq) and Nucleosome Occupancy and Methylome Sequencing (NOMe-seq) to the *gold standard* Micrococcal Nuclease Sequencing (MNase-seq; Buenrostro, Wu, Chang, & Greenleaf, 2015; Kelly et al., 2012; Struhl & Segal, 2013). In MNase-seq protocols, an endo-exonuclease from *S. aureus* is used to digest linker DNA, unprotected sequence in between two nucleosomes. Selecting and sequencing protected DNA fragments corresponding to the length of one nucleosome, around 150-200 bp, will reveal the underlying genomic location of the nucleosome. The first maps revealed the nucleosomes are not randomly positioned throughout the genome of *S. cerevisiae* (Yuan et al., 2005). A certain genomic region has positioned nucleosomes when MNase-seq reads map to a discreet location of around 150 bp. This indicates most cells across the population had a nucleosome on that position at the time of chromatin isolation (Struhl & Segal, 2013). In contrast, genomic regions with not positioned nucleosomes could have the same number of mapped nucleosomal reads, but not restricted to one position (Struhl & Segal, 2013). Nucleosome occupancy, on the other hand, refers to MNase-seq coverage of a certain genomic location in comparison to the rest of the genome, regardless of the nucleosome position. Regulatory regions, such as enhancers and promoters, seem to have lower nucleosome occupancy, yet nucleosomes usually occupy preferred position within those regions (Teif et al., 2012). Promoters of active genes show a distinct depletion of nucleosomes referred to as the nucleosome-free region (NFR). The NFR is flanked by two highly position nucleosomes, referred to as +1 and -1 nucleosomes, in respect to their orientation to the transcriptional start site (Jiang & Pugh, 2009; Struhl & Segal, 2013; Teif et al., 2012). Interestingly, the space in between the +1 and -1 nucleosomes

is the length of one additional nucleosome. This gap in the NFR might be filled with a nucleosome to inhibit transcriptional initiation when repression is required. In fact, the phosphate-regulated yeast *PHO5* promoter contains an array of highly positioned nucleosomes flanking the TSS when the *PHO5* gene is transcriptionally inactive (Svaren & Hörz, 1997). In low phosphate conditions, the nucleosomes –1 to –4 are evicted from the promoter to expose binding sites for transcription factors necessary to activate the *PHO5* gene (Bryant et al., 2008). The nucleosome eviction is facilitated by a complex network of five chromatin remodeling complexes (Musladin, Krietenstein, Korber, & Barbaric, 2014). Once the promoter site has been exposed, the transcriptional activator Pho4 and the TATA-box binding protein (TBP) bind their motifs within the promoter and activate *PHO5* expression. Following the reversal to high phosphate concentrations, repression of *PHO5* is reestablished simply by chromatin reassembly. Histone chaperone Spt6 was shown to facilitate this chromatin reassembly (Adkins & Tyler, 2006) and blocking Spt6-mediated histone deposition cases the *PHO5* promoter to be continuously active. Therefore, *PHO5* activation is a model example illustrating how specific nucleosome position and occupancy inhibits transcription factor binding and downstream gene activation in yeast. The typical structure of the NFR is conserved in higher eukaryotes where regulating transcriptional outcome by nucleosome positioning might be a way to regulate cell-type specificity. Teif et al. identified cell-type specific nucleosome positioning in proximity to certain transcription factor motifs in mouse embryonic stem cells (Teif et al., 2012). When bound, most transcription factor binding sites are flanked by an array of positioned nucleosomes, with CCCTC-binding factor (CTCF) motifs having the most pronounced positioning in their vicinity. It is yet to be determined whether this positioning is a result of active chromatin remodeling. Alternatively, transcription factor binding itself might act as a barrier against which nucleosomes are placed, resulting in an array of positioned nucleosomes (Mavrich et al., 2008).

1.3.3 Histone Modifications, Histone Variants and DNA Methylation

Histone modifications

Histone proteins within the nucleosome structure have an unstructured N-terminal tail that protrudes from the center of the nucleosome, making it accessible for effector proteins to either modify or read the already established modification (Alberts, 2008). Chromatin is modified by the activity of modifying complexes that covalently transfer a functional group to the histone tail. These modifications have an impact on chromatin structure and transcription both in *cis* and *trans* (Allis, 2007). Two main mechanisms of epigenetic regulation by histone modifications have been described. Certain histone modifications, such as acetylation and phosphorylation, decrease the net positive charge of the histone octamer resulting in reduced electrostatic interaction between histones and DNA (Grunstein, 1997). Open chromatin structure increases accessibility for transcription factors, such as factors of the transcriptional machinery, to bind DNA. Alternatively, histone modifications recruit chromatin modifying complexes and interacting proteins which can lead to both activation or repression.

Histone acetylation was shown to exhibit both mechanisms. Histone acetyltransferases (HAT) promiscuously acetylate N-terminal tails of histones H2A, H3 and H4. Generally, higher levels of acetylation correspond to higher rates of gene activity (Dion et al., 2005). On the other hand, specific acetylation of histone H4 on lysine 16 (H4-K16Ac) modulates functional interactions of chromatin remodeling complex ACF and inhibits re-positioning of a nucleosome (Shogren-Knaak et al., 2006). Furthermore, acetyl-binding domains, so-called bromodomains, are found in many chromatin-associated proteins, eg. in subunits of the RNA polymerase II preinitiation complex and chromatin remodeling complex RSC (Allis, 2007). It is thought that bromodomains have a role in recruiting protein complexes

to chromatin. However, the function of bromodomains in gene regulation is not well characterized.

Advances in DNA sequencing technologies, namely chromatin immunoprecipitation sequencing (ChIP-seq), identified genome-wide profiles of chromatin modifications (Hoffman et al., 2013; Ram et al., 2011). Di- and tri-methylation of H3K4 are found enriched at promoters, irrespective of their transcriptional activity. Enhancers are marked by H3K4 mono-methylation (H3K4me1) and additionally, H3K27 acetylation. Transcribed regions are marked by tri-methylated of H3K36 in gene bodies. Ubiquitination of H2AK119 and tri-methylation of H3K27 are thought to be transiently silencing marks, as they are present at transcriptionally silent developmental genes that are switched on later in development. Constitutively silent genes are decorated by tri-methylated H3K9 (Hoffman et al., 2013; Ram et al., 2011).

These maps revealed chromatin domains are bimodal in their genomic distribution; modifications correlating with gene activity are found in regions permissive for transcription, and *vice versa*. The exception to that rule are so-called bivalent domains marks by both K27 tri-methylation, a silencing mark and K4 tri-methylation, the mark considered as activating (Allis, 2007; Kouzarides, 2007). The dual nature of bivalent domains is thought to pose genes for activation at subsequent stages in development and enable the rapid change in states. Chromatin maps performed in different tissues did reveal the dynamic nature of these chromatin marks, with dynamic gene expression patterns throughout development (Allis, 2007; Arvey, Agius, Noble, & Leslie, 2012; Bannister & Kouzarides, 2011; Graf & Enver, 2009; V. W. Zhou et al., 2011). To what degree particular chromatin modifications drive or maintain developmental decisions, and to what degree they are a mere consequence of a transcriptional state is a complex question and a part of an ongoing discussion. It is clear, however, from the loss-of-function studies, that the majority of chromatin modifying complexes are essential for mammalian development (Huang et al., 2009).

To date, over 50 different histone modifications have been identified including acetylation, methylation, ubiquitination, and phosphorylation (Kouzarides, 2007; V. W. Zhou et al., 2011). It is appropriate to hypothesize that many modifications will have overlapping functions and their function will be aggregated to only a few chromatin states. Indeed, Filion et al. identified principal components of chromatin state. By determining the binding profile of 53 proteins known to have a role in chromatin dynamics, they discovered the fly genome can be computationally segmented into only five principal chromatin types based on the protein composition (Filion et al., 2010).

Histone variants

Replacement of canonical histone proteins by histone variants is a dynamic process that changes chromatin properties. Well over a dozen histone variants have been described to date and their role in replication, transcriptional regulation and DNA damage has been well established (Henikoff & Ahmad, 2005; Talbert & Henikoff, 2016). The prominent examples it CENP-A, a H3 histone variant found in centromeres and essential for kinetochore assembly. CENP-A defines centromeres in all eukaryotes and strikingly, does not appear to require centromertic DNA sequence for assembly of centromeric nucleosomes and centromere identify (Andy Choo, 2001; Voullaire, Slater, Petrovic, & Choo, 1993). Of particular interest for this study are H3.3 and H2A.Z variants frequently found throughout the genome. H3.3 accounts for 15% to 20% of the total H3 histone pool (McKittrick, Gafken, Ahmad, & Henikoff, 2004) and H2A.Z represents 5 to 10 % of total H2A protein (West & Bonner, 1980). Their genomic location, however, is not random. Both variants are found preferentially in nucleosomes flanking the transcription start site, indicating their role in regulating transcription (Allis, 2007; Talbert & Henikoff, 2016). Canonical H3 deposition is restricted to the S phase of the cell cycle and coupled to replication foci. Strikingly, the difference in only four amino acids between the canonical form of H3 and the H3.3 variant results in deposition

of histone H3.3 in a replication-independent manner (Lennox & Cohen, 1988; Talbert & Henikoff, 2016). In line with this observation, H3.3 is found in coding regions of actively transcribed genes and higher levels of H3.3 in gene bodies correlate with higher transcriptional rates (Henikoff & Ahmad, 2005; McKittrick et al., 2004; Wirbelauer, Bell, & Schübeler, 2005). It suggests H3.3 merely replaces nucleosomes evicted during transcription in the absence of the canonical mechanisms for H3 incorporation. In fact, in addition to incorporation at promoters and coding regions, H3.3 compensates for histone loss at regulatory regions with high histone turnover, for instance enhancers and transcription factor binding sites (Dion et al., 2007). Measuring nucleosome dynamics by metabolic labelling revealed H3.3 patterns strongly overlap sites of high nucleosome turnover (Deal, Henikoff, & Henikoff, 2010).

In contrast to H3.3, H2A.Z is structurally divergent from its canonical counterpart and shares only 60% similarity (Thatcher & Gorovsky, 1994). H2A.Z is thought to have a role in establishing transcriptional competence, however, precise mechanisms of its effect on chromatin structure are not yet evident. H2A.Z is of particular interest as is it incorporated by a nucleosome remodeling complex SWR-C/SWR1 (Mizuguchi et al., 2004; Yen, Vinayachandran, & Pugh, 2013), a rare example of a role for remodeling complexes in histone variant dynamics. Moreover, the H2A.Z-H2B dimer is specifically removed by a related chromatin remodeling family, INO80 (Morrison & Shen, 2009; Yen et al., 2013). It is evident that histone variants have an important role in regulating various biological process, however, it will be interesting to clarify their role in regulating transcription. Histone variants such as H2A.Z and H3.3 might have a role in influencing dynamics of transcription factor binding.

DNA methylation

DNA methylation is perhaps the most studied and best characterized epigenetic modification. It is a covalent modification of cytosine by addition of the methyl at the 5’ position. In mammals, it mostly occurs in the CpG

context. DNA methylation is considered to maintain a repressed chromatin state and to stably silence genes. It is involved in mechanisms of transposable elements silencing, genomic imprinting, inactivation of the X chromosome and generally maintenance of the silent state of promoters throughout development (A. Bird, 2002). DNA methylation patterns are established early in mammalian embryonic development by *de novo* DNA methyltransferases Dnmt3a and Dnmt3b (Goll & Bestor, 2005). Once established, the methylation patterns are maintained through replication by Dnmt1 that preferentially recognizes hemimethylated DNA and methylates the newly synthesized unmethylated CpG (Goll & Bestor, 2005). Although this mechanism maintains DNA methylation patterns through numerous cell cycles, changes in DNA methylation occur during development and in adult tissues. In fact, DNA methylation plays an important role in enabling cell-type specific expression by silencing pluripotency-associated promoters during differentiation of embryonic stem cell to neuronal progenitors (Mohn et al., 2008). Surprisingly, not all inactive CpG island promoters are DNA methylated. Many CpG promoters are marked by H3K27 tri-methylation and differentially expressed during development depending on the cell-type (Deaton & Bird, 2011). Upon differentiation of embryonic stem cells to neuronal progenitors, many of the inactive H3K27me3-rich promoters acquire DNA methylation (Deaton & Bird, 2011; Mohn et al., 2008). It suggests that DNA methylation stably silences promoters in lineages that will not require reactivation.

The methylation state of single CpGs at base pair resolution can be inferred using bisulfite sequencing. Genome-wide methylation maps revealed that the majority of the genome is fully methylated with the exception of so-called CpG islands (Lister et al., 2009; Meissner et al., 2008; Stadler et al., 2011). Due to intrinsic high mutagenicity of methylated cytosines, mammalian genomes have undergone an evolutionary depletion of CpG dinucleotides. Parts of the genome that were unmethylated in the germline preserved their CpG density in the course of the vertebrate evolution and are referred to as CpG islands (Deaton & Bird, 2011). CpG islands are found in the majority

of mammalian promoters and once methylated, they correlate with the repressed state of the respective gene (A. Bird, 2002; Deaton & Bird, 2011). Genome-wide methylation maps also revealed that distal regulatory regions, mostly enhancers, exhibit intermediate methylation levels (Stadler et al., 2011). The intermediate methylation levels are a consequence of binding of transcription factors, namely CTCF and REST, indicating their role in regulating DNA methylation (Stadler et al., 2011). The opposite concept of DNA methylation being instructive for transcription factor binding has also been demonstrated; binding of CTCF and NRF1 has been shown to be blocked by DNA methylation (Domcke et al., 2015; Hark et al., 2000). Taken together, DNA methylation plays an important role in chromatin dynamics, through means of providing stable gene silencing and by modulating transcription factor binding.

1.4 Transcriptional Silencing by Polycomb Group Proteins

1.4.1 Regulation of Cell Fate by Polycomb Group Proteins

Cellular identity is created through cellular programming during development, starting from pluripotent embryonic cells that give arise to most adult cell lineages. Therefore, embryonic cells face the challenge of maintaining their self-renewal potential and executing cell-type-specific programs upon developmental decisions. The initial developmental switch, leading to a change in gene expression, is usually mediated by sequence-specific targeting of DNA-binding factors (Graf & Enver, 2009). However, once certain lineage-specific expression has been established, transcriptional information has to be mitotically heritable and maintained in the absence of the initial cue. Polycomb Group Protein (PcG) were one of the first group of proteins shown to be required for maintenance of cell-fate in *D. melanogaster* development (Lewis, 1978). They were identified as regulators of *HOX* genes, a group of a conserved family of genes that regulate body patterning during development. Particular *HOX* genes are

expressed in specific segments of the developing embryo during a specific time window (Pearson, Lemons, & McGinnis, 2005). Their misregulation leads to homeosis, transformation of one organ into another. The first identified regulator of *HOX* genes was *Polycomb* (*Pc;* Lewis, 1978; Paro & Hogness, 1991). Heterozygous mutant of *Pc* shows body segment transformations, a phenocopy of homoeotic transformations caused by ectopic *HOX* gene expression (McKeon et al., 1994). *Pc* was therefore defined as a *HOX* repressor. Subsequent studies identified other repressors of *HOX* genes and their antagonists, Trithorax Group (TrxG) proteins. TrxG proteins are responsible for maintaining the active state of *HOX* genes. Further studies showed that PcG proteins act on chromatin in the form complexes rather than autonomously, which set the basis for unveiling the mechanisms of PcG mediated silencing (Czermin et al., 2002; Shao et al., 1999). The complex composition of PcG proteins will be described in the following chapter (Chapter 1.4). Following the discovery of PcG proteins in *D. Melanogaster*, PcG homologues have been identified in the majority of other multicellular organisms. In mouse, mutations of PcG genes lead to homeotic-like transformations of vertebra, indicating a well-conserved mechanism of Poylcomb-mediated repression (Allis, 2007). Furthermore, PcG proteins have been implicated in the maintenance of pluripotency and cell-lineage specification in mammals (Beisel & Paro, 2011). Upon differentiation of mouse ES cells, a subset of inactive promoters occupied by PcG proteins become DNA methylated and permanently repressed (Margueron & Reinberg, 2011). However, not all Polycomb-decorated promoters in ES cells are maintained in their repressed state during development. A subset of Polycomb targets becomes re-activated in certain lineages, indicating a more complex mode of lineage regulation when compared to the *D. Melanogaster* model (Beisel & Paro, 2011; Jeffrey a Simon & Kingston, 2009). Polycomb-mediated repression could constitute a mechanism to reduce transcriptional noise while ensuring activation only upon strong developmental triggers (Mohn & Schübeler, 2009). A couple of studies identified a subset of upregulated genes in mouse ES cells lacking

PcG components. The PcG mutant ES cells are viable yet unable to undergo *in vitro* differentiation, emphasizing their key role in lineage commitment rather than self-renewal (Chamberlain, Yee, & Magnuson, 2008; Leeb et al., 2010).

The importance of PcG proteins in long-term transcriptional silencing was further confirmed in more recent studies. Both PRC1 and PRC2 complexes were shown to be critical for X-chromosome inactivation in mammals (Plath et al., 2003; H. Wang et al., 2004; Zhao, Sun, Erwin, Song, & Lee, 2008). During early development of female embryos, one X-chromosome is randomly chosen and inactivated to ensure dosage compensation between XX females and XY males. Inactivation of the X-chromosome is initiated by the long noncoding RNA *Xist* that recruits PcG proteins (Plath et al., 2003). The subsequent inactivation of the X-chromosome depends on PcG proteins (Plath et al., 2003; Zhao et al., 2008). This inactivation is irreversible in the lifetime of an organism, again emphasizing the role of Polycomb in mediating long-term transcriptional silencing. While it became clear that PcG proteins are essential for transcriptional repression of their target genes, the exact mechanism of repression remains unsolved.

1.4.2 Mechanisms of Silencing by Polycomb Group Proteins

Polycomb-group proteins exert their function, in part, by modifying histone tails. PcG proteins are mainly found in two classes of complexes; Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (Allis, 2007; Beisel & Paro, 2011). In particular, PRC1 activity results in monoubiquitination of lysine 119 on histone H2A (H2AK119ub) whereas the hallmark of PRC2 silencing is methylation of lysine 27 on histone H3 (H3K27me3; Beisel & Paro, 2011; Cao et al., 2002; Margueron & Reinberg, 2011; Müller et al., 2002; Wang et al., 2004).

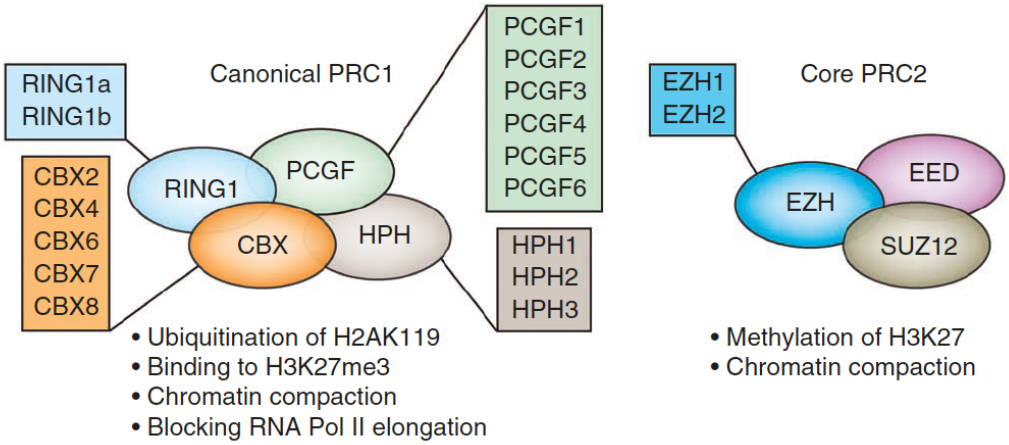


Figure 1.3 Composition of Polycomb complexes. The graphic depicts core components of canonical mammalian PRC1 complex and the PRC2 complexes. Subunit variants contribute to complex diversity (adapted from Di Croce & Helin, 2013).

Mammalian PRC2 consists of four core components; enhancer of zeste homologues 1/2 (Ezh1/Ezh2), suppressor of zeste 12 (Suz12), embryonic ectoderm development (Eed) and retinoblastoma-binding protein p4 (Rbbp4). With little variation in complex composition, the PRC2 core subunits are conserved in other organisms; *D. melanogaster*, *A. thaliana*, *C. elegans*. Ezh1/2 is the catalytic subunit of the complex. The methylation is established by the SET domain of Ezh2 that catalyzes mono-, di-, and trimethylation of H3K27 (Cao et al., 2002; Margueron & Reinberg, 2011; Müller et al., 2002). Eed has a structural role and is essential for the enzymatic activity of Ezh1/2 (Z. Han et al., 2007). Eed deficient cells show complete loss of H3K27me3 mark as shown by western blot (Leeb et al., 2010). Furthermore, Eed specifically binds the H3K27me3, indicating a role for Eed in propagation and spreading of the mark (Hansen et al., 2008). Suz12 is thought to facilitate the binding of the complex to DNA as it is the only component that harbors a DNA-binding domain (Schwartz & Pirrotta, 2007). Deletion of each PRC2 subunit is embryonically lethal, as expected due to their evident role in development (Boyer et al., 2006; Faust, Schumacher, Holdener, & Magnuson, 1995; Leeb et al., 2010; Riising et al., 2014).

The canonical PRC1 complex consists of four core components; (1) Ring finger subunits Ring1A/Ring1B, (2) chromobox protein homologue 4,6,7,8 (Cbx4,6,7,8), (3) Polyhomeotic-like 1-3 (Phc1-3) and (4) Polycomb group ring finger 1-6 (Pcgf1-6; Figure **1.3**; Beisel & Paro, 2011; Schwartz & Pirrotta, 2007). In contrast to PRC2, mammalian PRC1 complex has several homologs for each component resulting in assembly of multiple complex variants. Ring1A/B is the catalytic subunit shared among all variants (Gil & O'Loghlen, 2014). It is an E3 ubiquitin ligase that monoubiquitinates lysine 119 of histone H2A (H2AK119ub). The chromodomain of Cbx subunits recognizes the H3K27me3 mark (Fischle et al., 2003), indicating a recruiting mechanism of PRC1 to chromatin.

Despite their established importance in development and gene regulation, the precise mechanism of Polycomb-mediated silencing is not clear. It was suggested that PRC1 has a main role in Polycomb-mediated repression by inhibiting chromatin remodeling and transcription *in vitro*. Indeed, it was shown that chromatin compaction by PRC1 inhibits transcription factor binding and chromatin remodeling by SWI/SNF remodeling enzymes (Eskeland et al., 2010; Grau et al., 2011). Furthermore, a Polycomb recruiting element taken from the fly UBX gene was placed in the reporter system driven by the heat shock-inducible HSP26 promoter (Dellino et al., 2004). The UBX element repressed the reporter expression, however, binding of RNA polymerase II, TBP or heat shock factor was unaffected (Dellino et al., 2004). This results suggests the mechanism of silencing might be via blocking of transcriptional initiation.

The discovery of chromodomain-containing Cbx subunits of the PRC1 complexes led to the hypothesis that PRC1 recruitment is subsequent to H3K27me3 deposition. It suggests a role for PRC2 in recruiting PRC1 to facilitate stable silencing. However, non-canonical PRC1 complexes are deficient for Cbx subunits and are unable to bind H3K27me3 mark, suggesting a PRC2-independant mechanism of genomic targeting for PRC1 (Farcas et al., 2012; Wu, Johansen, & Helin, 2013). Yet, a recent study by Pengelly at al. demonstrated that the H3K27me3 mark itself is

indispensable for silencing of PcG target genes (Pengelly, Copur, Jäckle, Herzig, & Müller, 2013). Flies with a mutation in lysine 27 of histone H3 fail to silence PcG gene and exhibit a phenocopy of PcG mutants (Pengelly et al., 2013). To what degree PRC1 recruitment is effected is yet to be determined.

1.4.3 Genomic targeting of Polycomb Group Proteins in *D. Melanogaster*

Efficient silencing requires (1) targeting of chromatin modifiers to genes or genomic regions, (2) modifying chromatin on histone tails and (3) propagation of the silent chromatin state. In flies, Polycomb is specifically targeted to Polycomb response elements (PREs) that respond to Polycomb knockout by upregulating the associated gene (Ringrose & Paro, 2007). These elements are comprised of various transcription factor-binding motifs and are often depleted of nucleosomes, which is indicative of factor occupancy (Müller & Kassis, 2006). When placed in an ectopic location, PREs maintain the pattern set by an enhancer in the proximity. These 'swap' experiment provided first evidence that PREs function as epigenetic memory elements, and do not define body patterning autonomously (J Simon, Chiang, Bender, Shimell, & O'Connor, 1993). The patterning is defined by transcription factors that bind PREs very early in development and orchestrate the fate of each segment (Pearson et al., 2005). Transcription factors bind their targets in a very narrow window of time and the role of PcG and TrxG complexes is to maintain the transcriptional state in the absence of transcription factor binding (Allis, 2007). However, a very recent study showed the repressive state established by PREs is lost upon DNA replication (Laprell et al., 2017). After excision of PRE DNA from the *D. Melanogaster* genome, H3K27me3 levels decreased with each round of cell cycle (Laprell et al., 2017). Accordingly, repression of the reporter gene was lost. These new insights prompt us to revise the models derived from previous studies of PREs in *D. Melanogaster*.

PcG complexes are ubiquitously expressed in the fly embryo, yet they target PREs in a tissue-specific manner. Considering PcG complexes do not have DNA-binding properties, it was suggested that sequence-specific DNA-binding factors, triggered by external signaling, recruit PcG complexes to their target genes. Since their discovery, numerous studies tried to identify novel PREs in order to identify sequence determinants of PcG recruitment and subsequent gene silencing (Beisel & Paro, 2011; Ringrose & Paro, 2007). Transgenic experiments revealed three common characteristics of PREs; (1) they localize with H3K27 tri-methylation, (2) they repress a reporter gene in a transgenic setting, (3) the repression of the reporter is Polycomb-dependent (Bauer, Trupke, & Ringrose, 2016; Ringrose & Paro, 2007). However, there are no common consensus motifs within these elements that can suffice their function. It was suggested that PREs serve as a docking platform for DNA-binding factors that are capable of recruiting PcG complexes through protein-protein interactions. Accordingly, efforts were made to computationally identify transcription factors within PREs that drive the recruitment. Well over 100 PRE sequences have been computationally identified in the fly genome, with several PREs being experimentally validated to have the three defining PRE features (Ringrose, Rehmsmeier, Dura, & Paro, 2003). GAG, ZESTE, PSQ, and PHO were identified to be co-occurring in PREs (Ringrose et al., 2003). However, genome-wide mapping of PcG components with ChIP DamID assays clearly showed that only one fifth of predicted PREs account for PcG binding in the fly genome (Filion et al., 2010). It seems that transcription factors GAF, PHO and ZESTE define recruitment of only a subset of PREs.

Furthermore, until the breakthrough in genome-wide chromatin maps, pleiohomeotic (Pho) factor and its relative Pho-like (Phol) factor were thought to be the main contributors to PRE function (L. Wang et al., 2004). These DNA-binding proteins have been proven to facilitate PRC2 and PRC1 recruitment in flies. However, genome-wide maps revealed Pho-binding sites do not overlap with PRC2 and PRC1 components for the large fraction of genes. Furthermore, Pho was found to be present at genes

marked by the active histone modification H3K4me3 (Kwong et al., 2008; Oktaba et al., 2008). It was a clear indication that the recruitment model of PcG by transcription factors needs to be extended and revised.

PcG recruitment in *D. melanogaster* appears quite complex and might involve a combinatorial network of various DNA binding factors. It seems there is not a consensus sequence that drives PcG recruitment. Perhaps, a protein-function consensus of factors that bind PREs might be a more appropriate approach to identify recruiting mechanisms. When examining known PRE-binding factors, it is evident they entail a variety of chromatin-modifying functions. Gaf mediates displacement of nucleosomes and opens chromatin for transcription factors to bind; Pho was shown to bend DNA; several chromatin remodelers and high-mobility group proteins have been shown to bind PREs (Margueron & Reinberg, 2011; Schwartz & Pirrotta, 2007). Redundancy of function at different PREs might be adding complexity to the system making it difficult to decipher. Recruitment of PcG complexes is all the more complex in the mammalian system and will be discussed in the following chapter.

1.4.4 Genomic targeting of Polycomb Group Proteins in mammals

In mammals, PREs are not yet characterized, partially due to a lack of robust reporter assays for Polycomb-mediated repression. Only two large fragments that mimic *D. Melanogaster* PREs were recently described in mammalian cells, both containing ying and yang 1 (YY1) binding sites (Sing et al., 2009; Woo, Kharchenko, Daheron, Park, & Kingston, 2010). Both so-called mammalian PREs, *D11.12* and *PRE-kr*, are rather large fragments of 1.8 kb and 3 kb in size, respectively. Both sequences recruit PcG proteins when placed ectopically, repress activity of the associated reporter and the silencing is PcG-dependent. Furthermore, *D11.12* PRE is able to maintain the repression throughout differentiation (Woo et al., 2010). These results show, for the first time in mammals, a memory-based mechanism similar to fly PREs (Woo et al., 2010). However, identification of the two PREs failed

to explain recruitment mechanism of other PcG targets and did not enable prediction of other mammalian PREs.

These studies, and others, suggest that YY1, the mammalian orthologue of *D. Melanogaster* Pho factor, is required for Polycomb silencing in mammals as it is present in both mammalian PREs (Atchison, Ghias, Wilkinson, Bonini, & Atchison, 2003; L. Srinivasan & Atchison, 2004; Woo et al., 2010). However, the fact that genome-wide analysis shows no clear overlap of YY1 with H3K27 methylation questions its general recruiter properties (Vella, Barozzi, Cuomo, Bonaldi, & Pasini, 2012). Furthermore, YY1 was shown to mainly bind active regions of the genome making it highly unlikely to be a recruitment signal for PcG complexes (Vella et al., 2012). Other DNA-binding factors, JARID2 and AEBP2, were recently suggested to be required for recruiting PRC2 to a subset of Polycomb targets in ES cells. Both have been co-purified with core PRC2 components and shown to have an overlap with PRC2 binding targets, suggesting a role in PRC2 recruitment (Landeira & Fisher, 2011; Peng et al., 2009; Son, Shen, Margueron, & Reinberg, 2013). Yet, depletion of JARID2 shows only a mild effect on global K27me3 levels (Pasini et al., 2010).

Another example describing the role of transcription factors in PRC2 recruitment involves transcription factors Rest and Snail. It was shown that promoter sequences with Rest or Snail motifs are sufficient to recruit PRC2 (P. Arnold et al., 2013). Furthermore, deletion of the respective motif within the promoter sequences resulted in loss of H3K27me3. This clearly indicates Rest and Snail are involved in PRC2 recruitment. However, they cannot explain PRC2 recruitment at most other PRC2 targets that lack binding sites for Rest and Snail.

Interestingly, genome-wide studies unveiled that Polycomb targets in mammalian genomes are exclusively CpG islands (unpublished data from the group; Ku et al., 2008). The *D. Melanogaster* genome has no such features as CpG islands evolved only in genomes with highly abundant DNA methylation (Deaton & Bird, 2011). It is possible, although remains to be proven, that CpG islands are the sole recruiting signal for Polycomb in

mammals. Unlike in *D. Melanogaster*, mammalian H3K27me3 domains are often discreet and H3K27me3 perfectly overlaps with localization of PRC2 components (Suz12, Ezh2; unpublished data from our group; Pasini et al., 2010; Tiwari et al., 2012). K27me3-rich domains are rarely found in distal genomic regions, indicating that recruitment is mostly established through promoter blocking, inhibiting RNA Pol II elongation or other proximal *cis*-regulatory elements. To understand the mechanisms of Polycomb-mediated repression, it will be important to elucidate how PcG complexes are recruited to CpG islands. Surprising findings demonstrated that a CpG rich *E. coli* sequence is able to recruit the PRC2 complex in mouse ES cells (Jermann, Hoerner, Burger, & Schübeler, 2014; Mendenhall et al., 2010). In an elegant study by Riising et al., transcriptional inhibition was shown to induce genome-wide ectopic PRC2 binding to endogenous PcG target CpG islands found in other tissues (Riising et al., 2014). This study indicates that PRC2 is recruited by default to CpG islands in the absence of transcription. It is yet to be determined what feature of the CpG island is the recruiting signal and if a certain density of CpG motifs is sufficient to autonomously drive PRC2 to target CpG islands. JARID2 and AEBP2, CpG-binding factors that co-purify with PRC2, are the most likely candidates to autonomously recruit PRC2 to CpG islands. Furthermore, PRC2 recruitment could be mediated by more than one mechanism; one that is CpG dependent and one that is mediated by other transcription factors (Figure **1.4**). Part of the presented project will address aspects of that question.

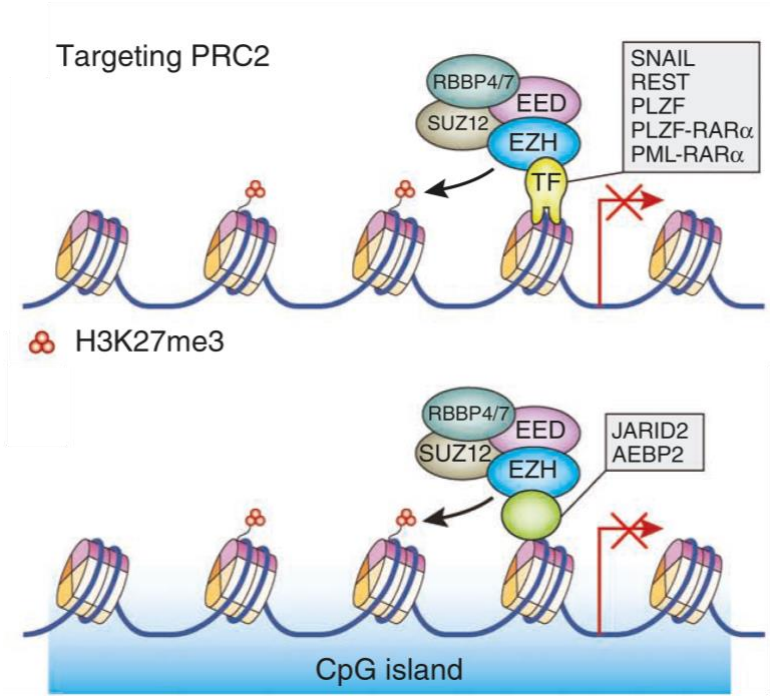


Figure 1.4 Models of PRC2 targeting in mammalian genomes. The graphic depicts two main models proposed for PRC2 recruitment. CpG-independent transcription factors have been shown to be involved in PRC2 recruitment (upper panel). H3K27me3 is recognized by Eed and might serve to propagate the PRC2 binding and subsequent spreading of the mark. An alternative model suggests CpG-dependant recruitment is mediated through CXXC-domain proteins, such as JARID2 and AEBP2 (adapted from Di Croce & Helin, 2013).

1.5 Nucleosome Remodeling and Chromatin-Remodeling Complexes

One fundamental question in epigenetics and gene regulation in eukaryotes is how to enable access to DNA when required yet retain the compact and repressive structure of chromatin. Chromatin repression in eukaryotes exists on two levels. (1) Chromatin can be modified to recruit machineries that repress certain genomic regions. This repression is established by folding of the nucleosome fiber into higher order structures or by directly preventing binding of activators (Allis, 2007). One such example is the PcG system described preciously (Chapter 1.4). (2) The other layer of chromatin repression is achieved by the mere nature of the chromatin fiber. Nucleosomes mask binding sites for transcription factors. An additional barrier to binding comes from DNA being strongly bent within the nucleosomes structure (Bai & Morozov, 2010; Struhl & Segal, 2013). To be able to maintain the compact structure yet ensure access when required, eukaryotic genomes use chromatin remodeling complexes. They use energy of ATP hydrolysis to evict, slide, insert or change nucleosomes to enable dynamic binding to chromatin (Allis, 2007). As they catalyze a fundamental process in chromatin dynamics, they are involved in every aspect of genome utilization; cell-fate regulation, transcription factor binding, genome stability, replication, DNA damage, etc (Allis, 2007). Being so instrumental in genome regulation, it is not surprising that mutations in chromatin remodeling complexes are amongst the most frequent ones in cancer (Kadoch & Crabtree, 2015). The next chapter will describe the role of chromatin remodelers, the ISWI family in particular, in chromatin dynamics.

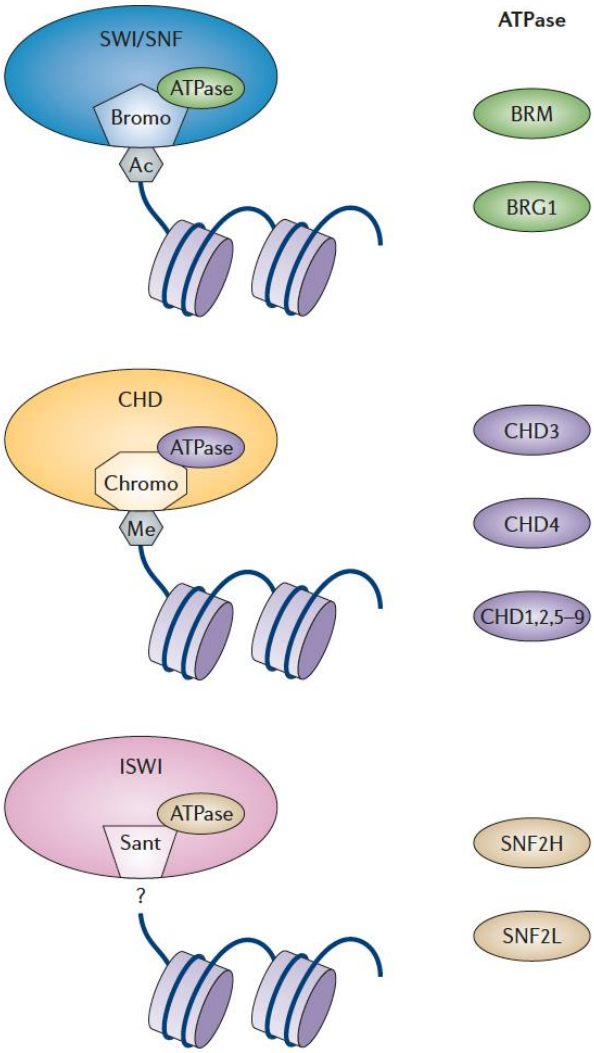


Figure 1.5 ATP-dependent chromatin remodelers. The illustration depicts the three best characterized chromatin remodeler families. Each family is represented by several ATPase subunits that form different complexes. In mammals, BRM and BRG1 of the SWI/SNF family form BAF, PBAF and ES-specific esBAF. SNF2H of the ISWI family is found in ACF, cHRAC, NoRC, RSF, and WICH complex. SNF2L of the same family is found in the NURF complex. CHD family consists of the NuRD and CHD1 complexes, with the NuRD complex harboring various ATPase subunits. Each family has a characteristic domain within the ATPase subunit (bromo, chromo, saint). These domains interact with specific chromatin substrates and are thought to be involved in their targeting and function. SANT-SLIDE domains of ISWI is the least characterized (adapted and combined from Clapier & Cairns, 2009; de la Serna, Ohkawa, & Imbalzano, 2006).

1.5.1 Diversity of Chromatin-Remodeling Complexes

To date, four divergent chromatin-remodeling families have been described; SWI/SNF, ISWI, CHD and INO80 (Clapier & Cairns, 2009). The unifying feature of all four families is the presence of a highly conserved catalytic ATPase domain within the ATPase subunit. Furthermore, all remodeler families share the ability to translocate DNA and disrupt the association of histones and DNA (Narlikar et al., 2013; C. Y. Zhou, Johnson, Gamarra, & Narlikar, 2016). What distinguishes different families is the composition of functional domains and subunits. Chromatin remodelers are multi-subunit

complexes, with each family having a unique set of subunits (Clapier & Cairns, 2009; de la Serna et al., 2006). The other distinguishing feature is the presence of distinctive domains within the ATPase subunit flanking the catalytic domain. Chromodomains, for instance, are found specifically in the catalytic subunit of CHD complexes, while bromodomains are characteristic for the SWI/SNF family (Clapier & Cairns, 2009; de la Serna et al., 2006). All remodeling complexes are conserved from yeast to human with variations in subunit composition (Clapier & Cairns, 2009). However, it is important to notice the key functional domains within the catalytic subunits are highly conserved, suggesting that the catalytic domain has a fundamental function.

SWI/SNF Family Remodelers

SWI/SNF complex, with the respective catalytic subunit Swi2/Snf2 (Swi2/Snf2 in yeast, Brm in fly, Brg1 and Brm in mammals), was one of the first chromatin remodeling complexes described (Allis, 2007; Clapier & Cairns, 2009). It was genetically identified in yeast; genes coding for subunits of the SWI/SNF complex were found required for mating-type switching (SWI) and for sucrose fermentation (SNF), hence the name SWI/SNF (switching defective/sucrose nonfermenting; Allis, 2007; Clapier & Cairns, 2009). SWI/SNF controls mating-type switching and sucrose fermentation by maintaining expression of the HO endonuclease gene and SUC2 invertase gene, respectively. Ever since its initial discovery, the SWI/SNF complex was thought to positively regulate transcription. This became clearer with the first purification of the complex from *S. cerevisiae*. Purified SWI/SNF was shown to disrupt nucleosomes in an ATP-dependent matter and to enable transcription factor binding *in vitro* (Fry & Peterson, 2001; Vignali, Hassan, Neely, & Workman, 2000). This disruption was followed by an increase in sensitivity to digestion by DNaseI in nucleosome arrays, indicating loss of nucleosome and formation of an open chromatin structure (Fry & Peterson, 2001; Vignali et al., 2000). It was not until later that SWI/SNF family was also shown to function as a transcriptional

repressor. Yeast SWI/SNF complex was shown to directly repress transcription of the SER3 gene in *S. cerevisiae*, and strikingly, only the SWI/SNF ATPase subunit Snf2 is required for the repression (Martens & Winston, 2002). This indicates the nucleosome remodeling process itself is causative of repression. Other studies provided additional examples of SWI/SNF mediating transcriptional repression (Clapier & Cairns, 2009). Furthermore, SWI/SNF was even reported to be mediating the switch between activation and repression at the same gene (Chi et al., 2002). Although SWI/SNF complexes had a well-defined role in gene activation, these studies redefined their function. The revised model would suggest SWI/SNF enables binding of transcription factors, both activators and repressors.

The main catalytic subunit of SWI/SNF complexes contains a bromodomain that binds acetylated histone tails (Clapier & Cairns, 2009; de la Serna et al., 2006). A long-standing question is whether the bromodomain has a function in recruiting the complex to chromatin or if its function is to increase remodeling efficiency. Several studies suggested that the bromodomain has a role in both aspects. A single bromodomain within the catalytic subunit of SWI/SNF complex in yeast is necessary for the retention of the complex at the SUC2 gene (Hassan et al., 2002). Also, another subunit of the SWI/SNF-family remodeler RSC was shown to be stimulated by H3K14 acetylation leading to gene activation (Kasten et al., 2004). If acetylation is one of the recruiting mechanism for SWI/SNF, the complex should localize to acetylated regions of the genome, such as active promoters and enhancer. Indeed, genome-wide ChIP–seq studies showed that esBAF, the mammalian SWI/SNF complex found in ES cells, is found at one-quarter of all promoters (Ho et al., 2011; Kidder, Palmer, & Knott, 2009). Interestingly, binding of the complex positively correlates with the expression level of genes (Ho & Crabtree, 2010).

Having such an essential role in chromatin regulation, it is expected that deletion of SWI/SNF subunits will have a tremendous effect on development. Indeed, deletion of Brg1, the catalytic subunit of mammalian

SWI/SNF, causes pre-implantation lethality in mice (Bultman et al., 2000). Deletion of other non-catalytic subunits of the SWI/SNF complex have various effects on the phenotype (Ho & Crabtree, 2010; Kim et al., 2001; Klochendler-Yeivin et al., 2000). Remarkably, they all seem to show a developmental phenotype in a specific tissue or lineage (Ho & Crabtree, 2010). This is of particular interest as studies of the mammalian SWI/SNF family show that these complexes change their subunit composition substantially during development. It is interesting to speculate that the diversity of assemblies in different cell types enables the complex to adapt its targeting to a new set of genes. For example, two catalytic subunits are found in the mammalian SWI/SNF complexes, Brg1 and Brm. An ES cell specific SWI/SNF complex is found in mouse, called esBAF, which contains Brg1 but not Brm subunit and Baf155 but not Baf170 subunit (Ho et al., 2011; Yan et al., 2008). During ESCs differentiation into neuronal progenitors, Brm and Baf170 are incorporated into the complex, potentially directing the complex to new targets (Ho et al., 2011; Yan et al., 2008). It was shown that esBAF has an essential role in regulating the core pluripotency network of in ES cells (Ho & Crabtree, 2010). It is, for instance, required for silencing of *Nanog*, a transcription factor that maintains the ES cells in the pluripotent state (Ho et al., 2009; Kidder et al., 2009; Yan et al., 2008). A role in development was shown for Brahma (dBrm), the Brg1/Brm homolog in flies (Brown, Malakar, & Krebs, 2007). Interestingly, dBrm was identified in genetic screen for suppressors of homeotic transformations caused by mutations in PcG genes, as mentioned previously. It was classified as TrxG protein member and shown to be essential for body segmentation further emphasizing the important role of SWI/SNF complexes in development.

CHD Family Remodelers

CHD (Chromodomain, Helicase and DNA-binding protein) remodeling family has been identified in most eukaryotic organisms (Allis, 2007). Its characteristic features are the two tandem chromodomains in the catalytic

subunit (Allis, 2007; Clapier & Cairns, 2009; de la Serna et al., 2006). Additional structural domains and subunits further differentiate the CHD family into distinct complexes; CHD1 and NuRD are the most well-studied and others complexes have not yet been fully characterized. CHD1 is often associated with active transcription as it co-localizes with the active form of RNA Pol II (S. Srinivasan et al., 2005). Moreover, CHD1 was shown to interact with the histone acetyltransferase complexes in mammals (Pray-Grant, Daniel, Schieltz, Yates, & Grant, 2005), which demonstrates the role of CHD1 in promoting an open chromatin structure.

The other CHD complex, NuRD, includes ATPase subunits CHD3 and CHD4 (referred to as Mi-2α and Mi-2β, respectively; Clapier & Cairns, 2009). Unlike CHD1, they lack a DNA binding domain and contain a pair of PHD domains (Clapier & Cairns, 2009). It has been shown that the NuRD complex is targeted to specific genes by its interaction with transcription factors. The Mi-2 ATPase subunit of *D. Melanogaster* NuRD complex is recruited by hunchback (Hb), a transcription factor important for repression of *HOX* genes during development (Kehle et al., 1998). This is another example of the role of chromatin remodelers in development. Indeed, NuRD was shown to be involved in transcriptional repression during development in *C. elegans*, *D. melanogaster* and mammals (Lai & Wade, 2011). Purification of NuRD complexes revealed that they contain, alongside the Mi-2 ATPase subunit, histone deacetylases 1 and 2 (HDAC1 and HDAC2; Zhang, LeRoy, Seelig, Lane, & Reinberg, 1998). Presence of HDACs in the complex indicates a repressive role for NuRD. Furthermore, some NuRD complexes contain methyl-CpG-binding-domain proteins (MBD; Le Guezennec et al., 2006). It is thought that MBD proteins recruit NuRD to remodel and deacetylate nucleosomes with methylated DNA therefore establishing a repressive chromatin state. This model was, however, recently challenged by our group; Baubec et al. demonstrating that MBD2 and MBD3 proteins show methyl-CpG-idependant recruitment (Baubec et al., 2013).

Remodeling by Mi-2 might be necessary for histone deacetylase subunits to gain access to histones to efficiently deacetylate and repress target genes. The proposed mechanistic model suggests; (1) NuRD is recruited to target genes by transcription factors and MBD proteins, (2) Mi-2 remodels nucleosomes at target genes exposing histone tails for deacetylation, (3) which creates a more compacted chromatin structure of repressed chromatin.

1.5.2 Mechanism of Remodeling and Transcriptional Regulation by Imitation Switch Family

ISWI family remodelers use two catalytic subunits to build numerous complexes. The ISWI catalytic subunits in yeast are Isw1a/b and Isw2, in *D. Melanogaster* ISWI and in mammals Snf2l and Snf2h (Clapier & Cairns, 2009; de la Serna et al., 2006; Vignali et al., 2000). In addition to catalytic subunits, specialized non-catalytic factors are found in ISWI complexes. They often contain chromatin-binding domains; DNA-binding histone fold motifs (Chrac15 and Chrac17), plant homeodomains (PHD), bromodomains (Bptf and Acf1), and additional DNA-binding motifs (HmgI; Clapier & Cairns, 2009). These highly diverse subunits together with Snf2h form five remodeler complexes in mammals; ACF, CHRAC, NoRC, WHICH and RSF (Clapier & Cairns, 2009; D. F V Corona & Tamkun, 2004). Unlike ACF and CHRAC, the other Snf2h-containing complexes (NoRC, WHICH and RSF) harbor only one non-catalytic subunit. Snf2l, on the other hand, forms one complex, NURF (Clapier & Cairns, 2009). Snf2h-containing complexes are generally thought to optimize nucleosome spacing and promote chromatin assembly in order to establish a repressive chromatin state. In an *in vitro* assay, ISWI is able to organize disordered nucleosomes into an array of evenly spaced nucleosomes (Längst & Becker, 2001). In fact, the catalytic subunit of ISWI is sufficient to perform this function (Davide F V Corona et al., 1999), indicating that other non-catalytic subunits have a role in modulating the activity or genomic targeting. The assembly of nucleosome arrays is thought to be exerted through nucleosome sliding by the

coordinated action of the ATPase and SANT-SLIDE domains. Moreover, the repressive chromatin state of Snf2h targets is further established by recruitment of co-factors (eg., HDACs, DNMTs; Corona et al., 1999). Indeed, ISWI in both yeast and *D. Melanogaster* was shown to interact with the HDAC complex (Rpd3) in order to promoter transcriptional repression (Burgio et al., 2008). Yeast ISWI ATPase Isw2 is thought to constitutively assemble nucleosomes in an array and slide nucleosomes over the TSS to promote transcriptional repression (Whitehouse & Tsukiyama, 2006). A recent study (Yen, Vinayachandran, Batta, Koerber, & Pugh, 2012) showed that yeast ISWI ATPases Isw1a and Isw2 in fact push nucleosomes towards the NFR regions within promoters, further demonstrating the repressive nature of ISWI complexes. Whether nucleosome repositioning towards the midpoint of the NRF comes prior to transcriptional silencing is still an open question. Nucleosome ejection and subsequent gene activation is established by antagonistic activity of other remodelers, such as SWI/SNF family members (Figure **1.6**).

The characteristic structural domain within ATPase subunits of the ISWI family is the SANT-SLIDE tandem domain (Clapier & Cairns, 2009; D. F V Corona & Tamkun, 2004). The SANT domain is found in many DNA-binding proteins and it facilitates the interaction with histones (Davide F V Corona et al., 1999). Juxtaposed to the SANT domain is the SLIDE domain that contacts nucleosomal DNA (Clapier & Cairns, 2009; Davide F V Corona et al., 1999). The tandem organization of the two domains is uniquely found in the ISWI remodelers. Interestingly, ATPase activity and subsequent nucleosome remodeling of the complex depends on both SANT and SLIDE domains, as shown by deletion mutants for *D. Melanogaster* ISWI (Grüne et al., 2003). Therefore, the tandem SANT-SLIDE domain acts to stimulate binding of the complex to the nucleosome and enhances ATPase activity.

As previously mentioned, the ISWI family entails at least six different complexes. CHRAC and ACF complexes were shown to organize nucleosome arrays and nucleosome spacing following replication-coupled chromatin assembly (Clapier & Cairns, 2009; D. F V Corona & Tamkun,

2004). In comparison to ACF, CHRAC contains two additional histone-fold subunits, Chrac15 and Chrac17 (Clapier & Cairns, 2009). The two additional subunits are thought to enhance sliding and remodeling activity. In fact, numerous studies show that the non-catalytic subunits regulate Snf2h activity in the CHRAC and ACF complexes by interacting with the linker DNA (Dang & Bartholomew, 2007). The N-terminal tail of histone H4 seems to be essential for nucleosome sliding by ISWI complexes. Interestingly, the hydrophilic patch (aa 17-19) in H4 that the ISWI activity depends on is part of H4 that interacts with nucleosomal DNA (Allis, 2007).

Another important ISWI complex is the NoRC complex with a well-established role in rDNA silencing (Clapier & Cairns, 2009). It was reported that the transcription terminator factor TTF-I recruits NoRC to RNA Polymerase I promoters to slide a promoter-bound nucleosome into a transcriptionally unfavorable position (Strohner et al., 2004). Interestingly, this transcriptional silencing of rDNA genes and the recruitment of HDAC and DNMT complexes is H4K16ac dependent. The PHD finger/bromodomain of Tip5 within the NoRC complex interacts with H4K16ac (Y. Zhou & Grummt, 2005). This again shows that the ISWI complexes are recruited by non-catalytic subunits to exert nucleosome remodeling where needed. The only Snf2l-harboring complex NURF is, however, involved in gene activation. The NURF complex disrupts nucleosomes to facilitate gene activation by assisting RNA Pol II activation (Clapier & Cairns, 2009).

1.5.3 Regulation of Chromatin Accessibility by SWI/SNF and ISWI in Mammals

It is apparent that chromatin remodelers have a role in nucleosome re-positioning *in vitro*. However, whether nucleosome mobility influences transcription factor binding *in vivo* is still unclear, especially in the mammalian system. It has been difficult to demonstrate the role of chromatin remodelers in genome accessibility due to the lack of deletion models and binding maps of their ATPases. One of the rare deletion model

described in mammalian system is a conditional deletion of Brg1 (Brg1 cko; Ho et al., 2011; Ho & Crabtree, 2010), a SWI/SNF family ATPase. *Brg1 cko* mouse ES cells lose pluripotency and die within days upon induction of the deletion (verbal communication with Prof. Gerald Crabtree, Ho et al., 2011; Ho & Crabtree, 2010). However, within a short time window after the induction, Ho et al. show that the loss of Brg1 reduces accessibility of functional binding sites for Stat3 (Ho et al., 2011). They show esBAF, the ESC-specific BAF complex, facilitates Stat3 binding and Stat3-activated transcription necessary for ES cell pluripotency (Ho et al., 2011). In the absence of Brg1, Stat3 is activated but is unable to bind to most of its target sites. This is likely due to decrease in chromatin accessibility as Stat3 target genes are silenced by PcG complexes in the Brg1 deletion (Ho et al., 2011). This is in line with the antagonizing nature of TrxG proteins (Brg1) and PcG proteins discussed previously (Chapter 1.4.1). However, esBAF seems to also cooperate with PcG at a subset of genes to facilitate *Hox* cluster repression (Ho et al., 2011). Taken together, these results indicate Brg1 creates an open chromatin structure permissive for binding of both, transcriptional activators and repressors. As a matter of fact, a very recent comprehensive mapping revealed the remodeler subunits Brg1, Chd1, Chd4, Chd6 and Chd8 highly correlate with open regions of the genome as defined by the DNaseI-seq (de Dieuleveult et al., 2016). This further indicates chromatin remodelers might not instruct transcriptional outcomes but rather enable binding of factors that do.

Compared to the SWI/SNF family, the *in vivo* role of ISWI family remodelers in chromatin accessibility is even less clear. A null mutation of Snf2h in mouse is lethal (pre-implantation lethality) since the inner-cell mass of embryos is apoptotic, restricting our tools to study the role of Snf2h *in vivo*. As previously described (Chapter 1.5.2), *in vitro* studies show ISWI family acts as a repressor by organizing evenly spaced nucleosome arrays. This might be established by creating a stiff structure of nucleosome arrays that reduce accessibility of DNA-binding factors to their respective motifs. In this

project, we aimed to test if the role of Snf2h *in vivo* is indeed restricting transcription factor binding by altering nucleosome positions.

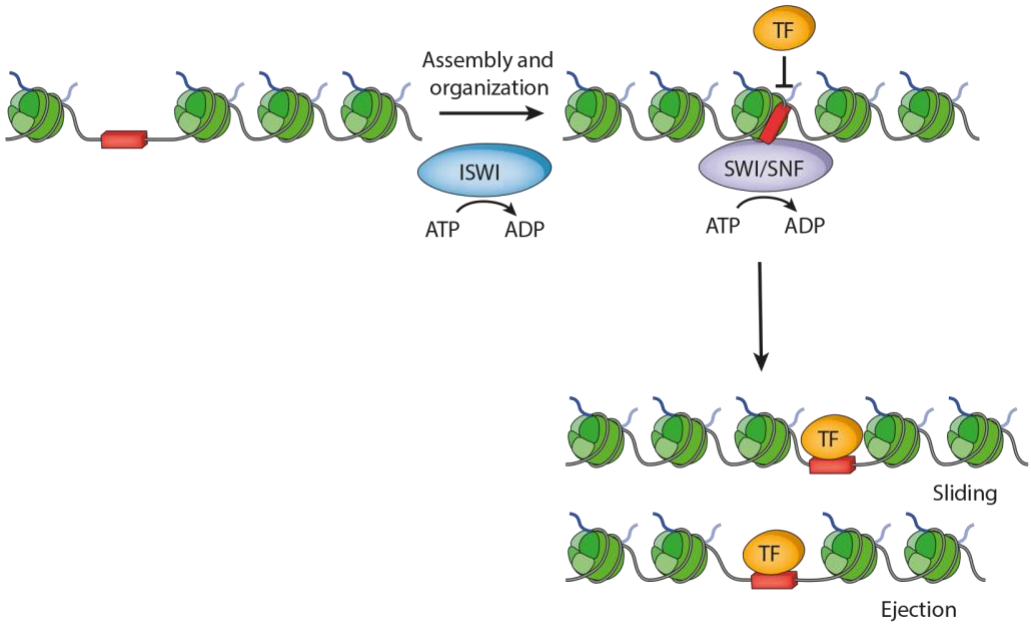


Figure 1.6 Model of chromatin reorganization by remodelers. Chromatin remodelers use energy of the ATP hydrolysis to mobilize nucleosomes. Nucleosomes can be repositioned and evicted. ISWI family remodelers (except for NURF and Isw1b) assemble nucleosome arrays and organize the repressive structure, blocking transcription factor binding. In contrast, SWI/SNF family remodelers slide or evict nucleosomes to enable transcription factor binding. These models are largely based on *in vitro* data (adapted from Cairns, 2009).

Chapter 2
Scope of the Thesis

Eukaryotic genomes are organized in the form of chromatin. Chromatin is established by wrapping the DNA around a histone octamer, which in turn compacts DNA and restricts its accessibility. The repressive nature of the chromatin structure provides a basal level of repression genome-wide. The eukaryotic cell therefore faces a challenge of how to enable access to DNA while preserving the repressive structure of chromatin. Nucleosomes themselves are highly stable and show limited mobility. Nucleosome remodeling and modifying complexes change that structure by either evicting, inserting or sliding a nucleosome or by changing the nucleosome structure itself. However, how these complexes are targeted to specific genomic locations and how they influence chromatin accessibility is not fully understood.

Chromatin modifying complexes, such as Polycomb group proteins, add or remove covalent modifications at specific residues on the histone proteins. For example, modification of chromatin by the Polycomb Repressive Complex 2 recruits other complexes, such as Polycomb Repressive Complex 1, that further change the chromatin structure (Margueron & Reinberg, 2011). However, how Polycomb group proteins are targeted in mammalian genomes is not fully understood. It is thought that transcription factors recruit Polycomb to their binding motifs (Schwartz & Pirrotta, 2007). Therefore, we aimed to decipher the mechanism of Polycomb recruitment by assaying many sequences with varying sequence properties for their ability to autonomously recruit Polycomb. This approach enabled us to identify transcription factor motifs that contribute to Polycomb recruitment. Furthermore, we aimed to link the Polycomb recruitment with gene silencing

by investigation the transcriptional response upon loss of Polycomb in mouse embryonic stem cell and during differentiation.

Chromatin remodeling complexes, unlike Polycomb group proteins, exert their function on chromatin in a more direct fashion. These complexes use energy from the ATP hydrolysis to change positions of nucleosomes, which might directly enable or prevent binding or transcription factors. However, how is nucleosome repositioning by chromatin remodelers linked to changes in chromatin accessibility is not clear. To ask if loss of remodeler activity changes nucleosome positioning and transcription factor binding, we deleted Snf2h, the ATPase subunit of ISWI remodeling complexes, in mouse ES cells. Following depletion, we monitored changes in nucleosome positioning by MNase-seq, transcriptional changes by RNA-seq and changes in transcription factor binding by either ChIP-seq or ATAC-seq. This approach enabled us to determine the role of ISWI complexes in regulating chromatin structure.

Chapter 3

Results

3.1 Recruiting and Silencing Mechanisms of Polycomb Repressive Complex 2

In this project, we focused on gaining insight into two key aspects of the Polycomb system. Namely, how is Polycomb recruited to specific sites and how does it mediate transcriptional repression in mammals? Even in flies, there are no common consensus motifs that have the potential to autonomously drive Polycomb recruitment. In light of this apparent complexity, an approach to test the ability of multiple sequences to drive Polycomb recruitment seems necessary. Such a system can be employed to systematically identify DNA sequence determinants of recruitment, and in a separate step, link recruitment with gene repression. Our working hypothesis suggests that Polycomb recruitment is encoded in the underlying DNA sequence (P. Arnold et al., 2013; Jermann et al., 2014). We aimed to identify DNA sequence features that determine Polycomb recruitment in mouse ES cells. Towards this aim, we investigated the contribution of transcription-factor motifs, CG density and DNA methylation to Polycomb recruitment as well as their impact on transcriptional repression. Furthermore, to link recruitment with repression, we investigated the transcriptional response to loss of Polycomb in different cellular systems.

3.1.1 Assay Development: Chromatin Immunoprecipitation on Hundreds of Integrated Genomic Sequences in Parallel

Our group recently showed how placing an endogenous Polycomb recruiting sequence (as short as 300bp) into a different locus reconstitutes Polycomb binding and endogenous H3K27me3 levels (P. Arnold et al., 2013; Jermann et al., 2014). Mutating specific motifs of transcriptional repressors within the sequence (such as REST or SNAIL) decreases Polycomb binding (P. Arnold et al., 2013). However, there are no consensus motifs or combination of motifs that have the potential to autonomously drive Polycomb recruitment. Therefore, we aimed to identify whether specific sequence features, such as transcription factors binding motifs or CpG density, have the potential to autonomously recruit PRC2. Building on recent technical advances and findings in the lab (Arnold et al., 2013; Krebs, Dessus-Babus, Burger, & Schübeler, 2014; Lienert et al., 2011), we decided to build a system to systematically test the ability of multiple sequences to drive Polycomb recruitment (Figure 3.1). The first step in the approach consisted of computationally designing primer sequences to amplify endogenous regions of interest (Krebs et al., 2014). The amplified control and PRC2 target regions were pooled and cloned into a vector to create the fragment library that is inserted into mouse ES cells. Fragments from the plasmid library were inserted into a defined genomic site using recombination-mediated cassette exchange (RMCE; Lienert et al., 2011; Stadler et al., 2011).

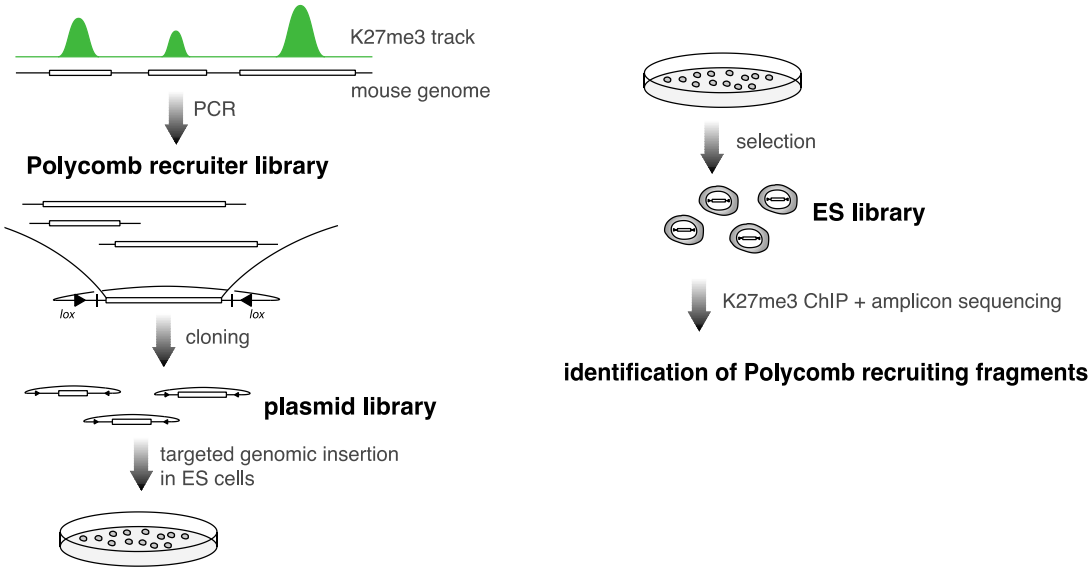


Figure 3.1 Library ChIP-seq workflow. Illustration of the workflow for identifying Polycomb recruiting fragments with parallel-targeted genomic integration coupled with chromatin immunoprecipitation (library-based ChIP-seq). PCR primers for regions of interest are computationally designed. Following the PCR, hundreds of fragments are pooled and cloned into a vector. Plasmid pool library is used to insert the fragments into mouse ES cells. Insertion is established using cassette exchange. Each fragment is inserted into the same, specific genomic location. ES cell pool with inserted library is used in a ChIP targeting a protein of interest (Suz12 or H3K27me3 in this case). ChIP enrichments are determined by PCR amplification of the insertion site and subsequent Illumina sequencing on MiSeq. Input fraction is used to normalize the data.

The genomic site for targeted insertion is located within the β-globin locus, which is deprived of CpG islands and repressive as well as activating chromatin modifications (Lienert et al., 2011). This inert chromatin environment is preserved during *in vitro* differentiation of ES cells to neuronal progenitors and terminal neurons (Bibel et al., 2004) since the locus is only active during erythropoiesis (Fromm & Bulger, 2009). Following library insertion, ChIP is performed on the ES library of inserted fragments. To only enrich for inserted fragments, the insertion site is PCR amplified, in both input and IP fraction, using a universal primer pair. The universal primer pair is used to amplify all inserted fragments simultaneously.

Following size selection and library preparation, amplicons were multiplexed and sequenced on the Illumina Miseq platform. A customized computational pipeline we developed was used to process and analyze the data. Enrichments of fragments from the library are shown as enrichment of library-size normalized IP read count over input read count.

Benchmarking library ChIP-seq

To test if our library ChIP-seq approach recapitulates binding of a protein of interest in a pool of inserted fragments, we first benchmarked our method. We assayed a pilot library of sequences that contain defined binding motifs for factors known to give high enrichments in ChIP experiments (CTCF and REST; Figure 3.2). The major experimental challenge of the approach was to enrich for the insertion fragment against endogenous targets in the pool of cells. By thoroughly optimizing protocols for amplification (varying PCR conditions, minimizing number of PCR cycles, etc.) we were able to quantify binding at the insertion site with high reproducibility (Figure **3.3**). Hierarchical clustering of ChIP-seq data (Figure **3.3**) showed perfect separation of fragments containing a CTCF motif from the fragments containing a REST motif or no motif (control regions). This demonstrated that the library ChIP-seq approach is able to specifically enrich for fragments bound by the transcription factor of interest.

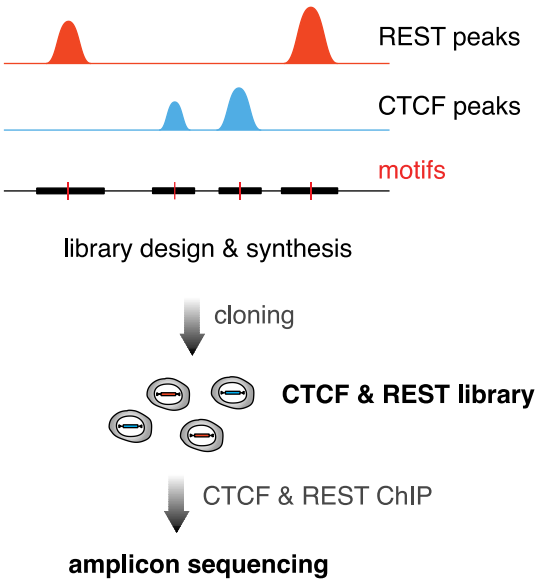


Figure 3.2 CTCF and REST library ChIP-seq workflow. Illustration of the workflow for benchmarking the library ChIP-seq approach using fragments with CTCF and REST motifs as well as control regions not harboring transcription factor motifs.

To further test quantitative accuracy of the library ChIP-seq approach, we performed quantitative PCR (qPCR) to quantify the enrichment for each library fragment individually. The dynamic range of fragment enrichment obtained using qPCR was not as large as the enrichment derived from next-generation sequencing, limiting our further analysis (Figure 3.4 A). The sensitivity of qPCR to quantitatively determine enrichments in the pool of fragments seemed to be the limiting step. However, two observations clearly indicated that the library ChIP-seq captures enrichments present in the qPCR; (1) REST ChIP enrichments of quantitative PCR compared to library ChIP-seq show a positive correlation (0.43; Pearson), (2) enriched fragments in the REST ChIP are shared between both methods (x and y values >0; Figure 3.4 A). Furthermore, qPCR enabled us to identify the critical fragment coverage in the input fraction required to confidently call enrichments (>11 in log2 space for REST ChIP; Figure 3.4 B). In summary, we established the first method that enables assaying ChIP enrichments for hundreds of inserted sequences in parallel. We further used the method to quantify enrichments in the library of putative Polycomb fragments.

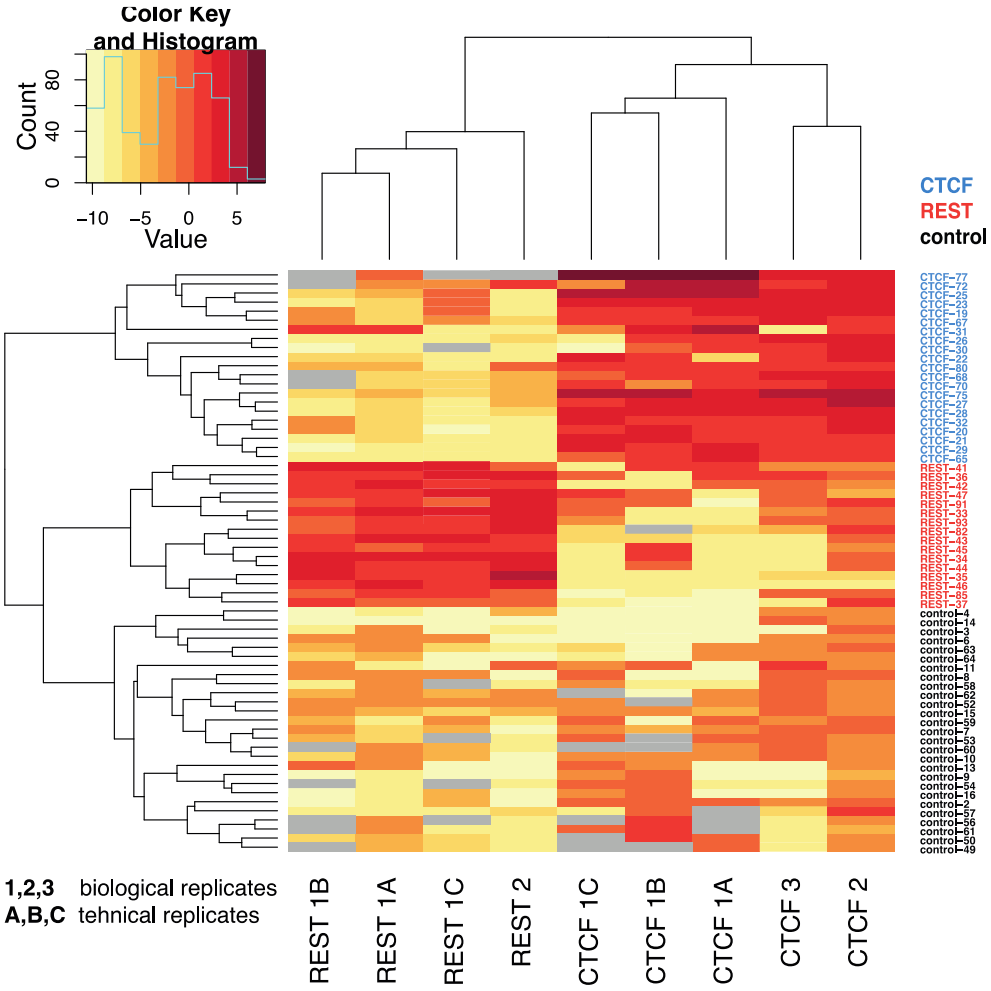


Figure 3.3 CTCF and REST library ChIP-seq enrichments. Enrichment heatmap showing ChIP-seq for the CTCF and REST insertion library. Each row represents one inserted fragment with a CTCF motif (blue), REST motif (red) or no motif (black). Color scheme of enrichment (log2; IP over input) is described in upper left corner. Clustering of samples shows robustness of the approach in enriching for binding events at the insertion site. Grey regions represent uncovered amplicons.

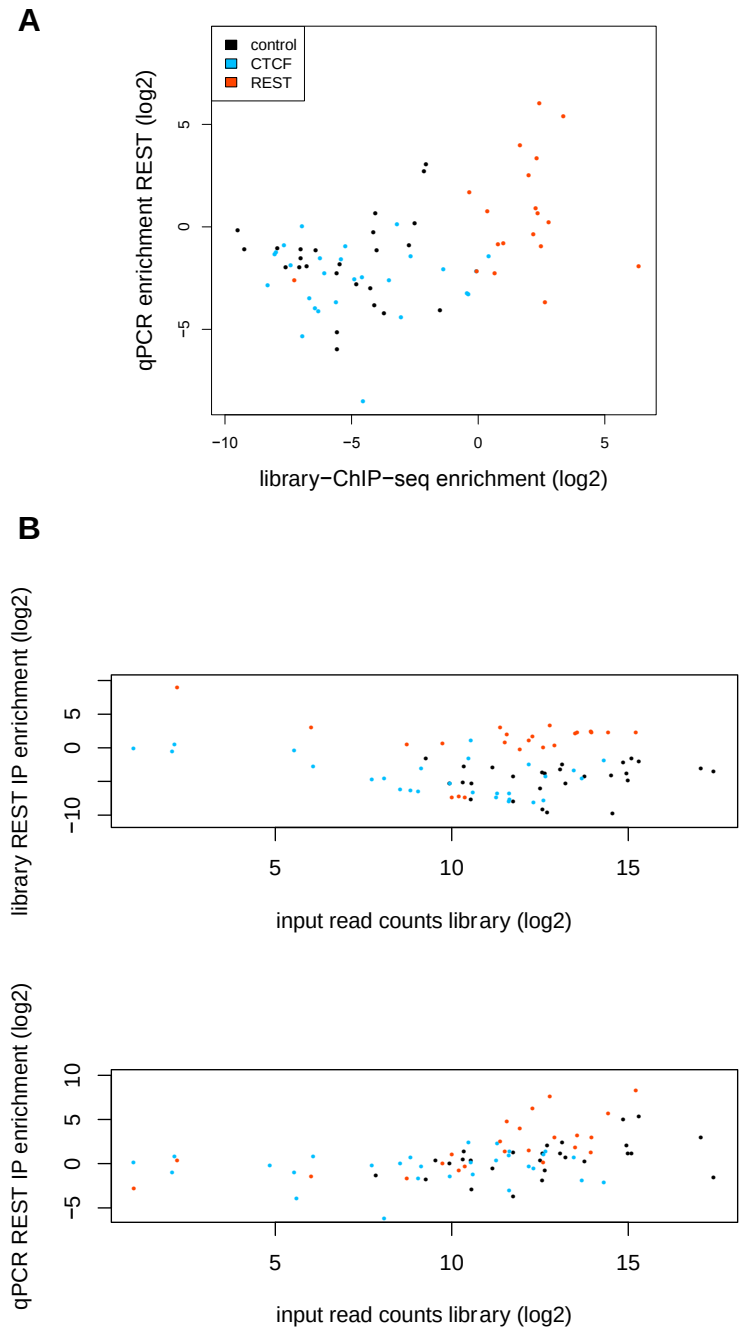


Figure 3.4 Quantitative PCR on individual fragments of the CTCF and REST library. A. Comparison of library ChIP-seq and quantitative PCR enrichments per fragment. The color code; CTCF motif (blue), REST motif (red) or no motif (black). B. Comparison of library ChIP-seq enrichments (upper panel) or quantitative PCR enrichments (lower panel) with the sequencing depth per fragment (input read count) in the CTCF and REST library.

3.1.2 CpG Density Drives PRC2 Recruitment in the Library of Endogenous Polycomb Sequences

To determine which endogenous Polycomb sequences reconstitute Polycomb binding when placed in an ectopic genomic context, we designed a library that contained hundreds of putative Polycomb targets as well as control regions that lack Polycomb enrichment. In the design of the library we included; (1) sequences with varying endogenous enrichment of PRC2 (Suz12), (2) both PRC2 targets and control regions with varying CpG density, (3) control sequences that are not bound by PRC2 in ES cells, but gain binding upon neuronal differentiation (Bibel et al., 2004). In mouse ES cells, H3K27me3 peaks are rather defined and narrow when compared to other tissues or organisms (*D. Melanogaster* embryos for example). Moreover, Suz12 and H3K27me3 peaks show a high overlap, making it easier to select regions both bound by PRC2 and enriched for the H3K27me3 mark. Nevertheless, we ensured that all selected Suz12-bound sequences show enrichment for the mark. The selected regions were used in the library ChIP-seq protocol to assess the recruitment of PRC2 and acquisition of the repressive histone mark H3K27me3, as well as the H3K4me3 histone mark. Surprisingly, we found that not all endogenous Polycomb recruiting sequences reconstitute Polycomb and H3K27 methylation when placed ectopically (Figure **3.5**). Moreover, some control regions that do not recruit Polycomb at the endogenous location were enriched for Polycomb binding at the ectopic location (Figure **3.5**). It appears that endogenous PRC2 enrichment is not predictive of the binding at the ectopic location. In all following library ChIP-seq experiments, Suz12 and H3K27me3 were performed in parallel and found to be highly correlated and interchangeable.

To show that observed library enrichments are PRC2-specific, we performed H3K4me3 ChIP on the Polycomb library. In ES cells, H3K4me3 is known to co-occupy PRC2 targets to create so-called bivalent domains (harboring both H3K4me3 and H3K27me3; Margueron & Reinberg, 2011). However, certain regions in the genome, namely regions of active

promoters, harbor only the H3K4me3 mark. We included those H3K4me3-exclusive regions in the design of the library. For these regions, we observed H3K4me3 specific enrichment in the Polycomb library-ChIP seq (Figure **3.5**, last two columns). Furthermore, as expected, library fragments found enriched for H3K27me3 were also enriched for H3K4me3, showing the ectopic fragments reconstitute the endogenous bivalent domain structure. When comparing Suz12 and H3K27me3 enrichments, we found them to be highly correlated (Figure **3.5**), resembling the endogenous situation. Taken together, it seems endogenous Polycomb enrichment is not indicative of binding at the ectopic site. Still, the bivalent structure seems preserved. We next aimed to determine sequence features that distinguish bound from the unbound fraction of fragments in the Polycomb library.

To determine if there are transcription factors that could account for the difference in the bound versus unbound fraction, we performed a motif search analysis (using the HOMER tool). We extracted sequences that are bound by PRC2 when inserted ectopically. Next, we asked if there are transcription factor motifs enriched in this set of sequences compared to sequences that are not bound in the library. The motif analysis did not reveal high enrichment of any specific known transcription factor motif. The highest score hits were two motifs associated with transcription factors TEAD1 and GMC1 (data not shown). However, the motifs for these two factors were just below the P-value threshold for being marked as enriched in the bound fraction and, moreover, both are reported to be transcriptional activators. Therefore, they are highly unlikely to be involved in the recruiting of PRC2. Taken together, we concluded that transcription factor motifs cannot discriminate between the bound and unbound fraction in the Polycomb library.

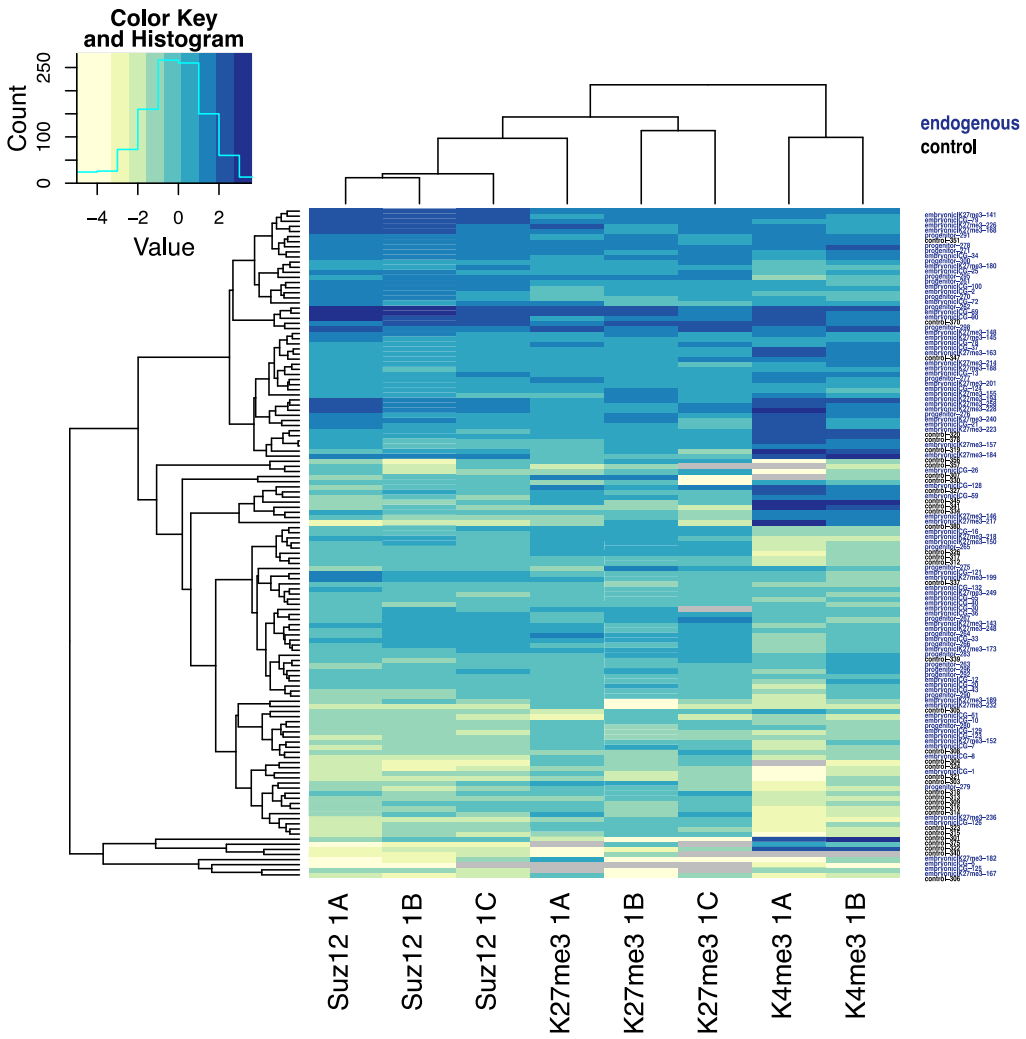


Figure 3.5 Enrichment heatmap showing Polycomb library ChIP-seq. Each lane represents one inserted fragment with an endogenous Polycomb enrichment (endogenous; blue) of without (control; black). Each column represents either Suz12, H3K27me3 or H3K4me3 ChIP on the Polycomb library in three (or two) biological replicated (1A, 1B, 1C). Color scheme of enrichment (log2; IP over input) is described in upper left corner. Grey represents regions that were not recovered in the sequencing run.

In light of these initial results, we hypothesized that increasing density of the CpG motifs alone could account for Polycomb recruitment at the ectopic location, regardless of the endogenous levels of Polycomb binding. To test this, we compared CpG density with the Suz12 enrichment in the library. Indeed, we observed that the enrichment of Suz12 as well as H3K27me3,

scaled with CpG density (observed over expected ratio; Figure **3.6**). It is known that PCR could preferentially amplify sequences with low GC content. To minimize the potential influence of such bias, we (1) normalized the IP fraction to input, since the input fraction should be subjected to a similar GC bias, and we (2) calculated observed over expected ratio when quantifying CpG content, which is normalized to GC content. However, to further ensure the observed correlation is not driven by the PCR bias, we correlated library enrichment with the inverted motif, GpC, which contains the same GC ratio as the CpG motif. Indeed, when inverting the motif, correlation of GpC content and Polycomb library enrichment is highly reduced (Figure **3.6**). Therefore, we concluded that CpG motifs might be the sole signal for Polycomb binding.

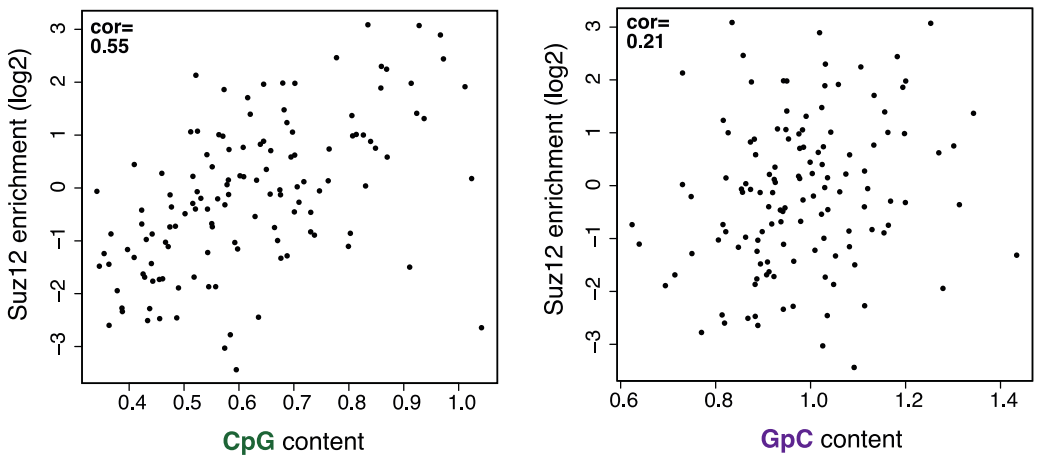


Figure 3.6 Comparison of CpG and GpC density to ChIP enrichment of fragments in the Polycomb library. Scatter plot of Suz12 enrichment at the insertion library with either CpG density (left panel) or GpC density (right panel). Each dot represents a distinct fragment in the library. Spearmen correlation is indicated in the upper left corner.

The fragments in the Polycomb library are endogenous sequences rich in transcription-factor motifs. Since transcription factors might obstruct or further help recruit PRC2 to sequences in the Poylcomb library, we aimed to further test the ability of isolated CpG motifs to autonomously drive Polycomb recruitment. Before testing the contribution of CpGs to PRC2

recruitment, we sought to determine the DNA methylation state of the Polycomb library. This is necessary to ensure DNA methylation does not influence PRC2 recruitment at the ectopic location because DNA methylation and Polycomb are known to influence each other (Di Croce & Helin, 2013; Margueron & Reinberg, 2011; Jeffrey a Simon & Kingston, 2009).

3.1.3 DNA Methylation Status of the Polycomb Library

DNA methylation and H3K27me3 are mutually exclusive since genome-wide mapping of both marks shows no overlap between the two (Brinkman et al., 2012; Meissner et al., 2008; Mohn et al., 2008; Reddington et al., 2013). Moreover, it is thought that DNA methylation blocks PRC2 recruitment to methylated CpG islands. Therefore, we wanted to ensure the lack of PRC2 enrichment at the fraction of sequences in the Polycomb library, is not driven by high DNA methylation levels. To investigate if binding of PRC2 at these fragments is blocked by DNA methylation, we performed amplicon bisulfate sequencing on the Poylcomb library (library BIS-seq). In the library BIS-seq protocol, genomic DNA is subjected to bisulfite conversion, subsequent PCR amplification and amplicon sequencing on the Illumina Miseq platform. Bisulfite treatment of genomic DNA and subsequent PCR step selectively converts unmethylated cytosines to thymines. By comparing the unconverted sequence to the converted sample, DNA methylation is presented as a percentage of converted (unmethylated) to unconverted (methylated) cytosines at a given position. Following the BIS-seq protocol on the Polycomb library, we compared the DNA methylation levels of fragments in the Polycomb library to their Suz12 or H3K27me3 enrichment. We observed low DNA methylation levels at majority of fragments that are not enriched in the Polycomb library (Figure **3.7**). Furthermore, fragments with high PRC2 enrichment also show low methylation levels, resembling the endogenous situation (Figure **3.7**). Taken together, we can confirm that the low PRC2 enrichment of putative

recruiters in the Polycomb library is not a consequence of the PRC2 complex being blocked by DNA methylation.

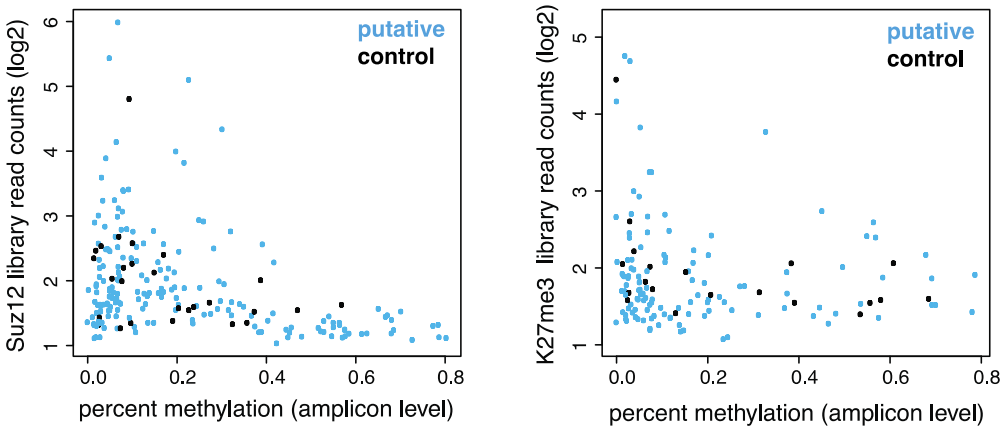


Figure 3.7 Comparison of DNA methylation levels to ChIP enrichment of fragments in the Polycomb library. Scatter plot of Suz12 (left panel) and H3K27me3 (right panel) enrichment at the insertion library with either DNA methylation levels calculated on the per fragment basis.

3.1.4 CpG Density Scales with PRC2 Recruitment in the Library of *E. coli* Sequences

Genome-wide studies unveiled that the vast majority of Polycomb targets in mammalian genomes are unmethylated CpG islands (unpublished data from the group; Ku et al., 2008). Furthermore, CpG density seems to be the best predictor of PRC2 enrichment when placed ectopically, as shown by previous experiments (Figure **3.5**). Moreover, a previous study in the group showed that exchanging the sequence of a Polycomb recruiter with random bacterial DNA but keeping only the number and position of CG dinucleotides still reconstitutes Polycomb binding and H3K27 methylation (Jermann et al., 2014). These results indicate that CpGs are sufficient to recruit PRC2 in an isolated context without mammalian transcription factor motifs. Motivated by these findings, we designed a library of fragments with varying CpG density in a neutral *E. coli* sequence (Krebs et al., 2014). In other words, we created a library that contained fragments of various CpG densities while lacking

mammalian transcription factor motifs. By assaying Polycomb recruitment to these sequences, we aimed to determine if there is a certain CpG density that autonomously drives Polycomb recruitment.

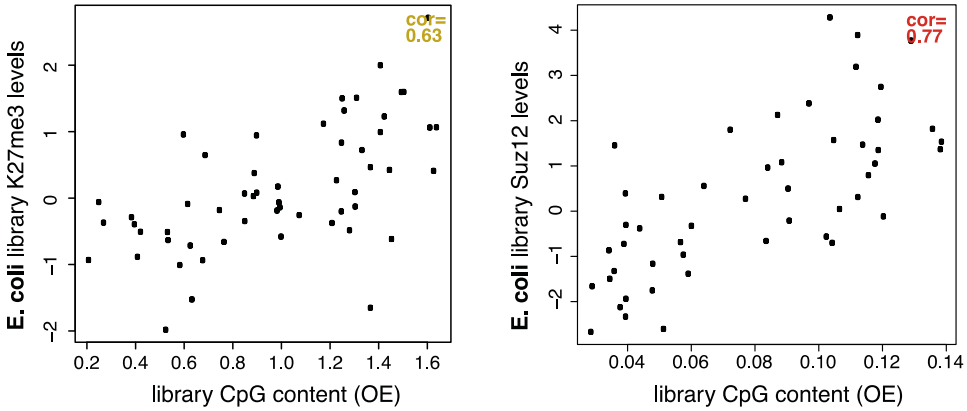


Figure 3.8 Comparison of E. coli library H3K27me3 and Suz12 ChIP enrichments with CpG density. Scatter plot of CpG density and either H3K27me3 ChIP enrichment (left panel) or Suz12 enrichment (right panel). Spearmen correlation is indicated in the upper right corner.

The *E. coli* library was inserted into mouse ES cells as described before (library ChIP-seq; Chapter 3.1.1). We determined H3K27me3 enrichments in the *E. coli* library by library ChIP-seq and observed that H3K27me3 enrichments scale with CpG density (with correlation of 0.77 for Suz12 and 0.63 for H3K27me3). Furthermore, CpG density seems to be a better predictor of PRC2 binding in the isolated context within the *E. coli* library when compared to putative fragments in the Polycomb library (correlation of Polycomb library enrichment and CpG density was 0.55 for Suz12 and 0.43 for H3K27me3). This indicates that CpG content is directly proportional to recruitment of PRC2 and is potentially mediated by CpG-binding proteins. What further supports this model is the observation that there is not a critical CpG density required for effective recruitment of PRC2. The recruitment seems to be continuous with increasing CpG content. It remains to be determined which transcription factors, presumably CpG-binding CXXC-domain proteins, are involved in the CpG-mediated PRC2 recruitment.

3.1.5 Polycomb Repressive Complex 2 is not Required for Transcriptional Silencing in Steady State Cellular Systems

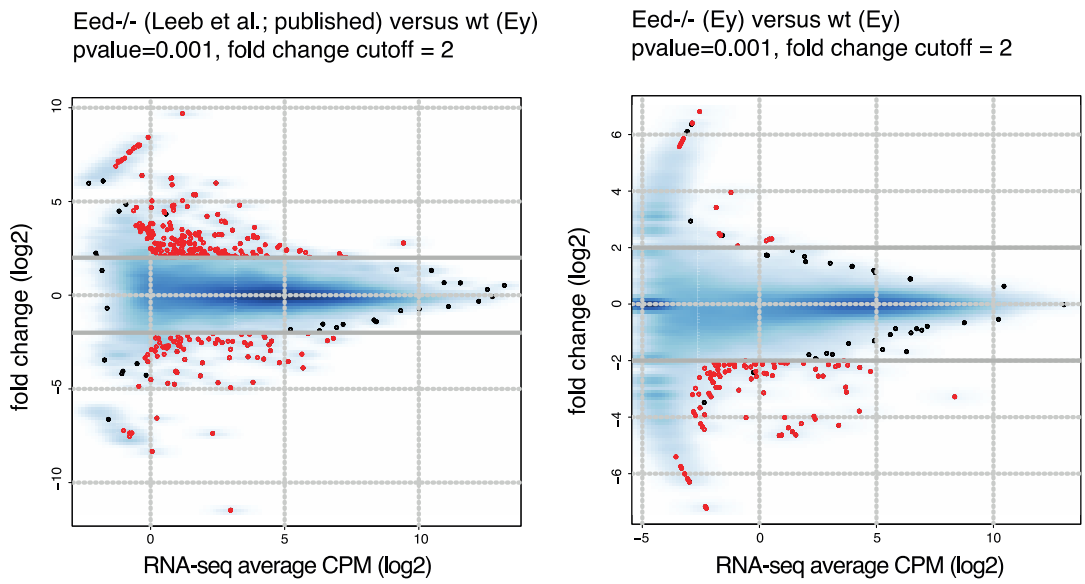
In order to investigate the impact of Polycomb recruitment on transcriptional repression, we decided to dissect promoters of genes upregulated in a Polycomb mutant. By comparing sequences that autonomously recruit PRC2 with Polycomb-responsive genes, we aimed to further define sequence features of Polycomb recruitment and link recruitment with transcriptional silencing. The first step in dissecting Polycomb-mediated repression was to confirm Polycomb-responsive genes in mouse ES cells. Several studies identified a set of genes upregulated in different mutants of PRC2 core components (Boyer et al., 2006; Leeb et al., 2010; Figure 3.9). However, the genes identified are not consistent between different studies, possibly due to technical differences between studies. Different genetic backgrounds or clonal differences might be another cause for observed discrepancies between these studies. In fact, genetic deletion (knockout; ko) clones in these studies have been derived after several rounds of genomic targeting. The identified set of differentially regulated genes might therefore reflect those differences and not the PRC2 (Eed) deletion phenotype. Indeed, a recent study showed that simply culturing ES cells over a long period of time will upregulate certain PRC2 targets (Riising et al., 2014). The origin for this deregulation is not known, but a certain proportion of cells undergoing spontaneous differentiation and loss of pluripotency might be the cause.

To test if PRC2 deletion in our ESC system will exhibit transcriptional changes comparable to the ones previously reported, we aimed to delete *Eed*, the gene encoding for the core PRC2 component that results in global loss of H3K27me3 levels. Deletion was generated using Cas9/CRISPR system in an wt-isogenic mouse embryonic stem cell line. We thereby excluded the influence of different genetic backgrounds. The ko strategy involved deleting the entire 1st exon of *Eed*. As CRISPR/Cas9 was a rather new method at the time, we investigated CRISPR deletion efficiency. We

observed 40% heterozygous deletions, 21% homozygous deletions, 100% allele mutation rate (n=48, deletion size=750bp). The high efficiency of CRISPR/Cas9 editing made it feasible to generate and characterize the *Eed ko* line in a short period of time.

The deletion of Eed resulted in a global loss of H3K27me3 as determined by western blot (data not shown). Therefore, we performed total RNA-seq to identify transcriptional changes. Unexpectedly, we observed no significant upregulation of gene expression levels in *Eed* ko (adjusted p-value cut-off 0.01, linear fold change cut-off 4; Figure **3.9**). These results indicate that recruitment of PRC2 and subsequent H3K27 methylation might not be required for gene repression of most of the target genes in mouse ES cells. As we observed no transcriptional response to *Eed* deletion, we were limited in ability to study Polycomb-mediated repression in mouse ES cells and its link to Polycomb recruitment. We therefore aimed to investigate the transcriptional response to loss of PRC2 in other cellular systems.

To ask if the lack of transcriptional response upon H3K27me3 depletion is specific to pluripotent stem cell stage, we inhibited PRC2 activity in mouse erythroleukemic cells (MEL cells). MEL cells are a representative system for highly differentiated cells. We used an Ezh2 inhibitor GSK343 recently developed by GlaxoSmithKline to inhibit PRC2 activity. GSK343 was shown to selectively target Ezh2 and reduce H3K27me3 levels in several cell types (Verma et al., 2012). Upon inhibition of PRC2 activity and subsequent depletion of genomic H3K27me3 mark, we observed a depletion of H3K27me3 to the *Eed ko* levels. However, using a total RNA-seq approach, we did not observe any significant changes of transcriptional activity in GSK343-treated MEL cells (adjusted p-value cut-off 0.01, linear fold change cut-off 4; data not shown). Therefore, we concluded that PRC2 recruitment is not required for repression once a stable stage is reached, such as in stable cell-line models. Rather, we hypothesized it might have a role in directly repressing target genes during cell-fate transitions. Moving to a differentiation system will enable us to further elucidate the role of PRC2 in transcriptional repression.



top upregulated genes	fold change
Zkscan4	77.03702044
Zic1	71.24205036
Tex13	57.27644318
Hoxa10	46.07524717
Hoxd10	45.96710808
Hoxd13	27.63203672
Hoxd11	25.98577173
Hoxa11	23.75270641
Nfia	18.40058846

top upregulated genes	fold change
Pitx2	4.307091469
Sfta2	4.056142321
Meis2	3.898666002
Gm525	3.735650367
Rab25	3.684485912
Camk2d	3.597013717
Chst5	3.574737676
Cdkn2a	3.499541593
Spink10	3.479301432

Figure 3.9 Transcriptional response upon deletion of *Eed* in mouse ESC. RNA-seq read count plotted against fold change between *Eed* ko and wild type. Read counts are shown as count per million (CPM). Fold change was determined using three biological replicates and calculated on the *per gene* basis. Left panel shows differentially expressed genes (red) in published *Eed* ko dataset, right panel shows our own CRISPR/Cas9-generated *Eed* ko data. Tables below represent top upregulated genes in both samples. Highlighted in blue are *Hox* genes known to be Polycomb targets.

3.1.6 Activity of Polycomb Repressive Complex 2 is Necessary for Silencing of Developmental Genes During *In Vitro* Differentiation

So far, we observed no transcriptional changes upon depleting global H3K27me3 levels in two different mouse cell-lines (ES cells, MEL cells; Chapter 3.1.5). This led us to hypothesize that PRC2 has a more pronounce

role in gene silencing during cell-fate transition. Therefore, we wanted to test if Polycomb recruitment is required for transcriptional repression during differentiation. To this aim, we differentiate *Eed* ko ES cells into neuronal progenitors. We used an established protocol for differentiation that generates pure population of neuronal progenitors and terminal neurons (Bibel et al., 2004). Initially, we intended to differentiate previously published *Eed* depleted cells (Leeb et al., 2010). However, we were unable to derive embryoid bodies, a first step in the neuronal differentiation protocol. Inability of *Eed* depleted ES cells to differentiate was also previously reported (Leeb et al., 2010). Nevertheless, we subjected the *Eed* ko generated with CRISPR/Cas9 to the neuronal differentiation protocol. Surprisingly, *Eed* ko cells were able to form neuronal progenitors and they appeared morphologically identical to wt cells.

Motivated by this observation, we determined the transcriptional response in *Eed* ko neuronal progenitors with RNA-seq. As PRC2 was shown to be essential for differentiation, we anticipated a high level of transcriptional misregulation during differentiation. Indeed, we observed 2719 genes deregulated in *Eed* ko neuronal progenitors when compared to wild type neuronal progenitors. This was despite using a rather stringent cut-off for characterizing a gene as deregulated (adjusted p-value cut-off 0.01, linear fold change cut-off; Figure **3.11**). The majority of deregulated genes were upregulated, which is consistent with the depletion of a repressor complex. Furthermore, 44% of upregulated genes have an enrichment of H3K27me3 at their promoter region in wt cells (within -2 kb to +1.5 kb distance from TSS), suggesting this is a direct consequence of PRC2 disruption. The relatively high transcriptional changes we observed in neuronal progenitors of *Eed* ko cells indicate that PRC2 is indeed important for the switch in transcriptional states during cell-fate transition rather than the maintenance of steady state repression.

To further confirm the hypothesis, we took advantage of the available GSK343 inhibitor to inhibit Ezh2 activity during differentiation (Figure **3.10**). Unlike the *Eed* deletion, inhibiting Ezh2 activity will show less secondary

effects as the integrity of the complex is preserved. We inhibited Ezh2 throughout the differentiation of wild type ES cells to neuronal progenitors as well as during only the last 48 hours of the differentiation protocol. If Ezh2 activity has a more pronounced role in cell-fate transition rather than cell state maintenance, we expected more transcriptional changes after continuous Ezh2 inhibition as compared to late stage inhibition. As expected, *Ezh2* inhibition throughout differentiation (starting 2 days prior to differentiation cue) shows a higher transcriptional response (110 genes deregulated) compared to almost non-responsive neuronal progenitors treated for 2 days (11 genes deregulated; Figure **3.11**). Additionally, the majority of the responding genes are Polycomb targets; 88% for the ten-day treatment and 100% for the two-day setting. Interestingly, *Ezh2* inhibition throughout differentiation compared to *Eed* deletion showed a more subtle effect on de-repression, more than 20-fold difference in the number of deregulated genes. This is potentially due to several reasons: (1) Ezh1, the Ezh2 homolog might compensate for the loss of Ezh2 activity, (2) the massive transcriptional response in *Eed* ko neuronal progenitors might come from secondary effects of loss of *Eed* or PRC2 complex disruption, (3) inhibition of Ezh2 might not be as efficient at different stages of cell cycle or throughout differentiation when fluctuations of *Ezh2* expression occur. Taken together, we showed that H3K27me3 depletion has a role in transcriptional silencing during differentiation, but not in the maintenance of gene repression in steady stage cell-line systems.

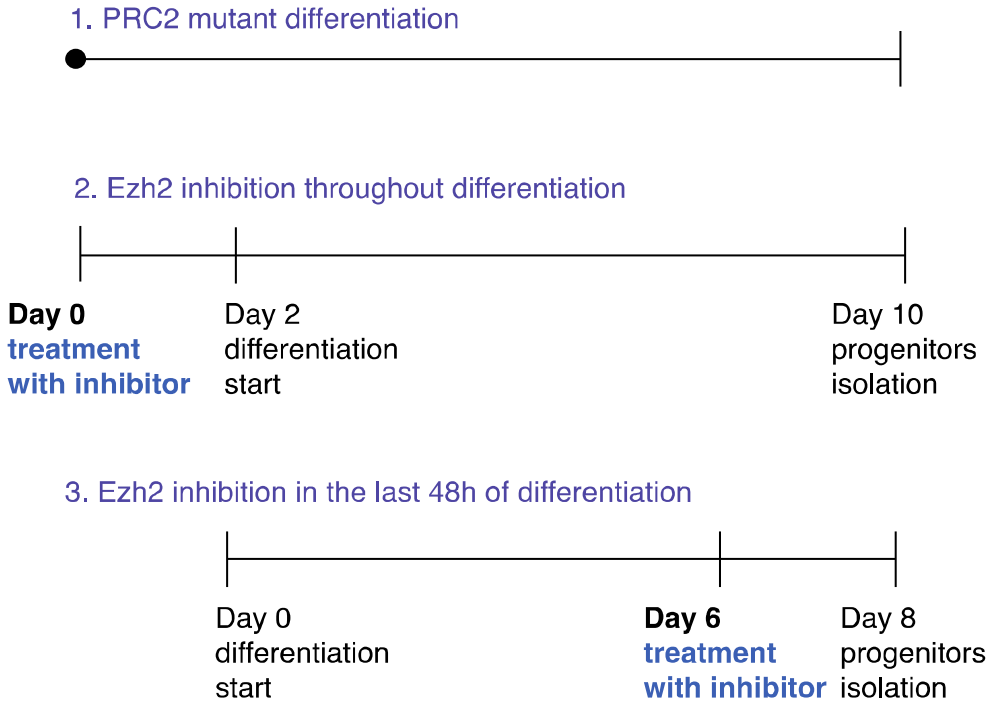


Figure 3.10 Experimental workflow summary for determining transcriptional response to *Eed* depletion and *Ezh2* inhibition during differentiation to neuronal progenitors. Three different systems were used to determine changes upon loss of H3K27me3; (1) constitutive Eed ko, (2) *Ezh2* inhibition during differentiation to neuronal progenitors (starting 2 days prior to differentiation), (3) *Ezh2* inhibition during the last 2 days of differentiation.

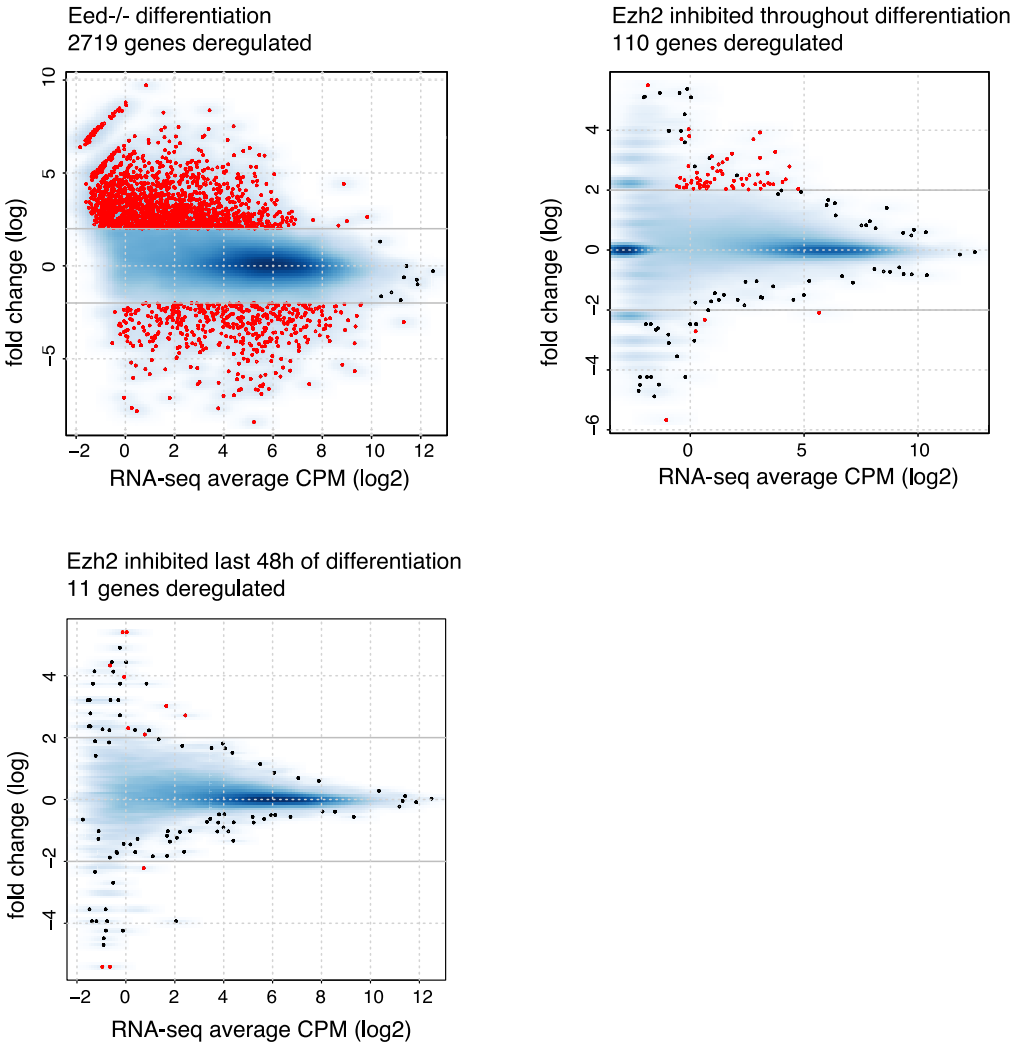


Figure 3.11 Transcriptional response during neuronal differentiation upon *Eed* depletion and *Ezh2* inhibition. RNA-seq read count plotted against fold change between *Eed* ko and wild type neuronal progenitors. Read count is shown as count per million (CPM). Fold change was determined using three biological replicates and calculated on the *per gene* basis. Upper left panel shows deregulated genes (red) in *Eed* ko neuronal progenitors dataset, upper right panel shows deregulated genes upon inhibition of *Ezh2* during the course of entire differentiation or (lower right panel) only the last 2 days of differentiation.

3.2 Function of Imitation Switch Complexes in Mouse Embryonic Stem Cells

In the eukaryotic genomes, nucleosomes represent a barrier for transcription factors and chromatin-binding complexes to access the DNA substrate. Chromatin remodelers play an essential role in regulating that accessibility by sliding, evicting or assembling nucleosomes (Clapier & Cairns, 2009). There are four main chromatin remodeling families: SWI/SNF, ISWI, CHD and INO80. While their ability to change chromatin structure has been demonstrated in *in vitro* biochemical assays (Clapier & Cairns, 2009), little is known about their *in vivo* function in mammals. It has been shown that the SWI/SNF family of remodelers are required for binding of certain transcription factors but not required for nucleosome positioning in mouse ES cells (Barutcu et al., 2016; Ho et al., 2011). However, the *in vivo* role of other mammalian remodeler complexes is not fully understood. In this project, we aimed to investigate how ISWI remodeling complexes change nucleosome positioning and accessibility for transcription-factor binding *in vivo*. The current model indicates that ISWI-family complexes induce chromatin compaction, organize nucleosomes in an array and maintain the equal linker lengths between nucleosomes (Clapier & Cairns, 2009; Saha, Wittmeyer, & Cairns, 2006). However, most of what we know about ISWI-mediated remodeling is derived from *in vitro* data. The little *in vivo* data available is restricted to yeast models. The reported cellular lethality of most known remodeler ATPase mutants and the lack of location maps impedes studying of these complexes in mammalian systems (Ho & Crabtree, 2010; Stopka & Skoultchi, 2003). To describe the role of mammalian ISWI complexes *in vivo*, we aimed to create a constitutive/conditional Snf2h deletion mouse ES cell line and track transcriptional and chromatin changes.

3.2.1 Mouse Embryonic Stem Cells are Viable and Exhibit Growth Phenotype Upon Snf2h Deletion

Snf2h Deletion Phenotype

To ask if ISWI complexes have a role in regulating nucleosome positioning and gene expression *in vivo*, we deleted Snf2h, the ATPase found in ISWI complexes. The Snf2h deletion (Snf2h ko) was created in a mouse ES cell line using CRISPR/Cas9, which introduced a single point mutation in the exon 6 of the *SMARCA5* gene that codes for Snf2h. The frameshift caused by a single-base deletion was sufficient to completely deplete the protein levels as measured by whole-cell lysate western blots (Figure **3.14**). The *Snf2h ko* cells appeared viable and showed unchanged morphology compared to wt ES cells. This was unexpected because the *Snf2h ko* mice were reported to die in the pre-implantation stage due to growth arrest and cell death of both the inner cell mass and the trophectoderm (Stopka & Skoultchi, 2003). Apoptosis of cells in the inner cell mass indicates that the embryonic stem cells are not viable and *Snf2h ko* was therefore considered to be lethal in cellular systems. In fact, this is the first reported viable ko ES cell line of a chromatin remodeler ATPase subunit. Therefore, we wanted to further characterize the *Snf2h ko* line.

One striking feature of *Snf2h ko* cells was their growth phenotype. Namely, the replication rate was determined to be 2.14 times slower than wt cells (as measured by cell counting 48h post seeding of the equivalent cell number). The observed growth phenotype might be caused by either (1) the reduction of the proportion of cells in one of the cell cycle phases, by (2) proportionally slower progression throughout the whole cell cycle or (3) a combination of the two.

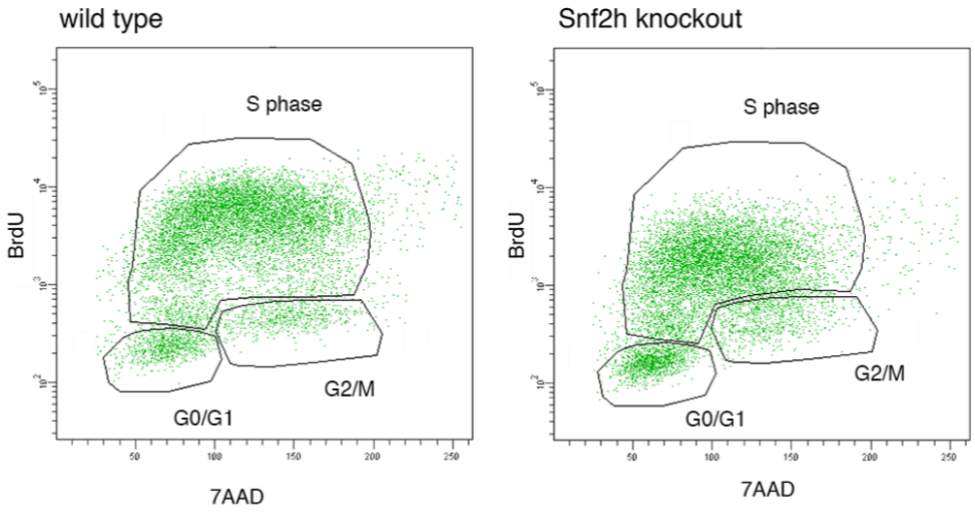


Figure 3.12 Cell cycle analysis in wt and *Snf2h ko* cells using BrdU and 7AAD labelling. BrdU signal is plotted against 7AAD (fluorescent intercalating DNA label) intensity to identify cell cycle stages. Scatter of the signal corresponding to a certain cell cycle stage was determined manually.

Changes in the proportion of cells in a certain cell cycle phase would confound analysis in further experiments. To test if *Snf2h ko* cells are delayed in progression through one of the cell cycle phases, we analyzed the cell cycle profile in *Snf2h ko* cells using a BrdU incorporation assay. In the BrdU labelling assay, the thymidine analog BrdU is incorporated into newly synthesized DNA. Cells entering and progressing through the S-phase of the cell cycle with a faster rate will proportionally incorporate more BrdU. Following 1h of labelling and detection of BrdU, we observed a decrease in overall BrdU levels in *Snf2h ko* cells compared to wild type (Figure **3.12**). This confirms our previous observation that *Snf2h ko* cells exhibit a slower division rate. Next, we asked if the growth phenotype is caused by a stall in the progression through a certain cell cycle phase or by the overall slower progression. We used BrdU and 7AAD staining to determine the number of cells in each cell cycle phase. Compared to wt, we observed only a slight decrease in the number of cells in S phase for the *Snf2h ko* (72.8% in S phase for *Snf2h ko* compared to 81.1% of wild type

cells). Co-occurring with a decrease in number of cells in the S phase, we observed a slight increase in the number of *Snf2h ko* cells in G1 and G2 phase (from wt 8.3% to ko 13.9% for G1; 6.1% to 8.6% for G2). However, the observed slight change in the cell cycle profile cannot account for the difference in the observed growth rate. Therefore, we concluded that the growth phenotype in *Snf2h ko* cells comes from an overall slower progression through the cell cycle and not from a major shift in the proportion of cells in a certain cell cycle phase.

To confirm this is a stable growth phenotype and not accumulation of secondary effects such as increase in DNA damage, we cultured *Snf2h ko* cells over a prolonged period of time. We observed the same growth rate and phenotype after 29 passages (counting from the first clonal expansion), confirming this is a stable phenotype. We next wanted to check the pluripotency potential of *Snf2h ko* cells. To do this, we performed the neuronal differentiation protocol previously mentioned (Bibel et al., 2004). However, *Snf2h ko* cells failed to form embryoid bodies and undergo differentiation, indicating *Snf2h* is required for this process. As *Snf2h ko* do not seem to be pluripotent, we asked if they still show ES cell characteristics. We observed that the *Snf2h ko* cells were able to form ES colonies, which is a feature of ES cells. So far, we demonstrated that *Snf2h ko* cells stably proliferate yet fail to differentiate into neuronal progenitors.

Transcriptional Response Upon Snf2h Deletion

To ask if the Snf2h ko cells still express pluripotency markers found in wt ES cells, we performed a transcriptional analysis using RNA-seq. We found the core pluripotency factors (*Nanog, Sox2, Oct4*) have unchanged expression levels in the *Snf2h ko* (Figure **3.13**), suggesting the *Snf2h ko* cells continue to show ES cell characteristics.

Furthermore, transcriptome analysis revealed substantial changes in gene expression. 532 genes were deregulated when using p-value cut-off of 0.001 and linear fold-change cut-off of 4 (Figure **3.13**). In the gene ontology analysis, deregulated genes were significantly enriched for the following

cellular processes; (a) positive regulation of cell proliferation, (b) skeletal development, (c) positive regulation of signal transduction, (d) organ morphogenesis, (e) tissue development (in the order of increasing p-value). Most of these processes are involved in tissue development. In fact, it has been shown that SWI/SNF complexes are involved in development and morphogenesis of various different tissues and organs (Ho & Crabtree, 2010). These results indicate that ISWI complexes might have a similar function. We next wanted to ask if (1) the observed cell cycle and transcriptional changes are reversible and dependent on the Snf2h ATPase activity (Chapter 3.2.2) and (2) if the observed transcriptional changes are linked to changes in nucleosome positioning (Chapter 3.2.4).

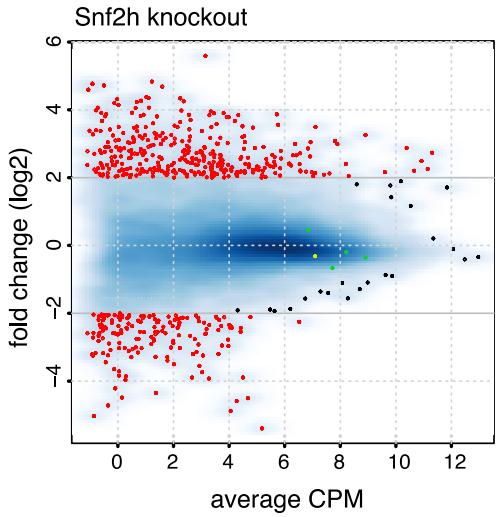


Figure 3.13 Transcriptional response upon Snf2h ko in mouse ESC. RNA-seq read count plotted against fold change between Snf2h ko and wild type. Read count is shown as count per million (CPM). Fold change was determined using three biological replicates and calculated on *per gene* basis. Red dots indicate differentially regulated gens (using linear fold change cut-off of 4, and p-value cut-off of 0.001), green dots indicate unchanged expression of pluripotency markers (*Nanog*, *Sox2* and *Oct4*). The yellow dot indicates unchanged expression of *CTCF*.

3.2.2 Lack of ATPase Activity Drives the Snf2h Phenotype

To further confirm the observed phenotype is caused by loss of Snf2h, we investigated if the observed phenotype is reversible. We addressed this question by stably re-expressing V5-tagged Snf2h into the ko background (Snf2h wt-ab). Upon Snf2h re-expression, we observed reversal of the growth phenotype to the wild type levels (Figure **3.15**). Furthermore, levels of incorporated BrdU in the *Snf2h wt-ab* were comparable to wt (Figure **3.15**). The number of cells in the S phase reverted back from ko 72.8% value to 80.2%, which very closely resembles wt levels (81.1%). Taken together, V5-Snf2h re-expression seems to rescue the growth phenotype. However, when determining the expression levels of V5-Snf2h protein in the *Snf2h wt-ab* cell, we observed it is not expressed to the wild type levels (as measured by western blot; Figure **3.14**). This could be due to inability of the ectopic promoter (synthetic CAG promoter) to drive V5-Snf2h expression to the especially high endogenous Snf2h levels. We observed that all re-expression clones express V5-Snf2h to the same level (data not shown). This is a good indication that the expression reached a plateau since we would expect position effects on expression levels when performing random genomic insertions. As we observed reversal of the growth phenotype in the *Snf2h wt-ab*, we nevertheless investigated if the lower expression levels are sufficient to rescue the transcriptional phenotype as well. We performed qPCR on four selected upregulated genes in the *Snf2h ko*. Interestingly, we observed reversal to wild type levels for all four genes (Figure **3.16**). To have a comprehensive overview of the degree of the rescue, we performed RNA-seq of *Snf2h wt-ab* cells. RNA-seq analysis showed majority of deregulated genes return to the range of their wild type levels, with exception of a few genes. Overall, we observed rescue of the phenotype when re-expressing V5-Snf2h in the ko background, further confirming the observed phenotype is Snf2h-dependent. Whether the low expression levels of the V5-Snf2h re-expression proteins is the cause of the incomplete transcriptional rescue is yet to be determined. We will generate Snf2h re-

expression lines that express Snf2h to wt levels and it will be further tested if this is sufficient to completely rescue the transcriptional phenotype. The V5 tag we introduced in the re-expression Snf2h protein might be effecting its expression levels. To test if this is the case, we will re-express a tag-free Snf2h protein.

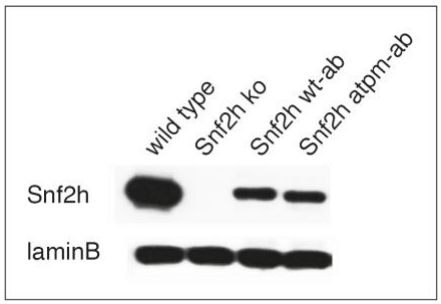


Figure 3.14 Expression levels of Snf2h in wild type, *Snf2h ko, Snf2h wt-ab and Snf2h atpm-ab* lines. *Snf2h wt-ab* and *atpm-ab* re-expression lines express V5-tagged version of the protein to a lower degree compared to wt.

Finally, to ask if the lack of Snf2h ATPase activity is driving observed transcriptional and cell cycle phenotypes, we re-expressed an ATPase mutant version of Snf2h into the *Snf2h ko* cells (Snf2h ATPase mutant add-back; Snf2h atpm-ab). We introduced a single point mutation in the ATP-binding domain shown to completely abolish ATPase activity of the protein, keeping the integrity of Snf2h-containitng complexes intact (Hakimi et al., 2002; Khavari, Peterson, Tamkun, Mendel, & Crabtree, 1993). We observed the *Snf2h atpm-ab* cells exhibit growth rate of the *Snf2h ko* cells (Figure **3.15**). Furthermore, in contrast to *Snf2h wt-ab* cells, *Snf2h atpm-ab* cells fail to rescue the transcriptional response as measured by qPCR (Figure **3.16**). The inability of the ATPase mutant to rescue the ko phenotype indicates that the impairment of ATP-dependent nucleosome remodeling drives the observed changes. To further link the transcriptional phenotype with nucleosome positioning, we aimed to describe changes in nucleosome positioning in *Snf2h ko* cells.

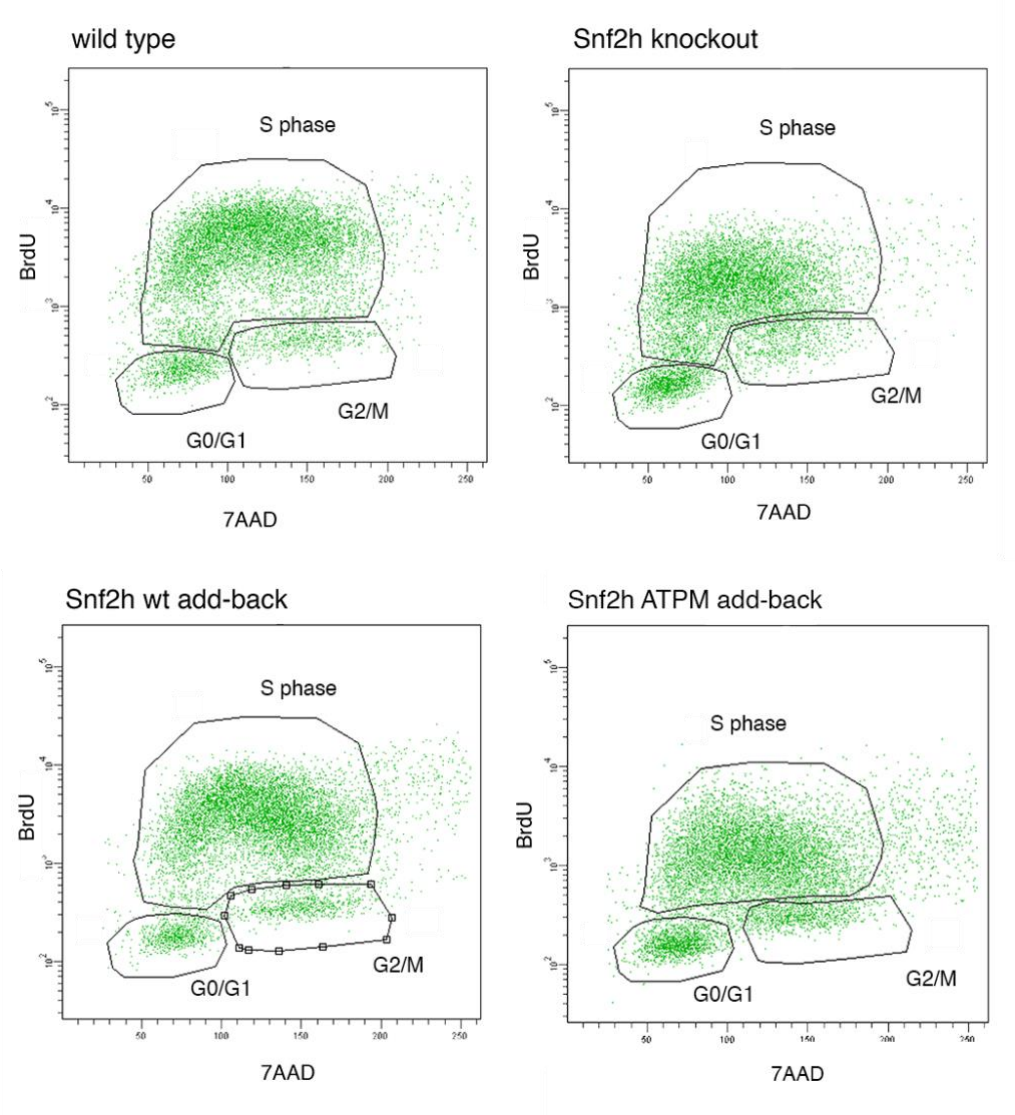


Figure 3.15 Cell cycle analysis in wild type, *Snf2h ko, Snf2h wt-ab and Snf2h atpm-ab* cells using BrdU and 7AAD labelling. BrdU signal is plotted against 7AAD intensity to identify cell cycle stages. Scatter of the signal corresponding to a certain cell cycle stage was determined manually and slightly differs between samples.

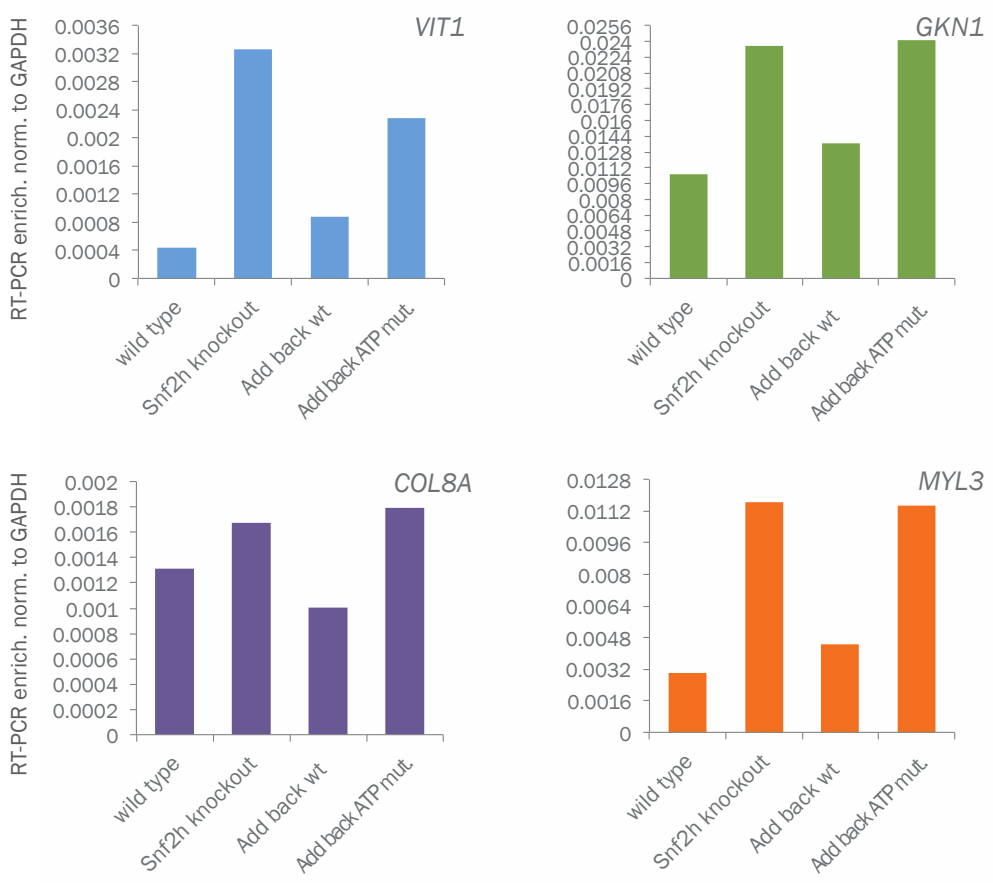


Figure 3.16 Transcriptional analysis of response to deletion and re-expression of Snf2h. RT-PCR was performed on four targets identified to be upregulated upon Snf2h depletion. Expression levels were normalized to *Gapdh* levels. In contract to Snf2h ATPase mutant re-expression (Add back ATP mut.), in wild type re-expression (Add back wt) we observe rescue of the transcriptional levels.

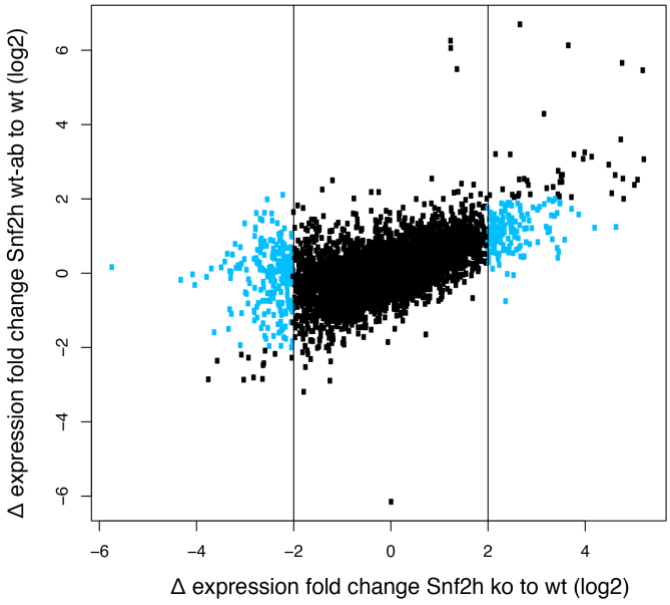


Figure 3.17 Rescue of the transcriptional phenotype upon re-expression of wt Snf2h protein into the ko background. The scatter plot indicates relationship between the transcriptional change of Snf2h depletion and re-expression of V5-Snf2h. Change in expression of ko to wild type (x axis) is plotted against change of expression in Snf2h wt-ab to wt (y axis). Inability of Snf2h wt-ab to rescue the phenotype would create a strong correlation on the diagonal, and the complete rescue would create a horizontal distribution around y=0. Points outside two vertical lines indicate differentially regulated genes in the ko. Blue dots indicate a subset of those genes rescued in the re-expression experiment. We observe the majority of differentially expressed gene are rescued. Small portion of genes are remain differentially expressed (upper left or bottom right corner of the plot).

3.2.3 Nucleosome Periodicity is Globally Reduced Upon Snf2h Depletion

To investigate if the changes in gene expression upon Snf2h knockdown are related to changes in nucleosome positioning, we performed MNase-seq on the Illumina Hiseq platform using single-end 51 cycles sequencing. In MNase-seq protocols, an endo-exonuclease from *S. aureus* digests linker DNA. Sequences protected by the digestion correspond to the nucleosomal DNA. Sequencing DNA fragments corresponding to the length of one nucleosome will, therefore, reveal the genomic location of nucleosomes genome-wide. To determine the best conditions for the MNase digestion, we performed low coverage (~ 30 mio. reads per sample) MNase-seq of eight conditions with varying either digestion time (30 min., 60 min.) or MNase concentration (2.5U, 5U, 7.5U, 10U).

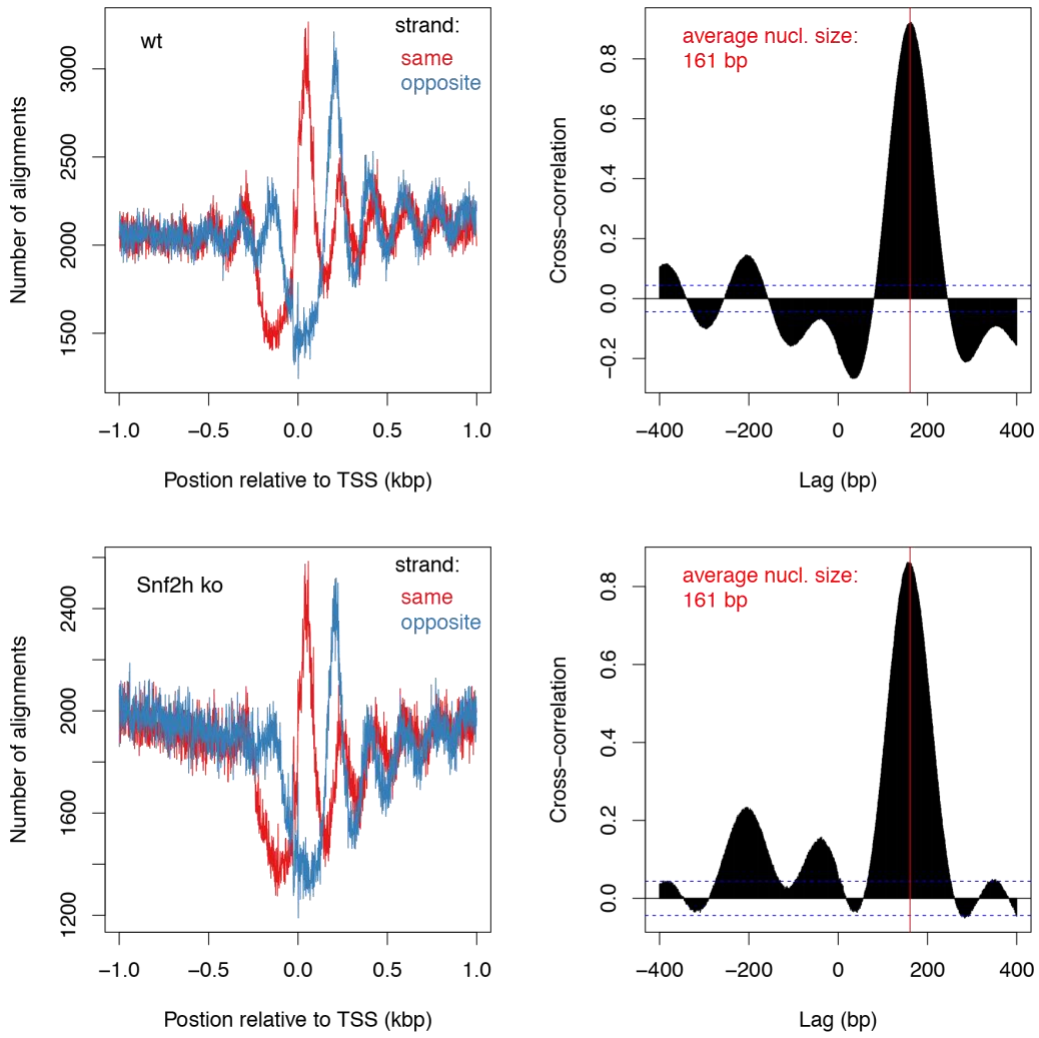


Figure 3.18 Nucleosome size inferred by fragment length in MNase-seq data for wt and *Snf2h ko* samples. Left panels shows alignment densities of MNase-seq fragments relative to TSS for both wt (upper) and *Snf2h ko* (lower panel). The reads are not shifted in position and indicate the MNase cut site, which on average corresponds to the edge of a nucleosome. Right panels indicate fragment size for MNase-seq data. Fragment size includes nucleosome length + part of linker. The global fragment size was defined using cross-correlation. Cross-correlation quantifies similarities of two series as a function of the displacement of one in relation to the other (displacement is indicated as lag in the graph). It was calculated between plus/minus strand alignment densities around TSS. Both wt and *Snf2h ko* cells show identical fragment size.

We determined quality of the data by analyzing the GC bias, which indicates over- and under-digestion of the sample (unpublished data from the group).

We identified the condition of 5U enzymatic treatment for 30 minutes as the one that lacks a GC bias. Following the MNase digestion of wt and *Snf2h ko* cells we sequenced the mononucleosomal fraction to a deeper coverage. The final sequencing depth was 600 mio. mappable reads per sample, amounting to ~ 30x in coverage for all nucleosomes in the mouse genome. Such high sequencing depth is essential to determine nucleosome positioning at the single locus level.

Snf2h was reported to have a role in maintaining linker length and organizing nucleosomal arrays *in vitro* (Clapier & Cairns, 2009). To investigate if Snf2h has a role in linker length maintenance *in vivo*, we determined MNase fragment length in both wt and ko conditions. MNase fragment length represents the DNA sequence that is wrapped around the nucleosome surface plus a part of the linker DNA. Changes in the fragment length will therefore indirectly reflect changes in the linker length. To determine the fragment length, we calculated cross-correlation between plus/minus strands of alignment densities around TSS. We used TSS as an anchor point as regions flanking TSS are rare genomic positions known to (1) exhibit very defined nucleosome positioning pattern and (2) occur frequently in the genome. Cross-correlation between plus/minus strands of MNase alignment densities was the highest for the length of 161 bp, indicating this is the fragment size in the MNase data. Unexpectedly, we observed identical fragment lengths of 161 bp in both wt and ko. This indicates that the fragment length is not altered in the absence of Snf2h (Figure **3.18**). However, fragment length is only an estimate and not a direct measure of linker length. Linker length would have to be further confirmed using other approaches.

We further wanted to investigate if nucleosome positioning is altered in the *Snf2h ko*. However, inferring nucleosome positioning on a single locus scale from MNase-seq data is rather challenging due to the nature of the assay. Assuming the nucleosome occupancy is unchanged, two samples with different nucleosome positioning pattern will still have the same number of alignments in a given genomic region. The difference in positioning will be

apparent only if enough fragments have a very defined location within the region. However, we know that the majority of the genome does not exhibit a very defined pattern of nucleosome positions, further complicating analysis of nucleosome positioning information from MNase-seq datasets. To circumvent this caveat, we developed an approach to infer nucleosome positions in such a complex dataset. As a first step, we applied the autocorrelation function (ACF) to our data. ACF is a computational approach to determine how data points in a given dataset are related. It is used to infer periodicity within the data. Most frequently, it is applied in a time series, which is in our case simulated by moving along the genomic position. For each genomic position, ACF generates a correlation of that genomic position with subsequent downstream positions.

The highest positive correlation is expected between mid-points of two nucleosomes, and the highest anti-correlation should occur between the nucleosome mid-point and the middle of the inter-nuleosomal distance. To confirm this is the case, we examined a single-locus example of two CTCF binding sites as they are known to position nucleosomes (Figure **3.19**). We observed that CTCF binding creates a clear pattern of alternating positive/negative correlations with a period of ~160 bp. Interestingly, these patterns are overall weaker in the *Snf2h ko*. This indicates nucleosome positioning is altered in *Snf2h ko*.

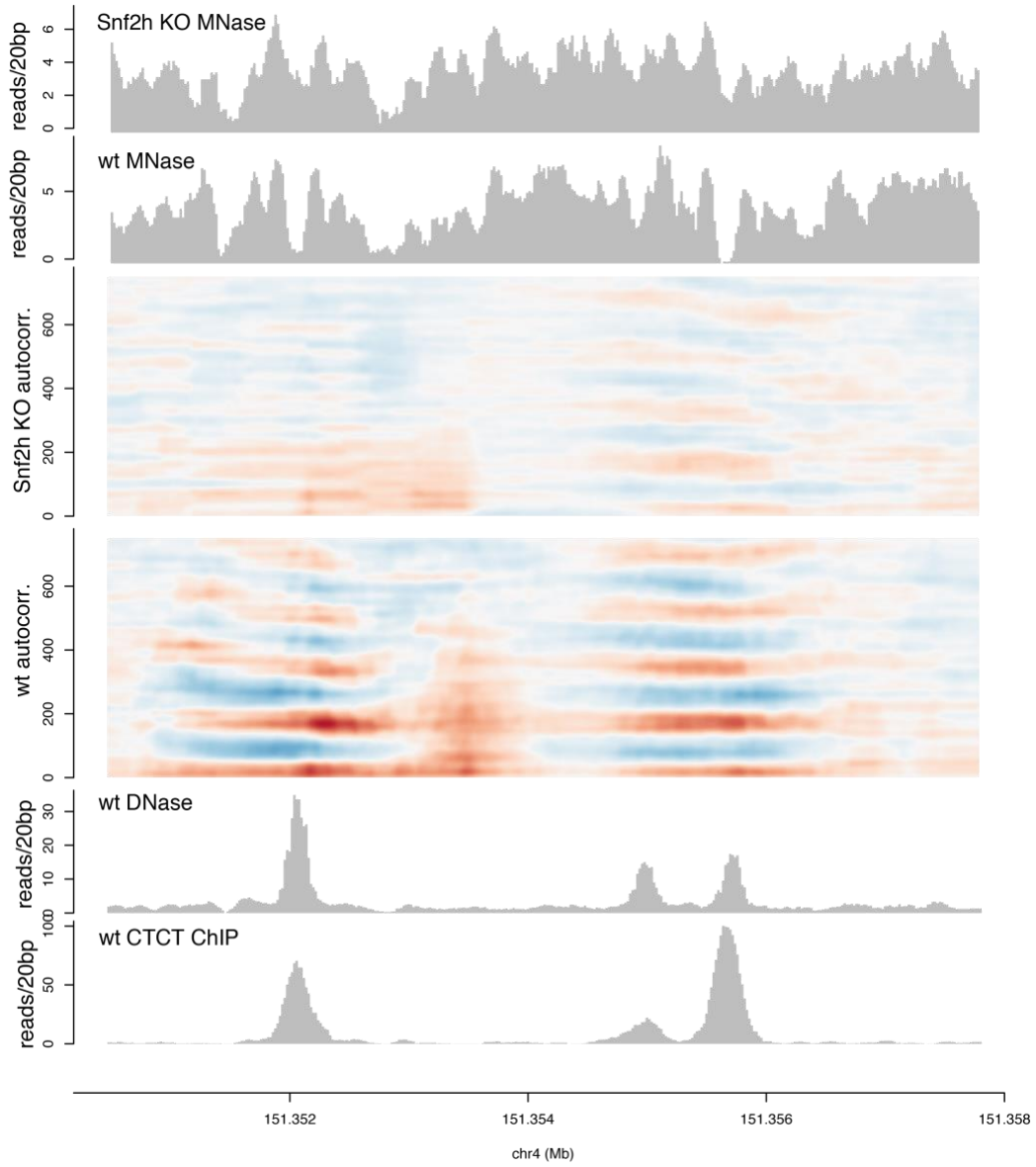


Figure 3.19 Autocorrelation function applied to the MNase-seq data. In the autocorrelation function (middle section), correlation coefficient is calculated between an anchor position (x axis) and the next downstream position. The correlation is then calculated between the anchor position and each subsequent downstream position up to 800 bp. Therefore, 0 to 800 on the y axis indicates autocorrelation coefficient of the anchor position on the x axis and its 800 bp neighboring downstream positions. Upper section of the plot represents alignments from the MNase-seq data (reads shifted by half the nucleosome size to depict nucleosome positions, rather than MNase-cut sites). Lower panel represents DNase-seq and CTCF ChIP-seq data for the given genomic location.

However, the autocorrelation approach has some drawbacks; periodic signal with other periods also creates a signal, some of which are artifacts (e.g. non-mappable regions). It therefore seemed necessary to combine this signal with a filter on the period expected for nucleosomes (~161 bp). To this aim, we applied short-time Fourier transformation on a sliding window of five times this length (805 bp).

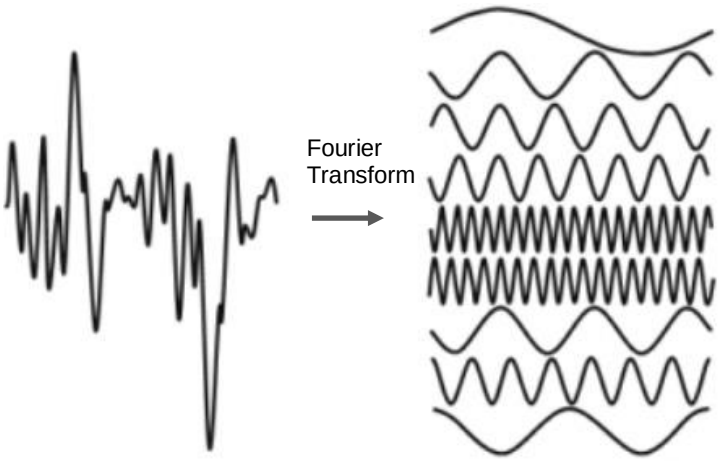


Figure 3.20 Depiction of Fourier transformation. In Fourier transformation, oscillating data is transformed to a wave function (function of time) and decomposed to a set of different frequencies. The sum of frequencies make up the original function. Each frequency has an amplitude that describes how much the given frequency contributes to the average power of the signal (adapted from Phuket, 2014).

Fourier transformation converts the data along the genome into a wave function. As we know that genomic features other than nucleosomes create periodic signal, we extracted the amplitude from the Fourier coefficients corresponding to a periodic signal of nucleosomes (161 bp). By doing this, we captured the periodic signal in MNase data resulting from multiple consecutive positioned nucleosomes. In other words, we transform the data into single amplitude values per 805 bp (five times the nucleosomal length) corresponding to nucleosome periodicity in the window.

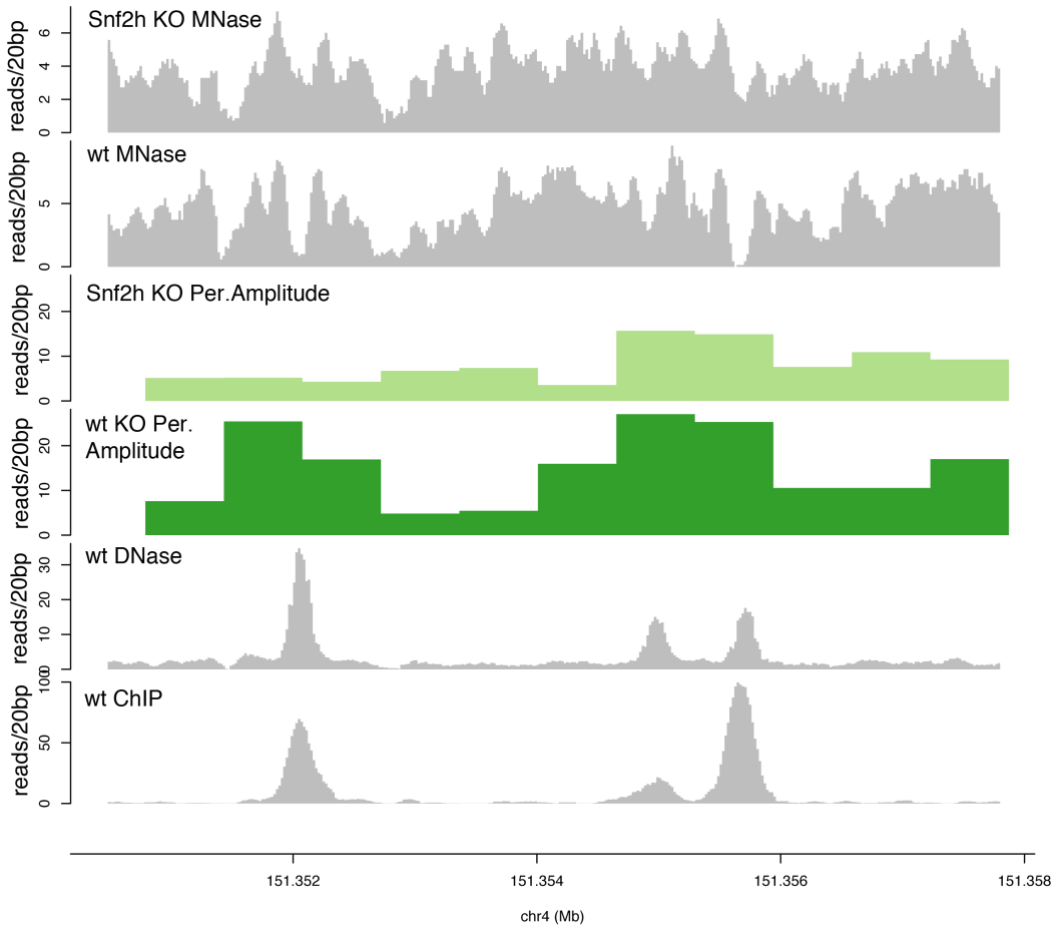


Figure 3.21 Single-locus nucleosome periodicity as calculated by Fourier transformed autocorrelation data (MNase-seq). Amplitude of the frequency corresponding to nucleosomal signal captures differences in nucleosome periodicity previously demonstrated by autocorrelation function. Upper section of the plot represents alignments from the MNase-seq data, lower panel represents DNase-seq and CTCF ChIP-seq data for the given genomic location.

Amplitude values at a single locus seem to recapitulate reduction in nucleosome periodicity previously observed in the autocorrelation data (Figure **3.21**). Having genome-wide amplitude information allows us to scan the genome for regions of altered nucleosome periodicity. We therefore asked if we observe genome-wide difference in nucleosome periodicity in the *Snf2h ko* compared to wt. To this aim, we applied a sliding window across the genome and compared amplitude values. We observed slight, but constant decrease in nucleosome periodicity in *Snf2h ko* cells genome-wide (Figure **3.22**). Furthermore, loss of amplitude/nucleosome periodicity is greater with increasing amplitude levels, suggesting that genomic regions with high nucleosome periodicity tend to proportionally lose more of positioning in the absence of Snf2h. This clearly demonstrates that Snf2h has a role in maintaining nucleosome periodicity genome-wide in mouse ES cells. We next wanted to ask if some regions of the genome are more affected by the loss of Snf2h.

It is known from previous studies that nucleosome positioning/periodicity is most defined in proximity to regulatory regions and transcription factor binding sites (Jiang & Pugh, 2009; Struhl & Segal, 2013). To determine if nucleosome periodicity is altered in those regions in the *Snf2h ko*, we examined amplitude levels of TSS-centered regions as well as DnaseI hypersensitive regions. We found a shift to lower amplitudes in the *Snf2h ko* for both promoter regions (TSS-centered) and transcription factor binding regions (as defined by DnaseI; Figure **3.23**). In contrast, we found no change in amplitude signal for regions that are known to have less positioned nucleosomes. When looking into repressive chromatin (Suz12 regions), the fraction of windows with higher amplitudes in wt is 54.3% (45.7% higher in *Snf2h ko*), indicating almost no difference in positioning for repressive chromatin. Furthermore, the decrease in the amplitude signal in promoters and DNaseI regions is proportionally larger for higher amplitudes, as observed on the genome-wide level. Taken together, it appears regions with high nucleosome positioning lose more periodicity in the *Snf2h ko*. We wanted to further examine the observed changes at promoters and

transcription factor binding sites in greater detail. Specifically, we wanted to investigate (1) how loss of nucleosome positioning/periodicity at promoters is linked to changes in gene expression observed in the *Snf2h ko* and (2) if certain transcription factors show different patterns of changes in nucleosome positioning. The following two subchapters will cover these topics.

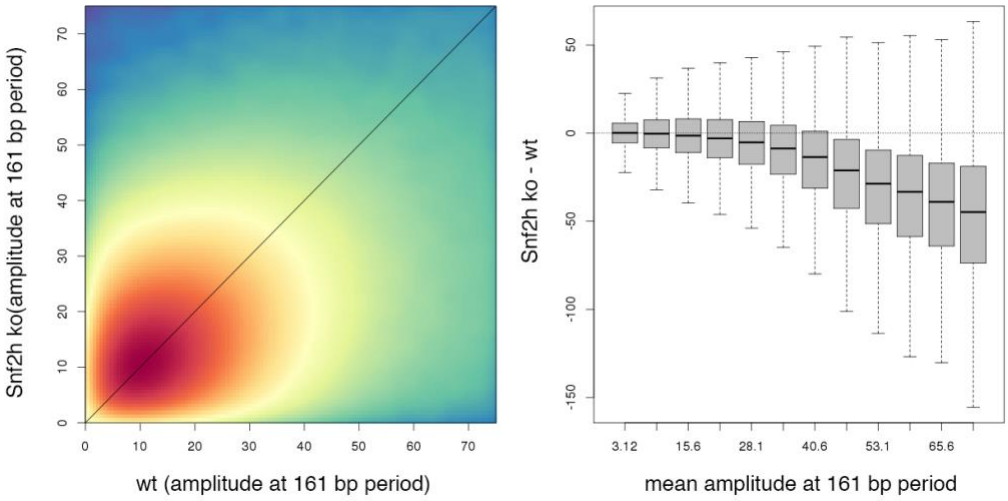


Figure 3.22 Comparison of nucleosome periodicity genome-wide in wt and *Snf2h ko* cells. Left panel; heatmap showing correlation of amplitudes in wt and *Snf2h ko* MNase-seq data. Amplitudes are calculated by extracting the value from the Fourier coefficients corresponding to a periodic signal with period of 161 bp. For the heatmap, amplitudes are calculated on a sliding window of 1416 bp (8 times the average of a nucleosome plus one linker). Neighboring windows are overlapping by 50%. A 708 bp sliding window is applied to the whole genome and amplitudes corresponding to a certain window are plotted for both wt (x axis) and *Snf2h ko* (y axis) samples. A slight increase in signal is observed in the wt sample compared to *Snf2h ko* for amplitudes with values >30. Right panel; boxplot showing difference in amplitude values (*Snf2h ko* to wt) for each bin of amplitudes (bins represent equal increase in amplitude values). Decrease in amplitudes/nucleosome periodicity is observed genome-side and the decrease is greater with higher amplitudes.

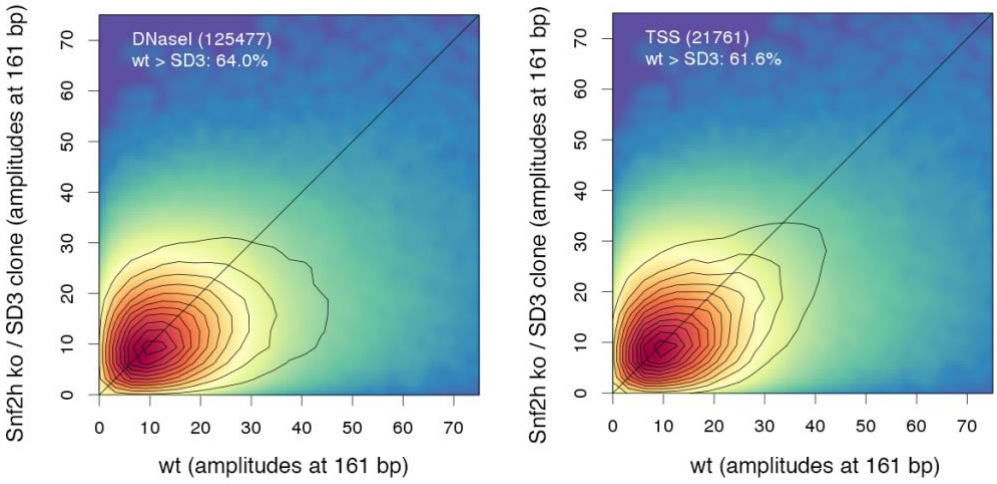


Figure 3.23 Comparison of nucleosome periodicity in wt and *Snf2h ko* cells for transcription factor binding sites (DNaseI) and promoter regions (TSS). Heatmap showing correlations of amplitudes in wt and *Snf2h ko* MNase-seq data. The amplitudes are calculated on a sliding window of 1416 bp (8 times the average of a nucleosome plus one linker). Neighboring windows are overlapping by 50%. The contours show amplitude windows that contain sites of a certain type as indicated in the legend on the top left. Percentage in the top left corner indicates the fraction of windows with higher amplitudes in wt than in the mutant. A slight decrease in signal for Snf2h ko is observed in both DnaseI and TSS regions (wt to *Snf2h ko* ratio of 64% and 61.6% for DNase I and TSS, respectively).

3.2.4 Nucleosome Positioning at Transcription Start Sites is Dependent on Snf2h ATPase Activity

Previous studies demonstrated that chromatin remodelers can modulate transcription by promoter-proximal nucleosome repositioning and eviction in yeast (Chapter 1.3.2). In mammals, dependence of positioning on Snf2h has not been studied in detail. Moreover, whether changes in positioning rely on ATPase activity *in vivo* is not clear, due to lack of cellular ko models to test it. We therefore aimed to investigate if the changes in gene expression we observed upon Snf2h deletion are linked to changes in promotor-proximal nucleosome positioning. For this purpose, we looked into average MNase-seq signal +/- 1kb from TSS across all mouse promoters.

In *Snf2h ko* cells, we observed a slight decrease in positioning of nucleosomes flanking TSS (Figure 3.1). This effect is evident on both TSS downstream and upstream nucleosomes. Furthermore, the extent of the observed changes is comparable to yeast Isw1 and Isw2 mutant strains (Figure 3.1). This indicates that the *in vivo* function of ISWI is conserved from yeast to mammals at promoter regions. To determine if the observed changes in nucleosome positioning are dependent on ATPase activity, we investigated promoter nucleosome positioning in V5-Snf2h re-expression lines. Re-expression of wt Snf2h, but not ATPase mutant rescues the nucleosome positioning phenotype (Figure **3.24**). This is a clear indication positioning effect we are observing is caused by the lack of ATP-dependent chromatin remodeling. Next, we asked if genes with different expression levels show differential positioning effect in the *Snf2h ko*. Although highly expressed genes had stronger promoter-proximal nucleosome positioning, the same degree of decrease in positioning was observed across all genes expressed in ES cells (data not shown). Finally, we asked if genes that show differential expression in the *Snf2h ko* also exhibit changes in positioning. Strikingly, all promoters irrespective of their transcriptional response in the ko show a decrease in positioning (Figure **3.25**). This indicates that Snf2h has a general role in maintaining promoter-proximal nucleosome positioning, but the effect on transcription is dependent on other downstream factors. Furthermore, the dependence on the ATPase activity itself was evident; *Snf2h wt-ab*, but not *Snf2h atpm-ab*, shows rescue of the phenotype at all promoters (Figure **3.25**). This indicates that changes in nucleosome positioning we see in all promoters are due to lack of ATPase-mediated remodeling. In summary, all active promoters show a loss of promoter-proximal nucleosome positioning in *Snf2h ko*, irrespective of their transcriptional change. We further wanted to examine if other regions of the genome show loss of nucleosome positioning.

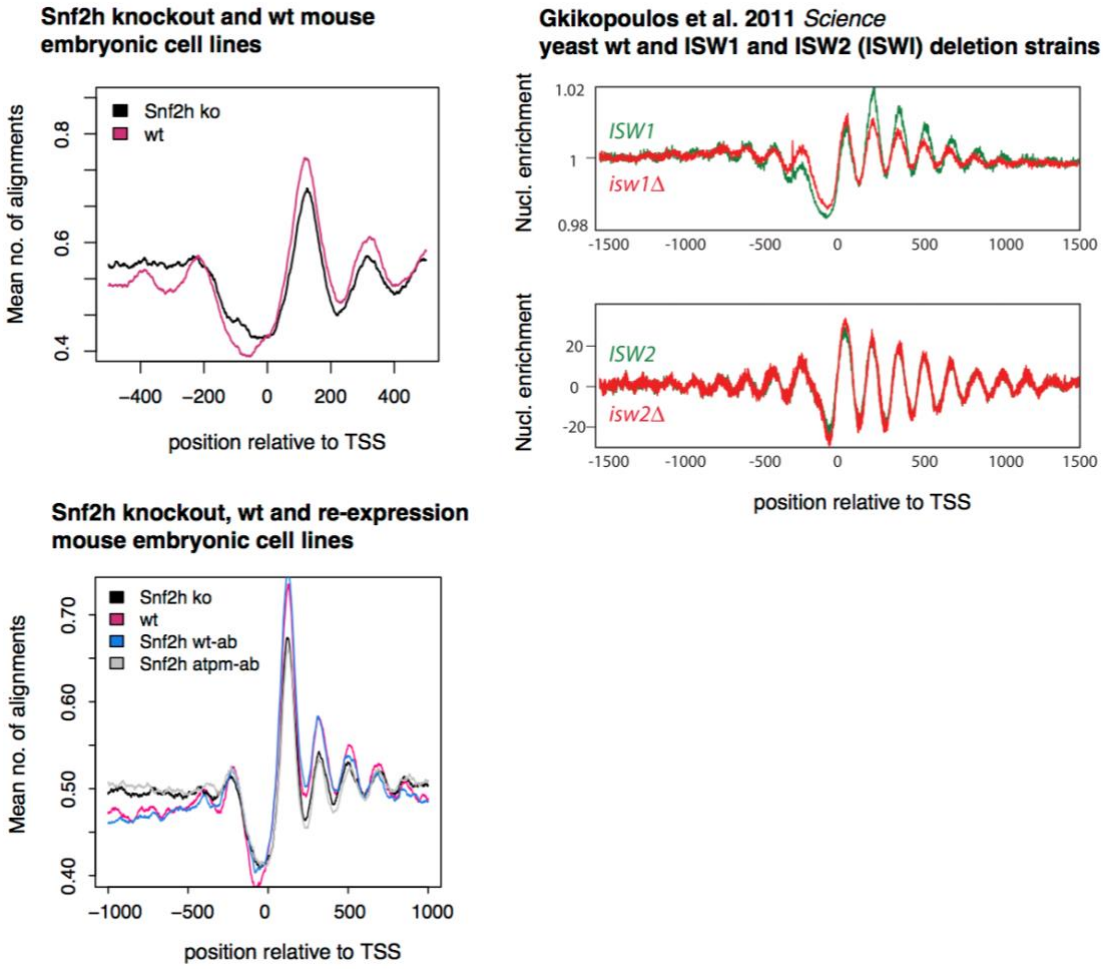


Figure 3.24 Nucleosome positioning within promoter regions for wt and Snf2h ko in both mouse ES cells and yeast. Left panel: average promoter plots of MNase-seq alignments anchored on TSS and extended 500 bp (upper) or 1000 kb (lower) in both directions. MNase reads are shifted by half of the nucleosome size to reflect nucleosome positions. All mouse promoters are plotted and smoothing of 20 bp is applied. Upper left panel shows nucleosomes flanking TSS are less positioned in *Snf2h ko*. Lower panel shows the positioning is rescued in wt but not ATPase mutant re-expression. Right panel; average promoter plot of MNase alignments in yeast ISWI homologues. Changes in nucleosome positioning around TSS are similar in mouse and yeast mutants.

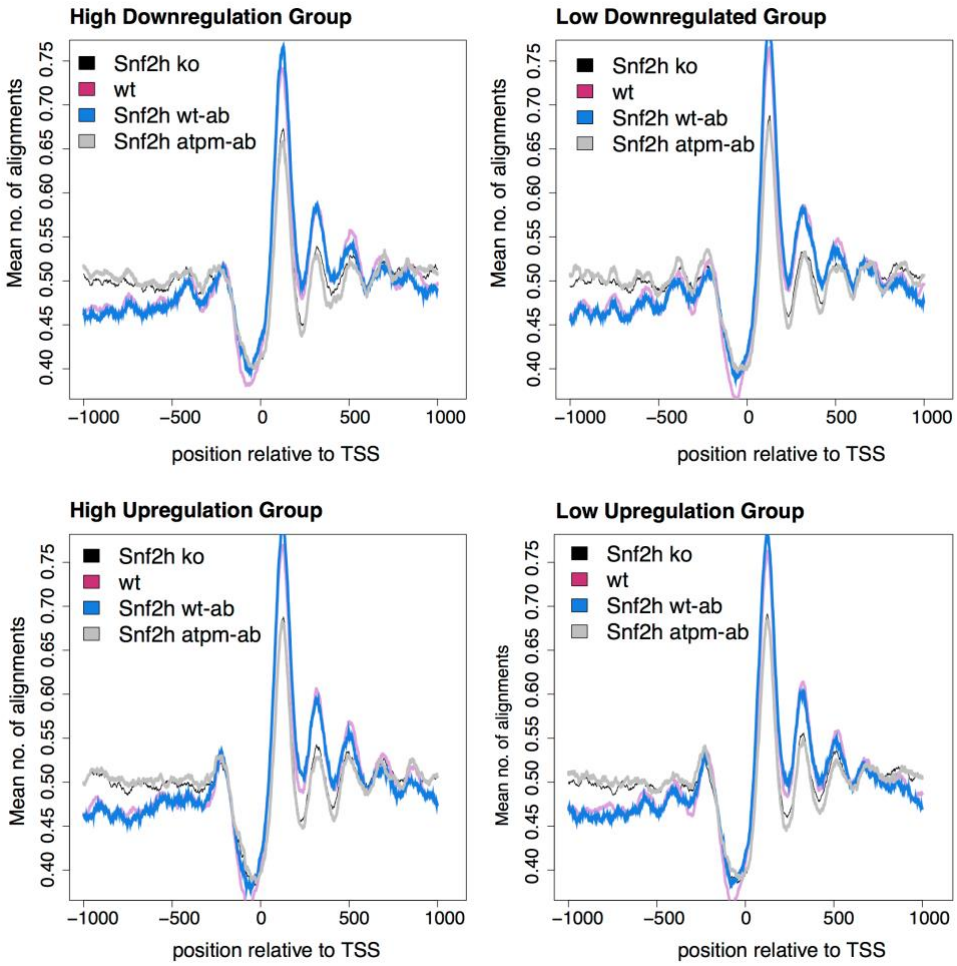


Figure 3.25 Nucleosome positioning within promoter regions stratified by level of transcriptional response in *Snf2h ko*. Average promoter plots of MNase-seq alignments anchored on TSS and extended 1000 bp in both directions. MNase reads are shifted by half of the nucleosome size to reflect nucleosome positions. Smoothing of 20 bp is applied. Promoters are stratified on the degree of transcriptional change in the *Snf2h ko*. No difference between different group is apparent; both upregulated and downregulated groups exhibit the same loss of positioning in *Snf2h ko*. All group follow the same trend in the re-expression cell lines (*Snf2h wt-ab* and *Snf2h atpm-ab*).

3.2.5 Nucleosome Positioning in Proximity to Transcription Factor Binding is Affected by Loss of Snf2h

To identify transcription factors that show a change in nucleosome positioning/periodicity, we scanned the genome for all known transcription factor motifs. We considered only factors expressed in ES cells. Two measures were used when examining changes over transcription factor motifs: (1) change in average amplitude levels over a 2 kb window around the motif and (2) changes in the MNase signal over the motif itself. Increase in nucleosome occupancy over the motif itself will result in an increase in MNase-seq coverage over the motif. This will likely reflect in loss of factor binding at the respective motif. Upon examining these parameters for over 50 different transcription factors and their respective motifs, we identified three different profiles of nucleosome positioning (three groups;
Figure **3.26**). In the group 1 we observed both; loss of nucleosome positioning up to +/- 2 kb from the motif and increase in nucleosome occupancy over the motif. CTCF, E2F1 and SP4 are among transcription factors that show this trend. It is likely these factors require Snf2h to reposition nucleosomes away from the motif in order to bind. The second class of factors (group 2) exhibits reduction in amplitude values and therefore loss of positioning/periodicity in the +/- 2kb window. However, this group shows no change in signal over the motif itself, indicating the factor binding is unchanged. In those examples, that include REST, STAT3 and SNAI2, binding is likely enabled by the action of another remodeler that repositions/evicts motif-bound nucleosome. In fact, both REST and STAT3 binding was shown to be dependent on Brg1, the ATPase subunit of SWI/SNF complexes (Ho et al., 2011; Ooi, Belyaev, Miyake, Wood, & Buckley, 2006). Nucleosomes upstream and downstream of the motif show a slight reduction in positioning for this group. Whether this loss of positioning is functionally relevant is still to be tested. Loss of positioning at those nucleosomes might, in fact, enable binding of other nucleosome-sensitive transcription factors in proximity. The third group shows the opposite effect; increase in nucleosome positioning over the +/- 2kb window

and no change in read coverage over the motifs. This is also likely a result of other remodelers being recruited to the binding site upon activation of the factor. This was shown previously to be the case for p53 (found in this group); p53 is bound to chromatin, but recruits co-activators only upon activation (Barlev et al., 2001). Activation of these factors upon Snf2h depletion is likely to be a secondary effect. However, observing both increase and decrease in nucleosome positioning of *Snf2h ko* cells further confirms the data is not compromised by potential under- or over-digestion of the MNase treatment.

Taken together, we identified three groups of transcription factors that behave differently in response to loss of Snf2h. We further focused on group 1, where we expect Snf2h to have a direct effect on transcription factor binding by repositioning of a motif-bound nucleosome. For factors in this group, namely CTCF, we aimed to test if the increase in nucleosome occupancy over the motif coincides with a loss of CTCF binding.

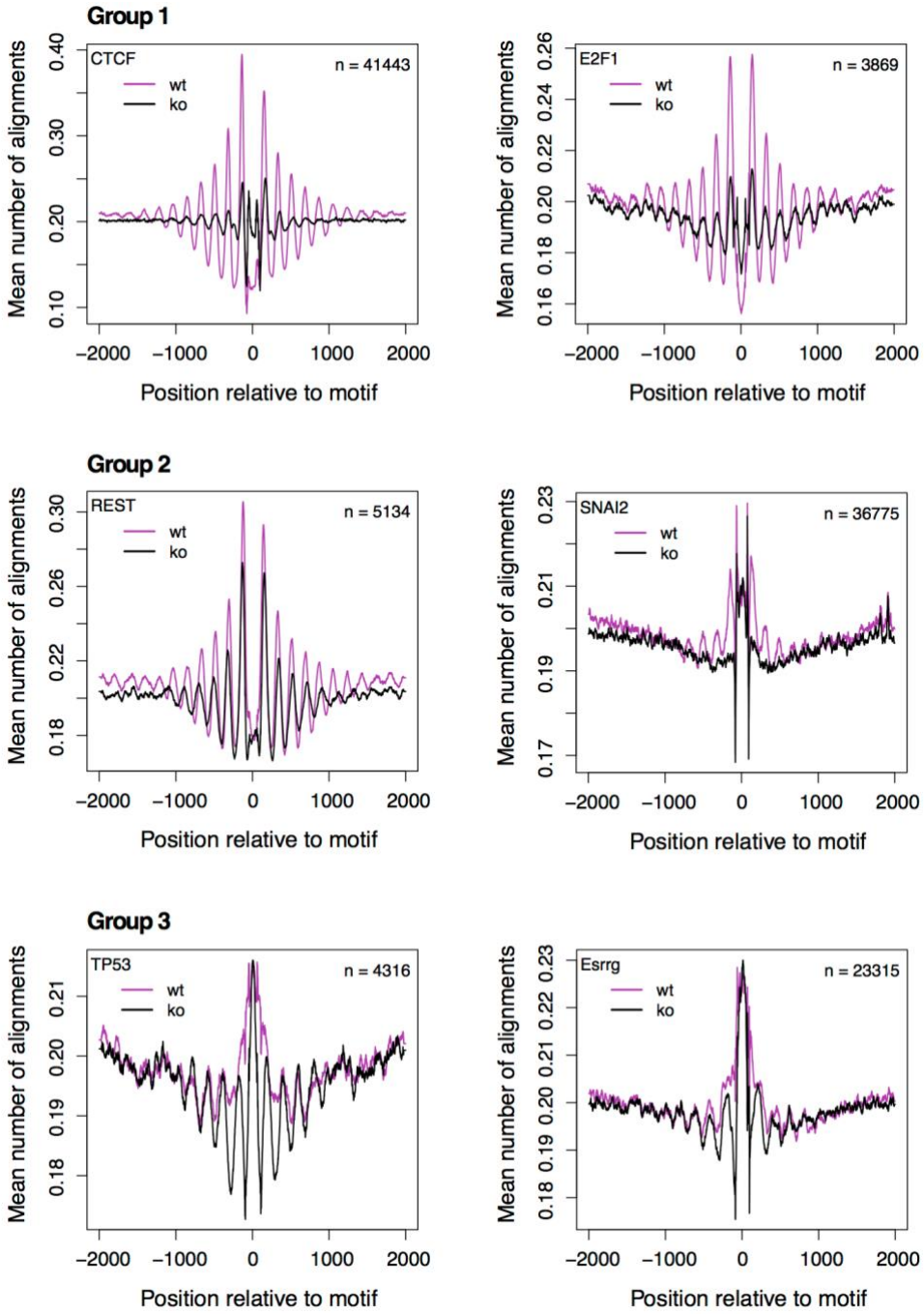


Figure 3.26 Nucleosome positioning flanking transcription factor motifs in wt and *Snf2h ko* cells. Average alignment plots of MNase-seq alignments anchored on transcription factor motifs and extended 2000 bp in both directions. 3 groups were identified based on (1) changes in nucleosome positioning as determined by amplitude values across the region and (2) change in nucleosome occupancy over the transcription factor motif. Group 1 shows reduction in amplitude over the region and increase in MNase signal over the motif, group 2 shows reduction in amplitude and no change in motif occupancy, group 3 shows increase in amplitude and no change in motif occupancy.

3.2.6 CTCF Binding is Globally Reduced in the Snf2h Knockout

CTCF ChIP-seq

So far, we have observed a striking loss of nucleosome positioning around CTCF motifs in the *Snf2h ko* cells. Furthermore, we observed an increase in nucleosome occupancy over the CTCF motif, indicating that CTCF binding might be dependent on Snf2h. To determine if this is the case, we performed ChIP-seq on wt and Snf2h ko cells, including the re-expression cell lines (*Snf2h wt-ab* and *Snf2h atpm-ab*). In *Snf2h ko*, we observed a global reduction in CTCF binding (Figure **3.27**; Figure **3.28**).

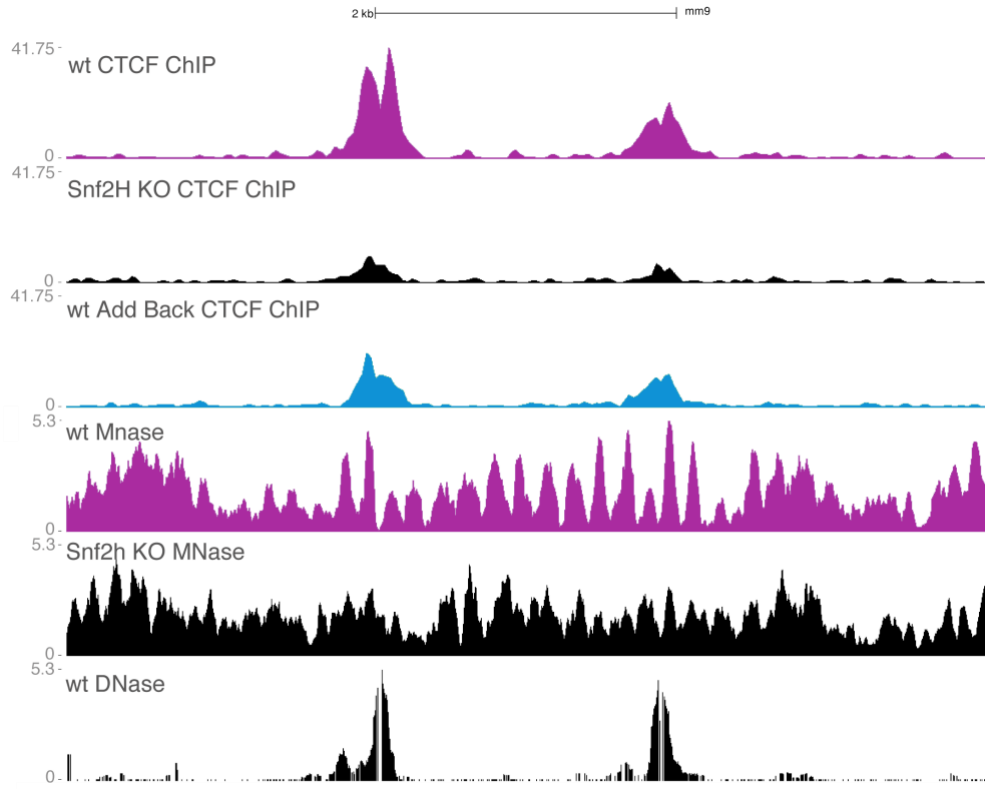


Figure 3.27 Single locus example of CTCF ChIP-seq profile in wt, Snf2h ko and Snf2h wt-ab. Snf2h ko line shows a decrease in CTCF ChIP signal. The loss of ChIP signal coincides with loss of positioning in proximity to the binding region. Re-expression of wt in the ko background shows rescue of the binding.

We next asked if loss of binding is restricted to certain sites/motifs, however, we observed all CTCF sites are affected by the loss of Snf2h (Figure **3.28**). When determining the degree of reduction in binding, we observed the binding is not lost, but reduced by a factor of two on average across all sites. This indicates that CTCF binding affinity is decreased upon loss of Snf2h. For the wt re-expression cell line, we observed reversal of the effect. *Snf2h wt-ab* highly correlates with wt (0.83) and the correlation is reduced when compared to the ko (0.64; calculated on 250 bp genomic windows around CTCF motifs; n=391,862). The rescue of CTCF binding is not complete. It is however, uniform across all CTCF binding sites, suggesting this is likely due to low levels of re-expression of the V5-Snf2h construct. Whether re-expression of V5-Snf2h to higher wt levels will fully rescue CTCF binding remains to be tested. To confirm the effect on nucleosome positioning we observe in the *Snf2h ko* is not due to a decrease in levels of CTCF, we checked for CTCF protein and RNA expression.

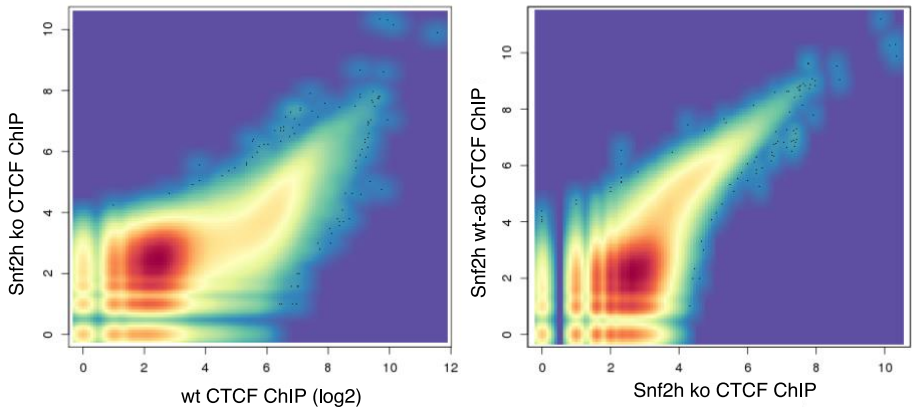


Figure 3.28 Genome-wide CTCF ChIP-seq profile in wt, *Snf2h ko* and *Snf2h wt-ab*. Density plots show CTCF binding for 250 bp fragments centered on CTCF motifs (both bound and unbound, n=391,862). Comparison of wt and *Snf2h ko* shows a global loss of signal across all regions (left panel). The effect is reversed when comparing ko to wt re-expression (right panel).

Both protein and RNA levels (as shown by RNA-seq and western blot) showed no change in CTCF expression (Figure 3.1 for RNA expression; protein expression data not shown), further confirming the observed effect in positioning is driven by Snf2h.

ATAC-seq

Finally, we wanted to further test if the observed changes are not caused by artifacts or technical biases in the methods we used (ChIP-seq, MNase-seq). This is especially relevant for the ChIP-seq experiments. In the ChIP-seq for *Snf2h ko*, we observe a global reduction in the ChIP-seq signal compared to wt, as mentioned previously (Figure **3.28**). The global reduction in the signal might come from lower efficiency of the immunoprecipitation specifically in the *Snf2h ko* cells. To ask if the reduction in the ChIP-seq signal is indeed caused by loss of CTCF binding, we performed ATAC-seq on the *Snf2h ko* and wt cell lines. ATAC-seq is an assay that measures chromatin accessibility (Buenrostro et al., 2015). Accessibility is measured as the frequency of the DNA adapter incorporation by the mutated hyperactive transposase Tn5. Tn5 preferentially incorporates in the stretches of exposed DNA, mostly in regions surrounding transcription factor binding sites and nucleosomes (linker regions; Buenrostro et al., 2015). ATAC-seq, therefore, provides a readout for both transcription factor binding and nucleosome positioning in a single assay. Following the ATAC-seq protocol, we observed reduction of ATAC-seq signal at the CTCF motifs (Figure **3.29**), indicating loss of binding. This further confirms our observations from ChIP-seq experiments. Furthermore, we observe a strong nucleosome positioning signal in proximity to CTCF motifs (Figure **3.29**). As expected, this positioning is decreased in the *Snf2h ko* compared to wt.

Taken together, ATAC-seq confirmed the loss of CTCF binding and nucleosome positioning in the *Snf2h ko*. In the ongoing analysis, we are using ATAC-seq to (1) identify other factors that lose binding upon Snf2h deletion, (2) determine if loss of positioning in proximity to CTCF creates new regions of open chromatin and enables binding of other transcription factors.

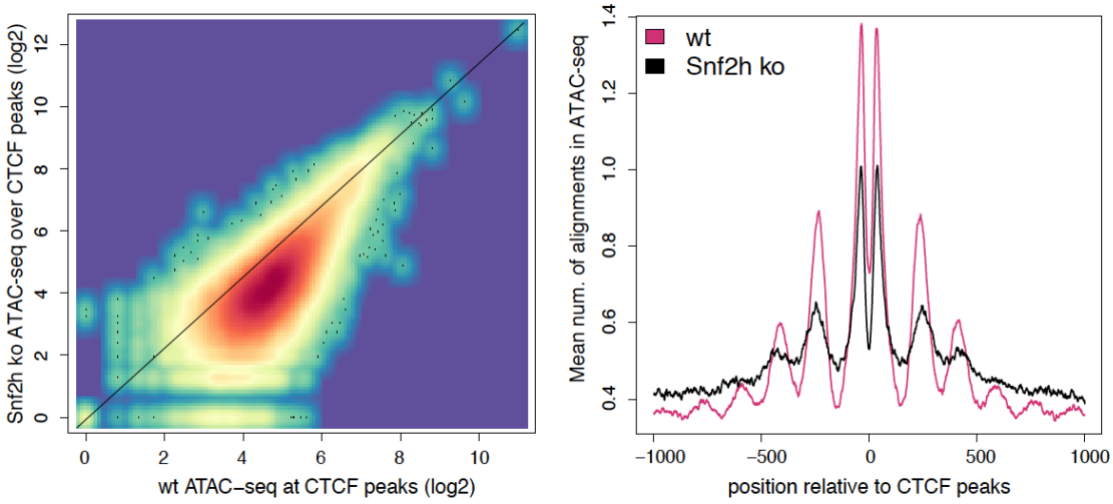


Figure 3.29 Chromatin accessibility at CTCF motifs in *Snf2h ko* and wt. Left panel; ATAC-seq alignments collected over CTCF bound motifs (peaks) in wt and *Snf2h ko,* n=31,153. A global shift in the signal from the diagonal is observed, indicating global reduction in accessibility in *Snf2h ko*. Right panel; average alignment plots of ATAC-seq alignments anchored on CTCF bound motifs (peaks) and extended 1000 bp in both directions (n=31,153). A reduction in both CTCF binding and nucleosome positioning is evident.

3.2.7 Highest Loss of Nucleosome Positioning is Found at Strongest CTCF Motifs

We next aimed to determine whether certain CTCF sites show differential nucleosome positioning in the *Snf2h ko*. First, we wanted to ask if certain CTCF sites show a differential behavior in the Snf2h re-expression rescue. We observed, however, neither wt nor ATPase mutant re-expression lines rescue the nucleosome positioning phenotype in proximity to CTCF motifs

(data now shown). As previously mentioned, we cannot directly compare wt sample with re-expression lines since Snf2h protein levels in the re-expression clones are expressed to a much lower level compared to wt. However, as the two re-expression constructs are expressed to the same levels (as measure by western blot; data not shown), we compared the two constructs in respect to each other. Indeed, when comparing nucleosome positioning around CTCF motifs, we observe a slight reduction in the ATPase mutant re-expression line compare to wt re-expression (Figure **3.30**). This indicates nucleosomes positioning flanking the CTCF motifs is at least partially dependent on the ATPase activity. To further confirm this, we will have to establish re-expression clones with wt-matching expression levels (ongoing).

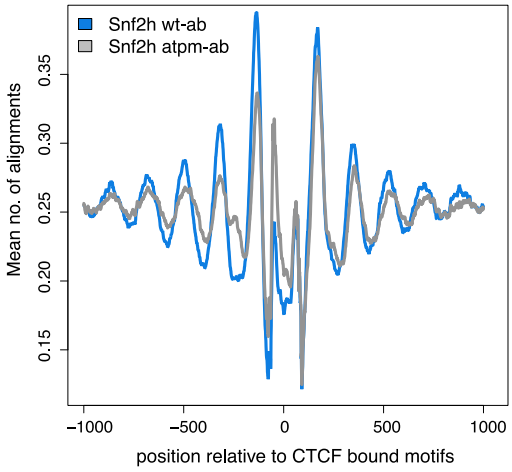


Figure 3.30 Nucleosome positioning flanking CTCF motifs in wt and ATPase mutant re-expression lines. Average plots of MNase-seq alignments anchored on transcription factor motifs and extended 1000 bp in both directions. Reduction in nucleosome positioning is apparent in the ATPase mutant compared to wt re-expression.

Next, we examined if CTCF motifs with different motif strengths show differential nucleosome positioning pattern in the *Snf2h ko*. CTCF motif strength has been shown to correlate with binding (Stadler et al., 2011). Out of 391,862 CTCF motif instance in the genome, we selected top ~ 10 percent with the highest motif score and further selected the motifs bound in wt (n=31,153). In the next step, we stratified those motifs into 5 bins of motifs score as determined by the position weight matrix (pwm) score and binding strength. We observed that nucleosome positioning increases with increasing motif strength as shown by average nucleosome positioning profiles, for both in wt and *Snf2h ko* (Figure **3.31**). Interestingly, we observed the highest score motifs lose more positioning in the ko compared to wt. In

fact, motif score seems to scale with loss of positioning. We observed the same trend in the CTCF ChIP data; regions with highest ChIP enrichment proportionally exhibit the highest loss of binding in the *Sfn2h ko*. Genomic regions of high transcription-factor occupancy, and respectively high motif score, are known to exhibit higher nucleosome turnover than other parts of the genome. Therefore, we want to further investigate if Snf2h impacts nucleosome turnover. H3.3 histone variant is known to be indicative of histone turnover (Chapter 1.3.3). We will assay H3.3 and total H3 levels by ChIP-seq in both wt and *Snf2h ko* (at the time of the thesis submission, experiments and data analysis have not been completed).

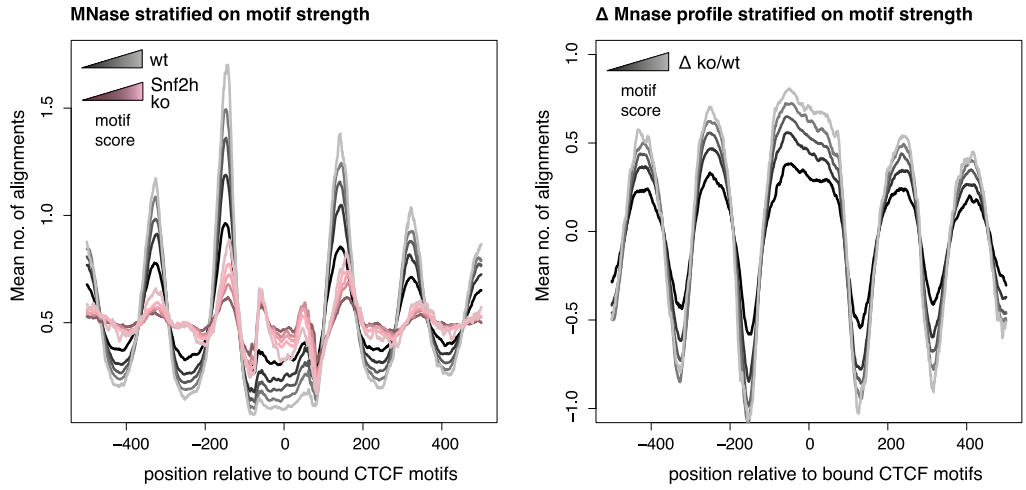


Figure 3.31 Nucleosome positioning profiles flanking CTCF binding sites stratified on motif strength. 5 bins of equal motif-strength span have been created. From highest to lowest motif score bins, they represent; 690, 4985, 9638, 9313 and 6537 CTCF motifs. In both wt and Snf2h ko, strongest nucleosome positioning change is observed in high scoring motifs (right panel). Furthermore, biggest decrease in signal upon Snf2h deletion is observed at high-score CTCF motifs (highest scoring motifs lose more positioning).

Chapter 4

Discussion and Outlook

4.1 Genomic Targeting and Transcriptional Regulation by PcG Proteins

Ever since PcG proteins were found to be important for maintaining developmental decisions in flies, efforts were made to understand how these complexes are recruited to target sites. In flies, PREs were identified as genetic elements sufficient to recruit PcG proteins and induce H3K27me3-mediated repression (Beisel & Paro, 2011). However, PREs have not yet been characterized in mammals. Here, we investigated sequence determinants of PcG recruitment in mouse embryonic stem cells. Furthermore, we described the role of PRC2 in transcriptional repression in two mammalian cell lines and during *in vitro* differentiation.

4.1.1 DNA Sequence Determinants of PRC2 Recruitment in Mouse Embryonic Stem Cells

Since PREs in flies harbor various specific motifs, transcription factors are thought to drive recruitment of PcG proteins. In mammals, PREs are not as well defined. However, several studies showed that transcription factors do contribute to PRC2 recruitment in mammals (P. Arnold et al., 2013; Tien et al., 2015). For example, mutations of either Rest or Snail motifs reduced H3K27me3 levels in mouse embryonic stem cells (P. Arnold et al., 2013). Yet, in both flies and mammals, a single transcription factor was not proved to be sufficient to autonomously recruit the PRC2 complex. We postulate that PRC2 recruitment is rather a combinatorial effect of various transcription factors. Indeed, factors that have been shown to bind PREs in flies execute very diverse chromatin functions from nucleosome positioning

to DNA bending (Margueron & Reinberg, 2011; Schwartz & Pirrotta, 2007). This indicates that PRC2 mediated repression is a multi-step process of chromatin changes. Furthermore, PRC2 targets are mainly promoter proximal regions, rich in binding motifs for variety of transcription factors (Margueron & Reinberg, 2011). This further supports the idea that PRC2 is recruited by more than one specific transcription factor. In this project, we aimed to determine whether combinations of transcription factor motifs facilitate recruitment of PRC2 in mammals. This was carried out by investigating the contribution of both complex transcription factor motifs and those composed of only CpG dinucleotides. Our approach consisted of integrating hundreds of endogenous PRC2 sequences with various sequence properties into an ectopic genomic locus and asking which sequences reconstitute PRC2 binding and H3K27me3 mark. Within this experimental setup, we did not identify a specific transcription factor motif or a combination thereof that has the potential to autonomously recruit PRC2. However, motif enrichment analysis cannot identify factors that bind low complexity motifs such as the CpG motif. We therefore asked if CpG motif frequency scales with PRC2 enrichment in our insertion library. Interestingly, we observed that CpG density is the best predictor of PRC2 recruitment in our library of ectopic insertions. We further tested this idea by inserting a library of bacterial (*E. coli*) sequences into the same ectopic location. These bacterial sequences do not harbor known complex mammalian transcription factors motifs, but contain low complexity motifs such as CpG dinucleotides. We observed that PRC2 binds to many of these bacterial sequences. Moreover, sequences with higher CpG density show higher PRC2 enrichments, for both endogenous PRC2 sequences and sequences derived from *E. coli*. This leaves us with the notion that CpGs might be the sole recruiting signal for PRC2 in mammals. Since non-methylated CpGs are recognized by CXXC-domains found in many chromatin-modifying complexes, it is likely that PRC2 recruitment is mediated through CXXC-domain proteins. As a follow-up of this part of the project, we want to test whether depletion of certain CXXC-domain proteins

will reduce binding of PRC2 in both our insertion libraries. To do so we would test several candidate factors by either CRISPR/Cas9 knockout or shRNA knockdowns. JARID2 and AEBP2 would be attractive candidates, as they have been implicated in PRC2 recruitment (Landeira & Fisher, 2011; Peng et al., 2009; Son et al., 2013).

CXXC-domains are in fact found in various proteins that regulate lysine methylation. Such examples would include CFP1 (CxxC finger protein 1), MLL (mixed lineage leukaemia protein), KDM (lysine demethylase) 2A and KDM2B (Di Croce & Helin, 2013). In fact, KDM2B was reported to be involved in the recruitment of the PRC1 complex (Farcas et al., 2012; He et al., 2013; Wu et al., 2013). ShRNA-induced depletion of KDM2B was shown to reduce genome-wide levels of PRC1 and subsequent de-repression of a portion of PRC1-target genes (Farcas et al., 2012; He et al., 2013; Wu et al., 2013). Whether PRC2 is recruited to CpG islands through PRC1 and KDM2B is still an unanswered question. Historically, PRC1 recruitment was thought to act downstream of H3K27me3 deposition by PRC2. Discovery of the PRC2-independent mechanism of PRC1 targeting recently led to a new hypothesis that PRC1 precedes PRC2 deposition. Interestingly, in our library of endogenous Polycomb sequences we observed no enrichment for PRC1's catalytic subunit Ring1B (data not shown). This suggests that in an isolated genomic context, PRC2-recruiting sequences target PRC2 independently of PRC1. However, our insertion library identified two groups of putative PRC2 sequences; one that recapitulates binding at the ectopic location and one that does not. The fraction of endogenous sequences that do not reconstitute binding ectopically might be dependent on PRC1. What are other distinguishing features of the two identified groups and whether they exhibit distinct functions is yet to be determined.

Taken together, it seems CpG density is the driving signal for PRC2 binding in an isolated, ectopic context. The observation that most PRC2 targets are CpG islands further supports this concept. However, CpG islands only occur in vertebrates that show broad DNA methylation of their genome. *D. melanogaster* for example lacks DNA methylation and CpG islands, yet PcG

group proteins are highly conserved from flies to mammals. How the PcG system evolved to be targeted to a novel genetic feature in mammals is an interesting question. It is tempting to speculate that PRC2 is simply targeted to all genomic regions with high non-methylated CpG density and that the differential regulation is established through the opposing activity of TrxG proteins. The recent observation that PRC2 binding increases at all CpG islands after genome-wide inhibition of transcription further supports this idea (Riising et al., 2014). According to this model, the role of PRC2-mediated repression might be to ensure that activation occurs only in the presence of a strong activating signal. This model would suggest that PRC2 has a role in reducing transcriptional noise.

4.1.2 Transcriptional Regulation by PRC2 in Mouse Embryonic Stem Cells and during Differentiation

Early studies reported that Ezh2 and Eed are required for self-renewal and pluripotency of mouse ES cells (Boyer et al., 2006; Leeb et al., 2010). However, we showed that deletion of Eed does neither alter self-renewal nor does it seem to alter differentiation of mouse ES cells. During the course of our study, Riising et al. published a similar finding (Riising et al., 2014). In their study, they deleted Suz12 and Ezh2, the two other core subunits of PRC2 and observed no transcriptional response (Riising et al., 2014). The discrepancy between earlier and recent studies might be explained by several factors. (1) Specific growth conditions or other technical differences might produce the differences observed. However, we cultured *Eed ko* ES cells provided by the Prof. Anton Wutz group (Leeb et al., 2010) with our standard culturing conditions (see Methods, Chapter 5.1). Transcriptional analysis of ES cells provided by the Prof. Anton Wutz group revealed the same genes to be upregulated in our culture conditions as they previously published (comparison of Figure **3.9** and Leeb et al., 2010). This is rather surprising considering that we and Leeb et al. used different technologies to determine transcriptional changes (RNA-seq and Affymetrix microarray, respectively). Yet, the results are very comparable. Taken together, it

seems the transcriptional phenotype is rather stable. This indicates that the discrepancy between different studies does not come from culturing conditions. (2) Differences in the transcriptional response might also come from clonal differences. In the earlier studies, the deletions were generated by using gene targeting methods based on homologous recombination. This requires that cells go through several rounds of clonal selection. In fact, *Eed ko* ES cells generated by Leeb et al. had four iterations of clonal selection. This in turn might lead to accumulation of clonal mutations and substantial differences in the transcriptional profile when compared to the parental line. We used the CRISPR/Cas9 technology to delete PRC2 components. CRISPR/Cas9 requires only one clonal selection to generate a homozygous mutant, making the comparison between ko and wt more controlled. Therefore, CRISPR/Cas9 generated PRC2 mutants are more isogenic to respective wt lines than mutants created with earlier genetic targeting approaches. This argues that the lack of transcriptional response we observe in our system reflects more closely the *Eed*-deletion phenotype. PRC2 ES cell mutants generated by previous studies might exhibit spontaneous differentiation, which in turn might mimic the PRC2 phenotype. In fact, Riising et al. showed that culturing cells over a prolonged period of time shows transcriptional deregulation of Polycomb targets (Riising et al., 2014). Taken together, we believe the CRISPR/Cas9 system enabled us to generate a more controlled system of PRC2 deletion than previously reported. We were therefore able to determine that there are no transcriptional changes in mouse ES cell upon PRC2 deletion.

The observed lack of transcriptional response in *Eed ko* cells indicates that PRC2 is dispensable for gene repression in ES cells. Other repressive mechanisms might be compensating for the loss of PRC2 in these cells. PRC1 might in fact still be targeted to PRC2 sites through recruiting mechanisms independent of H3K27me3. Furthermore, DNA methylation is another repressive modification that might be rescuing the loss of PRC2 repression in the *Eed ko*. In fact, reduction in genome-wide H3K27me3 levels was shown to induce a slight increase in DNA methylation at certain

PRC2 target regions (Thornton, Butty, Levine, & Boyer, 2014). Whether this is functionally relevant remains to be tested. Regardless of the compensation mechanism following depletion of PRC2, the lack of transcriptional response does not enable us to investigate mechanisms of PRC2-mediated repression in ES cells. Therefore, we investigated if PRC2 deletion exhibits a transcriptional phenotype during differentiation. We differentiated *Eed ko* mouse Es cells into neuronal progenitors. Contrary to previous reports, *Eed ko* formed embryoid bodies and neuronal progenitors morphologically similar to wt cells. Following the RNA-seq and transcriptional analysis of the *Eed ko* neuronal progenitors, we observed substantial changes in gene expression compared to wt neuronal progenitors. 2719 genes showed differential expression (adjusted p-value cut-off 0.01, linear fold change cut-off; Figure **3.11**). 44% of upregulated genes were PRC2 targets in wt neuronal progenitors, indicating that a large part of the phenotype might be a direct effect of loss of PRC2. It is not unexpected that PRC2 deletion shows a significant transcriptional phenotype during differentiation since PRC2 was shown to maintain developmental decisions in developing fly embryo (Schwartz & Pirrotta, 2007). Specifically, PRC2 exerts its repressive function upon embryo segmentation. Mouse ES cells are typically derived from the inner cell mass of the embryonic day E3.5 early blastocyst. At this stage, the embryonic cells have not yet started the segmentation and initial differentiation into the three germ layers (Bryja, Bonilla, & Arenas, 2006; Tam & Behringer, 1997). ES cells are therefore an unfit model for studying PRC2-mediated repression. This stage is likely too early in development to observe the repressive function of PRC2. The fact that we only see changes in gene expression upon differentiation further confirms this notion. Nevertheless, observing the transcriptional phenotype upon *Eed ko* differentiation will allow us to further investigate PRC2-mediated silencing. Moreover, it will enable us to link repression with recruitment of PRC2. We will be able to create a library of PRC2 target sequences that respond to loss of PRC2 by upregulation of the associated gene during differentiation. Inserting such

library into the ectopic location and monitoring PRC2 recruitment will allow us to further link determinants of PRC2 recruitment with repression. The transcription factors that recruit PRC2 might not be present in ES cells. In the differentiation system that shows a transcriptional response to loss of PRC2, we anticipate to see a bigger contribution of transcription factors to PRC2 recruitment than currently observed in ES cells. Furthermore, by coupling this library with a reporter gene we would be able to screen for mammalian PRE sequences. The identified genes in the *Eed ko* neuronal progenitors can be further used in other reporter assays to identify *trans*-acting factors involved in Polycomb recruitment.

4.2 Function of ISWI in Mouse Embryonic Stem Cells

Chromatin remodeling complexes have been implicated in many biological processes from DNA replication and DNA damage to regulation of gene activity. However, most of our knowledge about regulatory mechanisms of these processes has been limited to studies *in vitro* and in yeast. Although ISWI remodelers are structurally conserved from yeast to mammals, it is not clear if they exert the same *in vivo* function in mammals. Here, we generated and described the first viable chromatin remodeler ATPase mutant in a mouse ES cell line. Furthermore, we showed that loss of Snf2h affects nucleosome positioning in proximity to promoters and transcription factors motifs.

4.2.1 Role of Snf2h Complexes in Promoter-Proximal Nucleosome Positioning

Chromatin has an important role in most nuclear processes in eukaryotes. Most of these processes entail changes to chromatin structure by chromatin remodeling complexes. With chromatin remodelers having such a fundamental and widespread role, it is difficult to determine the mechanism of chromatin remodeling *in vivo*. It is particularly challenging to determine the causal relationship between events on chromatin and link them to remodeling activity. We therefore aimed to develop a cellular system that will allow us to investigate loss of function of ISWI remodelers and link the observed changes to the loss of ATPase activity. We deleted Snf2h, the ATPase of ISWI using CRISPR/Cas9. We show that *Snf2h ko* mouse ES cells are viable, stably proliferate but exhibit slower growth rate. The observed growth phenotype might potentially confound the interpretation of data derived from the *Snf2h ko* model. However, we showed that *Snf2h ko* cells do not seem to be stalled in one phase of the cell cycle. Rather, it seems *Snf2h ko* cells simply progress slower through all stages of the cell cycle. This suggests that observed changes are very unlikely to be a

reflection of changes in the cell cycle. However, we cannot exclude that alteration of the growth rate might have an impact on the interpretation of our findings. Further characterization of the growth phenotype thus seems necessary. This can be accomplished by using the BrdU pulse-chase labeling assay that provides insights into the cell cycle kinetics. Understanding kinetics of the changes will enable us to precisely determine alterations in the cell cycle progression and further link it to other observations.

In yeast, ISWI was shown to affect promoter-proximal nucleosome positioning (Figure **3.24**). We therefore asked if the loss of Snf2h would mimic the nucleosome positioning phenotype in yeast. Indeed, we observed a very similar effect on the TSS-proximal nucleosome positioning. We observed loss of nucleosome positioning upstream and downstream of the TSS. Furthermore, we observed that this positioning is ATPase dependent. This was shown by re-expressing the V5-Snf2h proteins into the *Snf2h ko cell* line. We observed that wt but not the ATPase mutant version of the protein was able to rescue the nucleosome positioning phenotype. This confirms that the observed changes are driven by ATPase-dependent remodeling. Whether loss of positioning at promoters is functionally relevant is something that needs to be further tested. Coupling an Snf2h-reponsive promoter with a reporter and being able to monitor nucleosome repositioning would allow us to address that question. To this aim, we would have to (1) identify Snf2h-responsive promoters that exhibit a nucleosomal shift in their NRF region upon loss of Snf2h. This identification will have to be performed on a single locus level. We would then need to (2) verify that the respective promoter is bound by Snf2h. Snf2h location maps are not available or of very poor quality. We have put efforts in mapping Snf2h in mouse ES cells, with antibody-ChIP and tag-ChIP approaches. Both methods did not yield significant enrichments for Snf2h (data not shown). The inability of chromatin remodelers to be enriched in standard ChIP-seq assays has been well described in the field. We will further explore other methods, such as MNase-ChIP approaches (de Dieuleveult et al., 2016), to

determine genomic location of Snf2h. Once we identify Snf2h-responsive promoters that are both bound by Snf2h and exhibit nucleosome repositioning, we would (3) couple the promoter to a reporter gene and ask if the loss or gain of reporter activation is Snf2h dependent. This would enable us to link nucleosome positioning by Snf2h with transcriptional outcomes.

During the course of this project, a study was published describing an effect of Snf2h shRNA knockdown in a human cancer cell line (Wiechens et al., 2016). In their study, the authors reported no change in nucleosome positioning in promoter regions. For this positioning phenotype to be manifested, the Snf2h protein likely needs to be fully depleted. It is likely that the knockdown of Snf2h in their study was insufficient to observe those changes. They do, however, observe loss of positioning around binding sites for the transcription factor CTCF (discussed in the following chapter).

4.2.2 Role of Snf2h in Transcription Factor Binding

Several studies described strong nucleosome positioning in proximity to transcription factor binding motifs in mammals (Teif et al., 2012; Valouev et al., 2011; Wiechens et al., 2016). However, factors responsible for maintenance of that positioning have not been described. Here, we showed that Snf2h is required for a nucleosome positioning pattern in proximity to numerous transcription factors expressed in mouse ES cells. The strongest loss of positioning in Snf2h depleted cells is found in proximity to CTCF motifs. A physical and functional interaction of Snf2h and CTCF has been previously reported, but not extensively characterized (Dluhosova et al., 2014). Our results suggest that Snf2h is required for binding of CTCF. Indeed, we observe both global loss of CTCF binding in the *Snf2h ko* cells and increase in nucleosome occupancy over CTCF motifs. Reduction of CTCF binding caused by loss of Snf2h might have functional consequences. CTCF is known to contribute to the three-dimensional folding of chromosomes and to influence the structure of topologically associated domains (TADs; Bouwman & de Laat, 2015). A TAD is a region

of the genome where interactions of DNA within the region occur relatively frequently (Bouwman & de Laat, 2015). In contrast, inter-TAD interactions occur less frequently. Interestingly, CTCF is often found at the TAD boundaries (Bouwman & de Laat, 2015). In fact, it was reported that knockdown of CTCF disrupts the TAD structures and increases inter-TAD contacts (Zuin et al., 2014). Disruption of individual CTCF binding sites was reported to change expression of genes adjacent to the TAD boundary and to the respective CTCF site (Nora et al., 2012). To determine if the reduction of CTCF binding caused by loss of Snf2h has functional consequences, we want to determine changes of chromosomal interactions in the *Snf2h ko* cells. We would be using one of the chromosome conformation capture methods (Van Berkum et al., 2010). It might enable us to link transcriptional changes observed upon loss of Snf2h to changes in chromosome conformations through loss of CTCF. This is of particular interest since it is not fully understood how a modulated decrease of CTCF binding impacts organization of TADs.

It has been reported, in both yeast and mammals, that certain transcription factors require chromatin remodelers for binding (Ho et al., 2011; Ooi et al., 2006). In fact, CTCF does not seem to be the only factor that requires Snf2h for binding. We identified other factors that follow the same trend, such as Sp4 and E2f1. It appears that Snf2h might rather have a general role in enabling transcription-factor binding by maintaining nucleosome positions at a distance from transcription factor motifs. Our data is compatible with the model that in the absence of Snf2h, motif-bound nucleosomes cannot be repositioned which impairs binding of certain transcription factors. We would expect, that Snf2h is recruited by these transcription factors. To answer this crucial question is rather challenging, as we would expect the interaction to be mediated through non-catalytic subunits of ISWI complexes. It would require investigating physical interactions of the given transcription factor and a variety of ISWI-related subunits found in complexes with Snf2h. Purification of Snf2h complexes has provided limited evidence for interaction with transcription factors arguing that such

interactions are likely of transient nature. Taken together, our results suggest that Snf2h enables transcription factors binding by nucleosome repositioning.

Historically, Snf2h was considered to be a repressor. In *in vitro* experiments Snf2h was shown to assemble randomly positioned nucleosomes into an equally-spaced array (referred to as a nucleosome array; Clapier & Cairns, 2009; D. F V Corona & Tamkun, 2004). Our findings do not necessarily go against the previously reported *in vitro* observations. Once transcription factors such as CTCF would recruit Snf2h to enable its binding, Snf2h would maintain the nucleosomal array in proximity to the binding site. This maintenance of nucleosomal arrays in the motif proximity could ensure that the nucleosome does not invade the motif. For example, spontaneous intrusion of the motif-adjacent nucleosome into the binding site would require repositioning of each following downstream or upstream nucleosome to maintain the same nucleosomal distance. It is therefore energetically favorable to pull back the motif-adjacent nucleosome away from the motif to its original location. In the absence of Snf2h, the array structure might not be preserved and the nucleosomes could occupy the binding site more frequently. To be able to test this hypothesis, we would need to have a readout for nucleosome positioning and transcription factor binding on the same molecule. NOMe-Seq is a high resolution single-molecule method that provides both a nucleosome and a transcription factor footprint (Kelly et al., 2012). In the NOMe-seq approach, chromatin is treated with a GpC methyltransferase enzyme which exogenously methylates GpC dinucleotides that are not protected by nucleosomes or transcription factors. The GpC methylation information is then used to infer occupancy on chromatin. NOMe-seq would therefore be a suitable approach to address this question. In summary, we showed that Snf2h is required for nucleosome positioning and transcription factor binding in mouse ES cells. The precise mechanism of how nucleosome positioning by ISWI enables transcription factor binding requires further investigation.

Chapter 5
Material and Methods

5.1 Cell Culture

In the study, we used wild type mouse ES cell lines of 129S6/SvEvTac background. We cultured mouse ES cells as previously described (Mohn et al., 2008). Briefly, cells were maintained in Dulbecco's Modified Eagle Medium (Invitrogen), supplemented with 15 % Fetal Calf Serum (Invitrogen), L-Glutamine (Gibco) and Non-essential amino acids (Gibco), beta-mercaptoethanol (Sigma) and homemade leukemia inhibitory factor (LIF). Differentiation of ES cells to neuronal progenitors was performed as previously described (Bibel et al., 2004; Mohn et al., 2008), a protocol known to yield a pure population of Pax6-positive radial-glial neuronal progenitor and terminally differentiated glutamatergic pyramidal neurons (Bibel et al., 2004). ES cells were trypsinized with 0.05% Trypsin-EDTA (Gibco), neutralized with cell culture medium and pelleted prior to plating. Cell were grown on plates coated with 0.2 % gelatin, with the exception of *Snf2h ko*. *Snf2h ko* cells were grown on feeders (inactivated mouse fibroblasts) for at least two passages after thawing. Cells were passaged every second day and medium was exchanged every day. Every cell line, except *Snf2 ko*, was diluted prior to plating in the 1/5 ratio. *Snf2h ko* cells were diluted 1/2.

5.2 Generation of Deletion Cell Lines

Eed ko and *Snf2h ko* cell lines were generated using the CRISPR/Cas9 protocols previously described with modifications (Cong et al., 2013). Briefly, mouse 159 ES cells were co-transfected (Lipofectamine 2000 and 3000, Thermo Fisher Scientific) with plasmids expressing mammalian-

codon optimized Cas9, sgRNAs targeting the region of interest (regions flanking exon 1 of *Eed* or a region within exon 6 of *Smarca5*) and a plasmid expressing resistance against puromycin. Puromycin selection (2 µg/ml) was carried out one day after transfection for 24h. Following a 3-5 day recovery, individual colonies were genotyped by western blot (H3k27me3 or Snf2h).

5.3 Generation of Re-Expression Cell Lines

Codon-optimized murine wt or ATPase mutant Snf2h cDNA was cloned into the pCDNA6-CAG-V5-MCS-IRES-Blasticidin plasmid (Hakimi et al., 2002; Khavari et al., 1993). It was stably expressed in the *Snf2h ko* cells. Integration in the genome was not directed (random integration). Briefly, 40 µg of the respective plasmid was linearized with FspI (NEB), precipitated with ethanol and resuspended in TE buffer. 4 mio. cells were electroporated (Amaxa Nucleofection, Lonza). Selection of cells was started two days following the transfection. 5ug/ml Blasticidin (Invivogen) was used for 10-days. Individual clones were expanded and tested for expression by western blot probing with either α-V5 (R960-25, Thermo Fisher Scientific) or α-Snf2h (ab72499) antibodies.

5.4 Library ChIP-seq Method

384 or 96 pairs of primers were *in silico* designed to target regions in the mouse and *E. coli* genome, respectively. Following the PCR amplification (KAPA, KAPA Biosystems), the products were pooled. For targeted insertion of the Polycomb and *E. coli* libraries, DNA fragments were cloned into a plasmid containing a multiple cloning site flanked by priming regions for a pair of universal primers and two inverted L1 Lox sites (pL1-LPP1-1L). Libraries were inserted using the Recombinase-mediated Cassette Exchange (RMCE) insertion protocol with modifications (Lienert et al., 2011). ES cells were selected under hygromycin (250 µg/ml, Roche) for 10

days. 12 mio. cells were electroporated (Amaxa Nucleofection, Lonza). 75 µg of L1-library-1L library plasmid and 45 µg of pIC-Cre plasimd was used. Negative selection with 3 µM Ganciclovir (Roche, Switzerland) was started 2 days post-transfection and was carried out for 10 days. Library pools were genotyped for insertion by PCR using universal primers that target the flanking region of the insertion site. PCR amplification of the insertion site was carried out with low-bias amplification KAPA Hi-Fi Polymerase as described previously (C. D. Arnold et al., 2013). Sequencing libraries were prepared using only 4 cycles of PCR amplification (NEBNext ChIP-Seq Library Prep). Libraries were pooled and sequences on the Illumina MiSeq platform using MiSeq Reagent Kit v2 (50 cycles were used).

5.5 Library BIS-seq Method

Genomic DNA (2 µg) of ES cell libraries with inserted fragments was submitted to bisulfite conversion using the EpiTec Bisulfite Kit (Qiagen). Libraries were amplified by PCR (AmpliTaq Gold Life Technologies). Bisulfite compatible primers that target the flanking region of the insertion site were used. The PCR product was column purified (PCR Purification Kit, Qiagen). Sequencing libraries were prepared using 12 cycles of PCR amplification (NEBNext ChIP-Seq Library Prep). Libraries were pooled and sequenced on the Illumina MiSeq platform using MiSeq Reagent Kit v3 (150 cycles were used).

5.6 Western Blot Analysis

Detection of H3K27me3 protein levels was performed with the histone extraction protocol. Briefly, cells were harvested and washed twice with ice-cold PBS. Cells were resuspended in Triton Extraction Buffer (TEB: PBS containing 0.5% Triton X 100 (v/v), 2 mM phenylmethylsulfonyl fluoride (PMSF), 0.02% (w/v) NaN3) at a cell density of 10 mio. cells per ml. Cells were lysed on ice for 10 min. with gentle stirring. Lysis solution was

centrifuged at 6,500 x g for 10 min. at 4°C to spin down the nuclei. The nuclei were washed in half the volume of TEB and centrifuged as before. The pellet was resuspended in 0.2 N HCl at a density of 40 mio. nuclei per ml. Histones were extracted for 30 min. on ice. Protein content was determined using the Bradford assay. Detection of Snf2h, V5 and CTCF levels were performed with the whole-cell lysate protocol previously described (Baubec et al., 2013). The membrane was probed with mouse α-V5 (R960-25, Thermo Fisher Scientific), α-Snf2h (ab72499), α-CTCF (Santa Cruz Biotechnology, C-20) or α-H3K27me3 (Millipore,07-449) in combination with appropriate secondary antibodies.

5.7 Chromatin Immunoprecipitation Sequenicng

ChIP was carried out as previously described (Weber et al., 2007) with slight modifications. Changes to the protocol were following; (1) chromatin was sonicated for ~50 cycles of 30 sec. using a Diagenode Bioruptor, with 45 sec. breaks in between cycles, (2) protein A/G magnetic Dynabeads Magnetic beads (Thermo Fisher Scientific) were used, (3) DNA was purified using on-column purification instead of chloroform/phenol extraction (Qiagen PCR Purification Kit). Immunoprecipitated DNA and input DNA were submitted to library preparation (NEBNext Ultra DNA Library Prep Kit, Illumina). In the library preparation protocol, input sample were amplified using 4 PCR cycles, IP sample using 12 cycles. Antibodies used were α-H3K27me3 (Millipore,07-449), α-REST (Santa Cruz, H-290), α-H3K4me3 (Millipore, 17- 614), α-Snf2h (ab72499), α-CTCF (Santa Cruz Biotechnology, C-20), α-Suz12 (Cell Signaling, 3737).

5.8 RNA Isolation and Sequencing

Total RNA-seq was performed for three biological replicates using RNeasy Mini Kit (Qiagen) for all samples in the study. Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat; Illumina) was used to remove ribosomal RNA. Data alignment and analysis was performed within QuasR. RNA abundance was quantified as read per kilobase of exon per million mapped reads (rpkm).

5.9 ATAC Sequencing

ATAC-seq was performed according to the previously described protocol (Buenrostro et al., 2015) with modifications. Briefly, 50,000 cells were washed with cold PBS and resuspended in lysis buffer to extract nuclei. The nuclei were cold-centrifuged at 500 x g for 10 min. The nuclei pellet was incubated with transposition reaction buffer for 30 min. at 37 °C. The DNA was purified using the PCR Purification Kit (Qiagen). The eluted transposed DNA was submitted to library preparation (NEBUltra Library Preparation Kit, NEB). DNA was amplified with 12 cycles of PCR. Amplification saturation was monitored with real time PCR. The optimal amount of additional PCR cycles was determined based on the real time analysis. The optimal number of cycles was defined as one-third of maximum fluorescence intensity in the real time reaction after performing 20 PCR cycles. 12 cycles were sufficient to yield the necessary amount of DNA and not to reach saturation. The libraries were sequenced on the Illumina HiSeq platform. 50 cycles paired-end sequencing was used.

5.10 MNase Sequencing

A custom MNase protocol was used combining previously published methods (Cui & Zhao, 2012; Gaidatzis et al., 2014). Cells were resuspended in 1 mL of Buffer 1 w/ detergent (Buffer 1 + 0.02% NP-40) and incubated on

ice for 5 min. Nuclei were then pelleted at 300 x g for 5 min at 4ºC. Nuclei were gently resuspended in 1 mL of Buffer 2. Nuclei integrity was checked at this point using a cell analyzer. Clumping of nuclei indicates nuclei has lysed. We made sure to recover no less than 75% of nuclei compared to the starting cell number. Nuclei were then pelleted for 5 min. at 300 x g at 4ºC. Pallets were resuspnded in 400 µl MNase buffer (w/o MNase). Varying concentration of MNase (Roche) was added. Optional concentration was later determined to be 5U. Nuclei were then incubated for 30 or 60 min. at 37ºC. Reaction was stopped by adding 4 µl of EDTA at 5 mM (0.5M solution). 20 µl of SDS at 1% (20% stock solution) and 8 µl of Proteinase K at 200 µg/mL (10 mg/mL stock concentration) was added to the sample. Samples were incubated at 55ºC for 1h with shaking. DNA was purified using PCR Purification Kit (Qiagen). Purified DNA was loaded on a 1% agarose gel or on the Agilent Bioanalyzer to determine digestion efficiency. Mononculesomal fraction was isolated using AMPure XP beads (1/1 radio, Beckman Coulter). Libraries were prepared with NEBUltra Library Preparation Kit (NEB), using 4 PCR cycles and 1 µg of starting DNA. 50 cycles HiSeq Illumina sequencing was performed. Buffer 1: 0.3M Sucrose, 15mM Tris pH 7.5, 60mM KCl, 15mM NaCl, 5mM MgCl2, 2mM EDTA, 0.5mM DTT, 1x PIC, 0.2mM spermine, 1mM spermidine. MNase Buffer: 0.3M Sucrose, 50mM Tris ph 7.5, 4mM MgCl2, 1mM CaCl2, 1 x PIC + additional of 5U of MNase S7 micrococcal nuclease (Roche). Buffer 2 (Buffer 1 w/o EDTA): 0.3M Sucrose, 15mM Tris pH 7.5, 60mM KCl, 15mM NaCl, 5mM MgCl2, 0.5mM DTT*, 1 x PIC, 0.2mM spermine, 1mM spermidine.

5.11 Data Processing

Bowtie v1.12.0 was used for aligning the non-bisulfite reads in all experiments except library BIS-seq. ATAC-seq, MNase-seq, RNA-seq and ChIP-seq analysis were performed using R version 3.3.0 and the Bioconductor package QuasR 1.12.0 (Gaidatzis, Lerch, Hahne, & Stadler,

2015). In brief, FASTQ files were aligned within QuasR using Bowtie v1.12.0 (parameters: -m 1 --best --strata) against mm9 genome assembly. Quality control analysis was performed within QuasR. Reads were shifted by 75 bp (MNase) or 100 bp (ChIP). For CTCF ChIP-seq, reads were quantified in 250 bp windows over predicted CTCF motifs and normalized to the mapped library size. Peaks were called with MACS-1.4.2 using standard parameters.

Bismark/Bowtie 0.12.7 (Langmead et al., 2009; Krueger and Andrews, 2011) were used to align bisulfite reads against an *in silico* converted reference genome (C > T and G > A). Methylation was called per CG and fragment averages were derived using the previously established reference set of regions for the library.

List of Abbreviations

AA	Amino acid
ACF	ATP-utilizing chromatin assembly and remodeling factor
ATAC-SEQ	Assay for transposase-accessible chromatin with high throughput sequencing
ATP	Adenosine triphosphate
BAF	BRG1- or HBRM-associated factors
CENP-A	Centromere protein A
CHD	Chromodomain, helicase, DNA binding
CHIP-SEQ	Chromatin immunoprecipitation followed by sequencing
CHRAC	Chromatin remodeling and assembly complex
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
ES	Embryonic stem (cell)
EZH	Enhancer of zeste homologue
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HOX	Homeotic complex (genes)
INO80	Inositol requiring 80
ISWI	Imitation switch
LIF	Leukemia inhibitory factor
MBD	Methyl-binding domain
MNASE-SEQ	Micrococcal nuclease sensitive sites sequencing
NFR	Nucleosome free region
NRF1	Nuclear respiratory factor 1

NOME-SEQ	Nucleosome occupancy methylome with high throughout sequencing
NORC	Nucleolar remodeling complex
NURD	Nucleosome remodeling deacetylase
NURF	Nucleosome remodelling factor
PCG	Polycomb group
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
POL	Polymerase
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PRE	Polycomb reponse element
PWM	Position weight matrix
QUASR	Quantify and annotate short reads in R
RNA-SEQ	RNA sequencing
REST	RE1-silencing transcription factor
RSC	Remodeling Structure Chromatin (complex)
SUZ	Suppressor of zeste
SWI/SNF	Switching defective/Sucrose nonfermenting
TBP	TATA-binding protein
TET	Ten-eleven Translocation
TSS	Transcription start site
WICH	WSTF-ISWI chromatin remodeling

.

References

Adkins, M. W., & Tyler, J. K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Molecular Cell*, *21*(3), 405–416. https://doi.org/10.1016/j.molcel.2005.12.010

Alberts, B. (2008). Molecular Biology of the Cell. *The Yale Journal of Biology and Medicine*, 1601. https://doi.org/10.1024/0301-1526.32.1.54

Allfrey, V., Faulkner, R., & Mirsky, A. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences*, *315*(1938), 786–794. https://doi.org/10.1073/pnas.51.5.786

Allis, C. D. (2007). *Epigenetics*. *New York*.

Andy Choo, K. H. (2001). Domain Organization at the Centromere and Neocentromere. *Developmental Cell*. https://doi.org/10.1016/S1534-5807(01)00028-4

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M., Stark, A., … Stark, A. (2013). Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science*, *339*(6123), 1074–1077. https://doi.org/10.1126/science.1232542

Arnold, P., Schöler, A., Pachkov, M., Balwierz, P. J., Jørgensen, H., Stadler, M. B., … Schuẗeler, D. (2013). Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome Research*, *23*(1), 60–73. https://doi.org/10.1101/gr.142661.112

Arvey, A., Agius, P., Noble, W. W. S., & Leslie, C. (2012). Sequence and chromatin determinants of cell-type – specific transcription factor binding. *Genome Research*, *22*(9), 1723–1734. https://doi.org/10.1101/gr.127712.111.

Atchison, L., Ghias, A., Wilkinson, F., Bonini, N., & Atchison, M. L. (2003). Transcription factor YY1 functions as a PcG protein in vivo. *EMBO*

Journal, *22*(6), 1347–1358. https://doi.org/10.1093/emboj/cdg124

Bai, L., & Morozov, A. V. (2010). Gene regulation by nucleosome positioning. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2010.08.003

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Barlev, N. A., Liu, L., Chehab, N. H., Mansfield, K., Harris, K. G., Halazonetis, T. D., & Berger, S. L. (2001). Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. *Molecular Cell*, *8*(6), 1243–1254. https://doi.org/10.1016/S1097-2765(01)00414-2

Barutcu, A. R., Lajoie, B. R., Fritz, A. J., McCord, R. P., Nickerson, J. A., Van Wijnen, A. J., … Imbalzano, A. N. (2016). SMARCA4 regulates gene expression and higherorder chromatin structure in proliferating mammary epithelial cells. *Genome Research*, *26*(9), 1188–1201. https://doi.org/10.1101/gr.201624.115

Baubec, T., Ivánek, R., Lienert, F., & Schübeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the mbd protein family. *Cell*, *153*(2), 480–492. https://doi.org/10.1016/j.cell.2013.03.011

Bauer, M., Trupke, J., & Ringrose, L. (2016). The quest for mammalian Polycomb response elements: are we there yet? *Chromosoma*. https://doi.org/10.1007/s00412-015-0539-4

Beisel, C., & Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. *Nature Reviews Genetics*, *12*(2), 123–135. https://doi.org/10.1038/nrg2932

Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., … Barde, Y. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nature Neuroscience*, *7*(9), 1003–1009. https://doi.org/10.1038/nn1301

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes*

and Development. https://doi.org/10.1101/gad.947102

Bird, A. P., & Wolffe, A. P. (1999). Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell*, *99*(5), 451–454. https://doi.org/10.1016/S0092-8674(00)81532-9

Bouwman, B. A. M., & de Laat, W. (2015). Getting the genome in shape: the formation of loops, domains and compartments. *Genome Biology*, *16*, 154. https://doi.org/10.1186/s13059-015-0730-1

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., … Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, *441*(7091), 349–53. https://doi.org/10.1038/nature04733

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., … Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, *478*(7369), 343–348. https://doi.org/10.1038/nature10532

Brinkman, A. B., Gu, H., Bartels, S. J. J., Zhang, Y., Matarese, F., Simmer, F., … Stunnenberg, H. G. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Research*, *22*(6), 1128–1138. https://doi.org/10.1101/gr.133728.111

Brogaard, K., Xi, L., Wang, J.-P., & Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. *Nature*, *486*(7404), 496–501. https://doi.org/10.1038/nature11142

Brown, E., Malakar, S., & Krebs, J. E. (2007). How many remodelers does it take to make a brain? Diverse and cooperative roles of ATP-dependent chromatin-remodeling complexes in development. *Biochemistry and Cell Biology*, *85*(4), 444–62. https://doi.org/10.1139/O07-059

Bryant, G. O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., & Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. *PLoS Biology*, *6*(12), 2928–39. https://doi.org/10.1371/journal.pbio.0060317

Bryja, V., Bonilla, S., & Arenas, E. (2006). Derivation of mouse embryonic stem cells. *Nature Protocols*, *1*(4), 2082–2087. https://doi.org/10.1038/nprot.2006.355

Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, *2015*, 21.29.1-21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., … Magnuson, T. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Molecular Cell*, *6*(6), 1287–1295. https://doi.org/10.1016/S1097-2765(00)00127-1

Burgio, G., La Rocca, G., Sala, A., Arancio, W., Di Gesù, D., Collesano, M., … Corona, D. F. V. (2008). Genetic identification of a network of factors that functionally interact with the nucleosome remodeling ATPase ISWI. *PLoS Genetics*, *4*(6). https://doi.org/10.1371/journal.pgen.1000089

Cairns, B. R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature*, *461*(7261), 193–198. https://doi.org/10.1038/nature08450

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., … Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science (New York, N.Y.)*, *298*(5595), 1039–43. https://doi.org/10.1126/science.1076997

Carey, M., Li, B., & Workman, J. L. (2006). RSC Exploits Histone Acetylation to Abrogate the Nucleosomal Block to RNA Polymerase II Elongation. *Molecular Cell*, *24*(3), 481–487. https://doi.org/10.1016/j.molcel.2006.09.012

Chamberlain, S. J., Yee, D., & Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells*, *26*(6), 1496–505. https://doi.org/10.1634/stemcells.2008-0102

Chi, T. H., Wan, M., Zhao, K., Taniuchi, I., Chen, L., Littman, D. R., & Crabtree, G. R. (2002). Reciprocal regulation of CD4/CD8 expression by SWI/SNF-like BAF complexes. *Nature*, *418*(6894), 195–199. https://doi.org/10.1038/nature00876\rnature00876 [pii]

Clapier, C. R., & Cairns, B. R. (2009). The Biology of Chromatin Remodelling Complexes. *Annual Review of Biochemistry*, *78*(73), 273–304. https://doi.org/10.1146/annurev.biochem.77.062706.153223

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., … Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas System. *Science*, *339*(February), 819–824. https://doi.org/10.1126/science.1231143 RNA-Guided

Corona, D. F. V, Längst, G., Clapier, C. R., Bonte, E. J., Ferrari, S., Tamkun, J. W., & Becker, P. B. (1999). ISWI is an ATP-dependent nucleosome remodeling factor. *Molecular Cell*, *3*(2), 239–245. https://doi.org/10.1016/S1097-2765(00)80314-7

Corona, D. F. V, & Tamkun, J. W. (2004). Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochimica et Biophysica Acta - Gene Structure and Expression*. https://doi.org/10.1016/j.bbaexp.2003.09.018

Cui, K., & Zhao, K. (2012). Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods in Molecular Biology*, *833*, 413–419. https://doi.org/10.1007/978-1-61779-477-3_24

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., & Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, *111*(2), 185–196. https://doi.org/10.1016/S0092-8674(02)00975-3

Dang, W., & Bartholomew, B. (2007). Domain architecture of the catalytic subunit in the ISW2-nucleosome complex. *Molecular and Cellular Biology*, *27*(23), 8306–8317. https://doi.org/10.1128/MCB.01351-07

de Dieuleveult, M., Yen, K., Hmitou, I., Depaux, A., Boussouar, F.,

Dargham, D. B., … Gérard, M. (2016). Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature*, *530*(7588), 113–6. https://doi.org/10.1038/nature16505

Deal, R. B., Henikoff, J. G., & Henikoff, S. (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science (New York, N.Y.)*, *328*(5982), 1161–4. https://doi.org/10.1126/science.1186777

de la Serna, I. L., Ohkawa, Y., & Imbalzano, A. N. (2006). Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nature Reviews. Genetics*, *7*(6), 461–73. https://doi.org/10.1038/nrg1882

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, *25*(10), 1010–22. https://doi.org/10.1101/gad.2037511

Dellino, G. I., Schwartz, Y. B., Farkas, G., McCabe, D., Elgin, S. C. R., & Pirrotta, V. (2004). Polycomb silencing blocks transcription initiation. *Molecular Cell*, *13*(6), 887–893. https://doi.org/10.1016/S1097-2765(04)00128-5

Di Croce, L., & Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. *Nature Structural & Molecular Biology*, *20*(10), 1147–55. https://doi.org/10.1038/nsmb.2669

Dion, M. F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., & Rando, O. J. (2007). Dynamics of replication-independent histone turnover in budding yeast. *Science*, *315*(5817), 1405–8. https://doi.org/10.1126/science.1134053

Dluhosova, M., Curik, N., Vargova, J., Jonasova, A., Zikmund, T., & Stopka, T. (2014). Epigenetic control of SPI1 gene by CTCF and ISWI ATPase SMARCA5. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0087448

Domcke, S., Bardet, A. F., Adrian Ginno, P., Hartl, D., Burger, L., & Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*, *528*(7583),

575–579. https://doi.org/10.1038/nature16462

Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., … Bickmore, W. A. (2010). Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Molecular Cell*, *38*(3), 452–464. https://doi.org/10.1016/j.molcel.2010.02.032

Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., … Klose, R. J. (2012). KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *eLife*, *2012*(1). https://doi.org/10.7554/eLife.00205

Faust, C., Schumacher, A., Holdener, B., & Magnuson, T. (1995). The eed mutation disrupts anterior mesoderm production in mice. *Development*, *121*(2), 273–285.

Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., … van Steensel, B. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell*, *143*(2), 212–224. https://doi.org/10.1016/j.cell.2010.09.009

Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., & Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. *Genes and Development*, *17*(15), 1870–1881. https://doi.org/10.1101/gad.1110503

Fromm, G., & Bulger, M. (2009). A spectrum of gene regulatory phenomena at mammalian beta-globin gene loci. *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire*, *87*(5), 781–790. https://doi.org/10.1139/O09-048

Fry, C. J., & Peterson, C. L. (2001). Chromatin remodeling enzymes: Who's on first? *Current Biology*. https://doi.org/10.1016/S0960-9822(01)00090-2

Gaidatzis, D., Burger, L., Murr, R., Lerch, A., Dessus-Babus, S., Schübeler, D., & Stadler, M. B. (2014). DNA Sequence Explains Seemingly Disordered Methylation Levels in Partially Methylated Domains of

Mammalian Genomes. *PLoS Genetics*, *10*(2). https://doi.org/10.1371/journal.pgen.1004143

Gaidatzis, D., Lerch, A., Hahne, F., & Stadler, M. B. (2015). QuasR: Quantification and annotation of short reads in R. *Bioinformatics*, *31*(7), 1130–1132. https://doi.org/10.1093/bioinformatics/btu781

Gil, J., & O'Loghlen, A. (2014). PRC1 complex diversity: Where is it taking us? *Trends in Cell Biology*. https://doi.org/10.1016/j.tcb.2014.06.005

Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu Rev Biochem*, *74*, 481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721

Graf, T., & Enver, T. (2009). Forcing cells to change lineages. *Nature*, *462*(7273), 587–594. https://doi.org/10.1038/nature08533

Grau, D. J., Chapman, B. A., Garlick, J. D., Borowsky, M., Francis, N. J., & Kingston, R. E. (2011). Compaction of chromatin by diverse polycomb group proteins requires localized regions of high charge. *Genes and Development*, *25*(20), 2210–2221. https://doi.org/10.1101/gad.17288211

Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B., Kullman, K., Leitch, I. J., … Bennett, M. D. (2007). Eukaryotic genome size databases. *Nucleic Acids Research*, *35*(SUPPL. 1). https://doi.org/10.1093/nar/gkl828

Grüne, T., Brzeski, J., Eberharter, A., Clapier, C. R., Corona, D. F. V, Becker, P. B., & Müller, C. W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Molecular Cell*, *12*(2), 449–460. https://doi.org/10.1016/S1097-2765(03)00273-9

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. *Nature*, *389*(6649), 349–352. https://doi.org/10.1038/38664

Hakimi, M.-A., Bochar, D. a, Schmiesing, J. a, Dong, Y., Barak, O. G., Speicher, D. W., … Shiekhattar, R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature*,

418(6901), 994–998. https://doi.org/10.1038/nature01024

Han, M., & Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in vivo. *Cell*, *55*(6), 1137–1145. https://doi.org/10.1016/0092-8674(88)90258-9

Han, M., Kim, U. J., Kayne, P., & Grunstein, M. (1988). Depletion of histone H4 and nucleosomes activates the PHO5 gene in Saccharomyces cerevisiae. *The EMBO Journal*, *7*(7), 2221–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=454566&tool=pmcentrez&rendertype=abstract

Han, Z., Xing, X., Hu, M., Zhang, Y., Liu, P., & Chai, J. (2007). Structural Basis of EZH2 Recognition by EED. *Structure*, *15*(10), 1306–1315. https://doi.org/10.1016/j.str.2007.08.007

Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., … Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. *Nature Cell Biology*, *10*(11), 1291–300. https://doi.org/10.1038/ncb1787

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., & Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, *405*(6785), 486–489. https://doi.org/10.1038/35013106

Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J., & Workman, J. L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell*, *111*(3), 369–379. https://doi.org/10.1016/S0092-8674(02)01005-X

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., & Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nature Cell Biology*, *15*(4), 373–84. https://doi.org/10.1038/ncb2702

Henikoff, S., & Ahmad, K. (2005). Assembly of Variant Histones into Chromatin. *Annual Review of Cell and Developmental Biology*, *21*(1), 133–153. https://doi.org/10.1146/annurev.cellbio.21.012704.133518

Ho, L., & Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature*, *463*(7280), 474–84. https://doi.org/10.1038/nature08911

Ho, L., Miller, E. L., Ronan, J. L., Ho, W. Q., Jothi, R., & Crabtree, G. R. (2011). esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nature Cell Biology*, *13*(8), 903–13. https://doi.org/10.1038/ncb2285

Ho, L., Ronan, J. L., Wu, J., Staahl, B. T., Chen, L., Kuo, A., … Crabtree, G. R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(13), 5181–6. https://doi.org/10.1073/pnas.0812889106

Hoffman, M. M., Ernst, J., Wilder, S. P., Kundaje, A., Harris, R. S., Libbrecht, M., … Noble, W. S. (2013). Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Research*, *41*(2), 827–841. https://doi.org/10.1093/nar/gks1284

Huang, H., Maertens, A. M., Hyland, E. M., Dai, J., Norris, A., Boeke, J. D., & Bader, J. S. (2009). HistoneHits: A database for histone mutations and their phenotypes. *Genome Research*, *19*(4), 674–681. https://doi.org/10.1101/gr.083402.108

Hughes, A. L., Jin, Y., Rando, O. J., & Struhl, K. (2012). A Functional Evolutionary Approach to Identify Determinants of Nucleosome Positioning: A Unifying Model for Establishing the Genome-wide Pattern. *Molecular Cell*, *48*(1), 5–15. https://doi.org/10.1016/j.molcel.2012.07.003

Jermann, P., Hoerner, L., Burger, L., & Schübeler, D. (2014). Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), E3415-21. https://doi.org/10.1073/pnas.1400672111

Jiang, C., & Pugh, B. F. (2009). Nucleosome positioning and gene

regulation: advances through genomics. *Nature Reviews. Genetics*, *10*(3), 161–72. https://doi.org/10.1038/nrg2522

Kadoch, C., & Crabtree, G. R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Science Advances*, *1*(5), e1500447–e1500447. https://doi.org/10.1126/sciadv.1500447

Kasten, M., Szerlong, H., Erdjument-Bromage, H., Tempst, P., Werner, M., & Cairns, B. R. (2004). Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. *The EMBO Journal*, *23*(6), 1348–1359. https://doi.org/10.1038/sj.emboj.7600143

Kehle, J., Beuchle, D., Treuheit, S., Christen, B., Kennison, J. a, Bienz, M., & Müller, J. (1998). dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science (New York, N.Y.)*, *282*(December), 1897–1900. https://doi.org/10.1126/science.282.5395.1897

Kelly, T. K., Liu, Y., Lay, F. D., Liang, G., Berman, B. P., & Jones, P. A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Research*, *22*(12), 2497–2506. https://doi.org/10.1101/gr.143008.112

Khavari, P. a, Peterson, C. L., Tamkun, J. W., Mendel, D. B., & Crabtree, G. R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature*, *366*, 170–174. https://doi.org/10.1038/366170a0

Kidder, B. L., Palmer, S., & Knott, J. G. (2009). SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells (Dayton, Ohio)*, *27*(2), 317–28. https://doi.org/10.1634/stemcells.2008-0710

Kim, J. K., Huh, S. O., Choi, H., Lee, K. S., Shin, D., Lee, C., … Seong, R. H. (2001). Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. *Molecular and Cellular Biology*, *21*(22), 7787–95. https://doi.org/10.1128/MCB.21.22.7787-7795.2001

Klochendler-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C., & Yaniv, M. (2000). The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. *EMBO Reports*, *1*(6), 500–506. https://doi.org/10.1093/embo-reports/kvd129

Knezetic, J. A., & Luse, D. S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell*, *45*(1), 95–104. https://doi.org/10.1016/0092-8674(86)90541-6

Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell*. https://doi.org/10.1016/j.cell.2007.02.005

Krebs, A. R., Dessus-Babus, S., Burger, L., & Sch??beler, D. (2014). High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife*, *3*, e04094. https://doi.org/10.7554/eLife.04094

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., … Bernstein, B. E. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genetics*, *4*(10). https://doi.org/10.1371/journal.pgen.1000242

Kwong, C., Adryan, B., Bell, I., Meadows, L., Russell, S., Manak, J. R., & White, R. (2008). Stability and dynamics of polycomb target sites in Drosophila development. *PLoS Genetics*, *4*(9). https://doi.org/10.1371/journal.pgen.1000178

Lai, A. Y., & Wade, P. A. (2011). NuRD: A multi-faceted chromatin remodeling complex in regulating cancer biology. *Nature Reviews. Cancer*, *11*(8), 588–596. https://doi.org/10.1038/nrc3091

Landeira, D., & Fisher, A. G. (2011). Inactive yet indispensable: The tale of Jarid2. *Trends in Cell Biology*. https://doi.org/10.1016/j.tcb.2010.10.004

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., … Szustakowki, J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Längst, G., & Becker, P. B. (2001). Nucleosome mobilization and

positioning by ISWI-containing chromatin-remodeling factors. *Journal of Cell Science*, *114*(Pt 14), 2561–2568.

Le Guezennec, X., Vermeulen, M., Brinkman, A. B., Hoeijmakers, W. A. M., Cohen, A., Lasonder, E., & Stunnenberg, H. G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Molecular and Cellular Biology*, *26*(3), 843–51. https://doi.org/10.1128/MCB.26.3.843-851.2006

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., & Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes and Development*, *24*(3), 265–276. https://doi.org/10.1101/gad.544410

Lennox, R. W., & Cohen, L. H. (1988). The production of tissue-specific histone complements during development. *Biochem Cell Biol*, *66*(6), 636–649. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3048335

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature*, *276*(5688), 565–70. https://doi.org/10.1128/AAC.03728-14

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., & Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genetics*, *43*(11), 1091–1097. https://doi.org/10.1038/ng.946

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., … Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–22. https://doi.org/10.1038/nature08514

Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, *469*(7330), 343–9. https://doi.org/10.1038/nature09784

Martens, J. A., & Winston, F. (2002). Evidence that Swi/Snf directly represses transcription in S. cerevisiae. *Genes and Development*, *16*(17), 2231–2236. https://doi.org/10.1101/gad.1009902

Mavrich, T. N., Ioshikhes, I. P., Venters, B. J., Jiang, C., Tomsho, L. P., Qi, J., … Pugh, B. F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Research*, *18*(7), 1073–1083. https://doi.org/10.1101/gr.078261.108

McKeon, J., Slade, E., Sinclair, D. A. R., Cheng, N., Couling, M., & Brock, H. W. (1994). Mutations in some Polycomb group genes of Drosophila interfere with regulation of segmentation genes. *MGG Molecular & General Genetics*, *244*(5), 474–483. https://doi.org/10.1007/BF00583898

McKittrick, E., Gafken, P. R., Ahmad, K., & Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(6), 1525–30. https://doi.org/10.1073/pnas.0308092100

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., … Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–770. https://doi.org/10.1038/nature07107

Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., … Bernstein, B. E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genetics*, *6*(12), 1–10. https://doi.org/10.1371/journal.pgen.1001244

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., & Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science (New York, N.Y.)*, *303*(5656), 343–8. https://doi.org/10.1126/science.1090701

Mohn, F., & Schübeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2008.12.005

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., … Schübeler, D. (2008). Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. *Molecular Cell*, *30*(6), 755–766. https://doi.org/10.1016/j.molcel.2008.05.007

Morrison, A. J., & Shen, X. (2009). Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nature Reviews Molecular Cell Biology*, *10*(6), 373–384. https://doi.org/10.1038/nrm2693

Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., … Simon, J. A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*, *111*(2), 197–208. https://doi.org/10.1016/S0092-8674(02)00976-5

Müller, J., & Kassis, J. A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Current Opinion in Genetics & Development*, *16*(5), 476–484. https://doi.org/10.1016/j.gde.2006.08.005

Murray, K. (1964). The Occurrence of iε-N-Methyl Lysine in Histones. *Biochemistry*, *3*(1), 10–15. https://doi.org/10.1021/bi00889a003

Musladin, S., Krietenstein, N., Korber, P., & Barbaric, S. (2014). The RSC chromatin remodeling complex has a crucial role in the complete remodeler set for yeast PHO5 promoter opening. *Nucleic Acids Research*, *42*(7), 4270–4282. https://doi.org/10.1093/nar/gkt1395

Narlikar, G. J., Sundaramoorthy, R., & Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*. https://doi.org/10.1016/j.cell.2013.07.011

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., … Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*, 381–385. https://doi.org/10.1038/nature11049

Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A. K., Furlong, E. E. M., & Müller, J. (2008). Dynamic Regulation by Polycomb

Group Protein Complexes Controls Pattern Formation and the Cell Cycle in Drosophila. *Developmental Cell*, *15*(6), 877–889. https://doi.org/10.1016/j.devcel.2008.10.005

Ooi, L., Belyaev, N. D., Miyake, K., Wood, I. C., & Buckley, N. J. (2006). BRG1 chromatin remodeling activity is required for efficient chromatin binding by repressor element 1-silencing transcription factor (REST) and facilitates REST-mediated repression. *Journal of Biological Chemistry*, *281*(51), 38974–38980. https://doi.org/10.1074/jbc.M605370200

Oudet, P., Gross-Bellard, M., & Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, *4*(4), 281–300. https://doi.org/10.1016/0092-8674(75)90149-X

Parnell, T. J., Huff, J. T., & Cairns, B. R. (2008). RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. *The EMBO Journal*, *27*(1), 100–10. https://doi.org/10.1038/sj.emboj.7601946

Paro, R., & Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proceedings of the National Academy of Sciences*, *88*(1), 263–267. https://doi.org/10.1073/pnas.88.1.263

Pasini, D., Cloos, P. A. C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J. V, … Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature*, *464*(7286), 306–310. https://doi.org/10.1038/nature08788

Pearson, J. C., Lemons, D., & McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nature Reviews Genetics*, *6*(12), 893–904. https://doi.org/10.1038/nrg1726

Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., & Wysocka, J. (2009). Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. *Cell*, *139*(7), 1290–1302. https://doi.org/10.1016/j.cell.2009.12.002

Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A., & Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-

modifying factor Polycomb. *Science*, *339*(6120), 698–699. https://doi.org/10.1017/CBO9781107415324.004

Phuket, B. I. S. (2014). IB Maths Resources from British International School Phuket. Retrieved from https://ibmathsresources.com/

Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., … Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science (New York, N.Y.)*, *300*(5616), 131–5. https://doi.org/10.1126/science.1084274

Pray-Grant, M. G., Daniel, J. a, Schieltz, D., Yates, J. R., & Grant, P. a. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature*, *433*(January), 434–438. https://doi.org/10.1038/nature03242

Probst, A. V, Dunleavy, E., & Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. *Nature Reviews. Molecular Cell Biology*, *10*(3), 192–206. https://doi.org/10.1038/nrm2640

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., … Bernstein, B. E. (2011). Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell*, *147*(7), 1628–1639. https://doi.org/10.1016/j.cell.2011.09.057

Reddington, J. P., Perricone, S. M., Nestor, C. E., Reichmann, J., Youngson, N. A., Suzuki, M., … Meehan, R. R. (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biology*, *14*(3), R25. https://doi.org/10.1186/gb-2013-14-3-r25

Riising, E. M., Comet, I., Leblanc, B., Wu, X., Johansen, J. V., & Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG Islands genome wide. *Molecular Cell*, *55*(3), 347–360. https://doi.org/10.1016/j.molcel.2014.06.005

Ringrose, L., & Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development (Cambridge, England)*, *134*(2), 223–32. https://doi.org/10.1242/dev.02723

Ringrose, L., Rehmsmeier, M., Dura, J. M., & Paro, R. (2003). Genome-

wide prediction of polycomb/trithorax response elements in Drosophila melanogaster. *Developmental Cell*, *5*(5), 759–771. https://doi.org/10.1016/S1534-5807(03)00337-X

Romero, I., Ruvinsky, I., & Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews. Genetics*, *13*(7), 505–516. https://doi.org/citeulike-article-id:10800937 doi: 10.1038/nrg3229

Saha, A., Wittmeyer, J., & Cairns, B. R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. *Nature Reviews. Molecular Cell Biology*, *7*(6), 437–447. https://doi.org/10.1038/nrm1945

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, *517*(7534), 321–326. https://doi.org/10.1038/nature14192

Schwartz, Y. B., & Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nature Reviews Genetics*, *8*(1), 9–22. https://doi.org/10.1038/nrg1981

Segal, E., & Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2009.01.004

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C. T., Bender, W., & Kingston, R. E. (1999). Stabilization of chromatin structure by PRC1, a polycomb complex. *Cell*, *98*(1), 37–46. https://doi.org/10.1016/S0092-8674(00)80604-2

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R., & Peterson, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, *311*(5762), 844–7. https://doi.org/10.1126/science.1124000

Simon, J. a, & Kingston, R. E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature Reviews. Molecular Cell Biology*, *10*(10), 697–708. https://doi.org/10.1038/nrm2763

Simon, J., Chiang, a, Bender, W., Shimell, M. J., & O'Connor, M. (1993). Elements of the Drosophila bithorax complex that mediate repression

by Polycomb group products. *Developmental Biology*. https://doi.org/10.1006/dbio.1993.1174

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., … Cordes, S. P. (2009). A Vertebrate Polycomb Response Element Governs Segmentation of the Posterior Hindbrain. *Cell*, *138*(5), 885–897. https://doi.org/10.1016/j.cell.2009.08.020

Son, J., Shen, S. S., Margueron, R., & Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes and Development*, *27*(24), 2663–2677. https://doi.org/10.1101/gad.225888.113

Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M., & Zaret, K. S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, *161*(3), 555–568. https://doi.org/10.1016/j.cell.2015.03.017

Srinivasan, L., & Atchison, M. L. (2004). YY1 DNA binding and PcG recruitment requires CtBP. *Genes and Development*, *18*(21), 2596–2601. https://doi.org/10.1101/gad.1228204

Srinivasan, S., Armstrong, J. a, Deuring, R., Dahlsveen, I. K., McNeill, H., & Tamkun, J. W. (2005). The Drosophila trithorax group protein Kismet facilitates an early step in transcriptional elongation by RNA Polymerase II. *Development (Cambridge, England)*, *132*(7), 1623–1635. https://doi.org/10.1242/dev.01713

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., … Schübeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*. https://doi.org/10.1038/nature10716

Stedman, E. (1950). Cell specificity of histones. *Nature*, *166*(4227), 780–781. https://doi.org/14780242

Stopka, T., & Skoultchi, A. I. (2003). The ISWI ATPase Snf2h is required for early mouse development. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(24), 14097–102. https://doi.org/10.1073/pnas.2336105100

Strohner, R., Németh, A., Nightingale, K. P., Grummt, I., Becker, P. B., & Längst, G. (2004). Recruitment of the nucleolar remodeling complex NoRC establishes ribosomal DNA silencing in chromatin. *Molecular and Cellular Biology*, *24*(4), 1791–8. https://doi.org/10.1128/MCB.24.4.1791-1798.2004

Struhl, K., & Segal, E. (2013). Determinants of nucleosome positioning. *Nature Structural & Molecular Biology*, *20*(3), 267–73. https://doi.org/10.1038/nsmb.2506

Svaren, J., & Hörz, W. (1997). Transcription factors vs nucleosomes: Regulation of the PHO5 promoter in yeast. *Trends in Biochemical Sciences*. https://doi.org/10.1016/S0968-0004(97)01001-3

Talbert, P. B., & Henikoff, S. (2016). Histone variants on the move: substrates for chromatin dynamics. *Nature Reviews. Molecular Cell Biology*. https://doi.org/10.1038/nrm.2016.148

Tam, P. P. L., & Behringer, R. R. (1997). Mouse gastrulation: The formation of a mammalian body plan. *Mechanisms of Development*. https://doi.org/10.1016/S0925-4773(97)00123-8

Teif, V. B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.-P., Marth, C., Höfer, T., & Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. *Nature Structural & Molecular Biology*, *19*(11), 1185–1192. https://doi.org/10.1038/nsmb.2419

Thatcher, T. H., & Gorovsky, M. A. (1994). Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic Acids Research*, *22*(2), 174–179. https://doi.org/10.1093/nar/22.2.174

Thornton, S. R., Butty, V. L., Levine, S. S., & Boyer, L. A. (2014). Polycomb repressive complex 2 regulates lineage fidelity during embryonic stem cell differentiation. *PLoS ONE*, *9*(10). https://doi.org/10.1371/journal.pone.0110498

Tien, C.-L., Jones, A., Wang, H., Gerigk, M., Nozell, S., & Chang, C. (2015). Snail2/Slug cooperates with Polycomb repressive complex 2 (PRC2) to regulate neural crest development. *Development (Cambridge, England)*, *2*(January), 1–10. https://doi.org/10.1242/dev.111997

Tiwari, V. K., Stadler, M. B., Wirbelauer, C., Paro, R., Schübeler, D., & Beisel, C. (2012). A chromatin-modifying function of JNK during stem cell differentiation. *Nature Genetics*, *44*(1), 94–100. https://doi.org/10.1038/ng.1036

Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z., & Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature*, *474*(7352), 516–20. https://doi.org/10.1038/nature10002

Van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., … Lander, E. S. (2010). Hi-C: A Method to Study the Three-dimensional Architecture of Genomes. *J. Vis. Exp*, (3910), 18693791–1869. https://doi.org/10.3791/1869

Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T., & Pasini, D. (2012). Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Research*, *40*(8), 3403–3418. https://doi.org/10.1093/nar/gkr1290

Verma, S. K., Tian, X., Lafrance, L. V., Duquenne, C., Suarez, D. P., Newlander, K. A., … Miller, W. H. (2012). Identification of potent, selective, cell-Active inhibitors of the histone lysine methyltransferase EZH2. *ACS Medicinal Chemistry Letters*, *3*(12), 1091–1096. https://doi.org/10.1021/ml3003346

Vignali, M., Hassan, A. H., Neely, K. E., & Workman, J. L. (2000). ATP-dependent chromatin-remodeling complexes. *Molecular and Cellular Biology*, *20*(6), 1899–1910. https://doi.org/10.1128/MCB.20.6.1899-1910.2000

Voullaire, L. E., Slater, H. R., Petrovic, V., & Choo, K. H. (1993). A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *American Journal of Human Genetics*, *52*(6), 1153–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7684888%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1682274

Waddington, C. H. (1942). The epigenotype. *International Journal of Epidemiology*, *41*(1), 10–13. https://doi.org/10.1093/ije/dyr184

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, *431*(7010), 873–8. https://doi.org/10.1038/nature02985

Wang, L., Brown, J. L., Cao, R., Zhang, Y., Kassis, J. A., & Jones, R. S. (2004). Hierarchical recruitment of polycomb group silencing complexes. *Molecular Cell*, *14*(5), 637–646. https://doi.org/10.1016/j.molcel.2004.05.009

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., & Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, *39*(4), 457–66. https://doi.org/10.1038/ng1990

West, M. H., & Bonner, W. M. (1980). Histone 2A, a heteromorphous family of eight protein species. *Biochemistry*, *19*(14), 3238–45. https://doi.org/10.1021/bi00555a022

Whitehouse, I., & Tsukiyama, T. (2006). Antagonistic forces that position nucleosomes in vivo. *Nature Structural & Molecular Biology*, *13*(7), 633–640. https://doi.org/10.1038/nsmb1111

Wiechens, N., Singh, V., Gkikopoulos, T., Schofield, P., Rocha, S., & Owen-Hughes, T. (2016). The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLoS Genetics*, *12*(3). https://doi.org/10.1371/journal.pgen.1005940

Wirbelauer, C., Bell, O., & Schübeler, D. (2005). Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes and Development*, *19*(15), 1761–1766. https://doi.org/10.1101/gad.347705

Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J., & Kingston, R. E. (2010). A Region of the Human HOXD Cluster that Confers Polycomb-Group Responsiveness. *Cell*, *140*(1), 99–110.

https://doi.org/10.1016/j.cell.2009.12.022

Wu, X., Johansen, J. V., & Helin, K. (2013). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. *Molecular Cell*, *49*(6), 1134–1146. https://doi.org/10.1016/j.molcel.2013.01.016

Yan, Z., Wang, Z., Sharova, L., Sharov, A. A., Ling, C., Piao, Y., … Ko, M. S. H. (2008). BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. *Stem Cells (Dayton, Ohio)*, *26*(5), 1155–65. https://doi.org/10.1634/stemcells.2007-0846

Yen, K., Vinayachandran, V., Batta, K., Koerber, R. T., & Pugh, B. F. (2012). Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell*, *149*(7), 1461–1473. https://doi.org/10.1016/j.cell.2012.04.036

Yen, K., Vinayachandran, V., & Pugh, B. F. (2013). XSWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. *Cell*, *154*(6). https://doi.org/10.1016/j.cell.2013.08.043

Yuan, G.-C., Liu, Y.-J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., & Rando, O. J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. *Science (New York, N.Y.)*, *309*(5734), 626–30. https://doi.org/10.1126/science.1112178

Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S., & Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell*, *95*(2), 279–289. https://doi.org/10.1016/S0092-8674(00)81758-4

Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J., & Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science (New York, N.Y.)*, *322*(5902), 750–6. https://doi.org/10.1126/science.1163045

Zhou, C. Y., Johnson, S. L., Gamarra, N. I., & Narlikar, G. J. (2016). Mechanisms of ATP-Dependent Chromatin Remodeling Motors.

Annual Review of Biophysics, *45*(1), 153–181. https://doi.org/10.1146/annurev-biophys-051013-022819

Zhou, V. W., Goren, A., & Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nature Reviews. Genetics*, *12*(1), 7–18. https://doi.org/10.1038/nrg2905

Zhou, Y., & Grummt, I. (2005). The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing. *Current Biology*, *15*(15), 1434–1438. https://doi.org/10.1016/j.cub.2005.06.057

Zuin, J., Dixon, J. R., van der Reijden, M. I. J. A., Ye, Z., Kolovos, P., Brouwer, R. W. W., … Wendt, K. S. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci USA*, *111*(3), 996–1001. https://doi.org/10.1073/pnas.1317788111